Evaluation of Surface Analyses and Forecasts with a Multiscale Ensemble Kalman Filter in Regions of Complex Terrain

BRIAN C. ANCELL,* CLIFFORD F. MASS, AND GREGORY J. HAKIM

University of Washington, Seattle, Washington

(Manuscript received 31 August 2010, in final form 8 December 2010)

ABSTRACT

Previous research suggests that an ensemble Kalman filter (EnKF) data assimilation and modeling system can produce accurate atmospheric analyses and forecasts at 30–50-km grid spacing. This study examines the ability of a mesoscale EnKF system using multiscale (36/12 km) Weather Research and Forecasting (WRF) model simulations to produce high-resolution, accurate, regional surface analyses, and 6-h forecasts. This study takes place over the complex terrain of the Pacific Northwest, where the small-scale features of the near-surface flow field make the region particularly attractive for testing an EnKF and its flow-dependent background error covariances. A variety of EnKF experiments are performed over a 5-week period to test the impact of decreasing the grid spacing from 36 to 12 km and to evaluate new approaches for dealing with representativeness error, lack of surface background variance, and low-level bias. All verification in this study is performed with independent, unassimilated observations.

Significant surface analysis and 6-h forecast improvements are found when EnKF grid spacing is reduced from 36 to 12 km. Forecast improvements appear to be a consequence of increased resolution during model integration, whereas analysis improvements also benefit from high-resolution ensemble covariances during data assimilation. On the 12-km domain, additional analysis improvements are found by reducing observation error variance in order to address representativeness error. Removing model surface biases prior to assimilation significantly enhances the analysis. Inflating surface wind and temperature background error variance has large impacts on analyses, but only produces small improvements in analysis RMS errors. Both surface and upper-air 6-h forecasts are nearly unchanged in the 12-km experiments. Last, 12-km WRF EnKF surface analyses and 6-h forecasts are shown to generally outperform those of the Global Forecast System (GFS), North American Model (NAM), and the Rapid Update Cycle (RUC) by about 10%–30%, although these improvements do not extend above the surface. Based on these results, future improvements in multiscale EnKF are suggested.

1. Introduction

The creation of accurate, high-resolution atmospheric surface analyses and short-term forecasts is important for many reasons. National Weather Service (NWS) forecasters require fine-scale surface analyses to provide awareness of current atmospheric conditions and to verify forecasts. High-quality, short-term (0–6 h) forecasts are useful for the effective use of alternative energy sources such as wind or solar power, and for the prediction of small-scale, high-impact events, such as convection, forest fires, and poor air quality.

In general, gridded analyses are produced by combining observations and a previous model forecast through a data assimilation system. Recent studies suggest that an ensemble Kalman filter (EnKF; Evensen 1994) has the potential to create accurate three-dimensional analyses and forecasts across a wide range of scales when assimilating surface data using numerical weather prediction models. For example, Hacker and Snyder (2005) found large background correlations between surface and boundary layer atmospheric variables within an EnKF using a one-dimensional (1D) column model. Subsequent 24–48-h forecast boundary layer errors were reduced when simulated surface observations were assimilated. In a similar study, Hacker and Rostkier-Edelstein (2007) showed that an EnKF could effectively spread information from real surface wind, temperature, and mixing ratio

* Current affiliation: Texas Tech University, Lubbock, Texas.

Corresponding author address: Dr. Brian Ancell, Atmospheric Science Group, Department of Geosciences, Texas Tech University, Box 42101, Lubbock, TX 79409.
E-mail: brian.ancell@ttu.edu

DOI: 10.1175/2010MWR3612.1

© 2011 American Meteorological Society
observations into the boundary layer during assimilation. Moreover, Hacker and Rostkier–Edelstein showed that boundary layer potential temperature, wind, and mixing ratio analysis errors were reduced by surface data assimilation. In ensemble data assimilation experiments with a global model, Whitaker et al. (2004) found a positive impact from the surface to the middle troposphere from assimilating surface data. In an effort to recreate the 500-hPa flow field in the early twentieth century, they showed that 500-hPa analysis errors were comparable with modern 2–3-day forecast errors when assimilating only sparse surface pressure observations. Both Hacker and Rostkier–Edelstein (2007) and Whitaker et al. (2004) showed clear analysis improvements for the EnKF over a three-dimensional data assimilation system (3DVAR). Although these studies present encouraging results regarding the assimilation of surface observations with an EnKF, Hacker and Snyder (2005) cite the need to expand this investigation using real observations from current, higher-density mesonet works.

Recent work has begun to evaluate EnKF forecasts in an experimental context more representative of current mesoscale models. Using the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5) EnKF at 30-km horizontal resolution, Meng and Zhang (2007) showed that an EnKF improved the forecasts of a mesoscale convective vortex (MCV) event, particularly when a multiphysics approach was employed. For the same MCV event and at similar resolution, Meng and Zhang (2008a) found that Weather Research and Forecasting (WRF) model EnKF forecasts outperformed those of a similarly configured WRF 3DVAR system. Fujita et al. (2007) showed that 0–18-h forecasts improved when using an MM5 model EnKF that assimilated real surface observations for a limited number of cases. Meng and Zhang (2008b) conducted WRF EnKF and 3DVAR experiments for 60 cases using radiosonde data for both assimilation and verification, and found that the EnKF on average outperformed 3DVAR for 60-h forecasts. Torn and Hakim (2008, hereafter TH08) found that EnKF forecasts at 45-km grid spacing were competitive with operational global models for up to 24 h, even when assimilating significantly less data.

The above results all suggest the potential of an EnKF to produce accurate regional analyses and forecasts. However, to our knowledge, no previous study includes both 1) an extended time period showing statistically significant, regional EnKF performance and 2) verification of both analyses and short-term forecasts against a large set of independent observations. Furthermore, we are not aware of the evaluation of EnKF analyses and forecasts at relatively high resolution (~10-km grid spacing) over large regions. The primary goals of this study are to evaluate the overall performance of a multiscale 36/12-km WRF-based EnKF in creating regional surface analyses and 6-h forecasts, and to perform an initial examination of the quality of three-dimensional fields in order to motivate further research on the performance of a regional, multiscale EnKF.

This study is conducted over the Pacific Northwest, an area characterized by complex terrain. Since complex terrain strongly influences low-level flow and produces small-scale features that are highly variable in space and time, an EnKF, with its flow-dependent covariances, could be of particular value in such regions. In this paper, analyses and forecasts produced with an EnKF will be compared to current operational products that utilize different data assimilation schemes, but that are run at similar resolution to the EnKF configuration used here: the Global Forecast System (GFS; ~36 km), the North American Model (NAM; 12 km), and the Rapid Update Cycle (RUC; 13 km). Section 2 provides a background on the modeling system and presents the experimental design. Section 3 gives experimental results and discussion, and a summary and conclusions are provided in section 4.

2. Background and methodology

The EnKF data assimilation method updates a short-term ensemble forecast, often called the background, with observations to obtain an analysis. The EnKF used in this study is the same formulation as that of TH08, which employs an ensemble square root filter (EnSRF) that assimilates observations serially (see Whitaker and Hamill 2002 for a full derivation of the EnSRF). A variety of data assimilation parameters control how the EnKF assimilates observations, such as the localization radius that determines the spatial extent to which observations influence the analysis (Gaspari and Cohn 1999) and the inflation of background covariance, which accounts for small ensemble size and helps mitigate filter divergence (Anderson and Anderson 1999).

This study uses a multiscale EnKF for an outer domain with 36-km grid spacing and a nested domain with a 12-km grid; each grid utilizes independent data assimilation parameters. The 36-km grid uses perturbed boundary conditions about the GFS forecast, drawn from a fixed covariance matrix (Torn et al. 2006), whereas the 12-km grid uses boundary conditions provided by the 36-km grid. Analysis variance inflation after the assimilation of observations is used on both domains (see TH08 for details). The 36-km grid uses the same variance inflation from TH08 (i.e., an analysis inflation factor of 0.17, which indicates the fraction of reduced variance during assimilation that is
retained at the end of the assimilation cycle), which was tuned to produce appropriate ensemble spread, whereas the 12-km grid uses a larger inflation factor (0.35, resulting in less variance inflation) since it is assumed large variance enters the 12-km domain from the 36-km boundary conditions. Tests varying the analysis variance inflation factor from 0.17 to 0.5 on the 12-km grid showed practically no change in resulting surface analyses (not shown). The localization radius also varies from the 36- to the 12-km grid. The 36-km grid uses a longer localization radius (2000 km, as in TH08) than the 12-km domain (200 km, chosen subjectively to represent local mesoscale features in the Pacific Northwest).

A 6-h assimilation cycle is applied, and each domain contains 80 ensemble members. The forecast model used in this study is the WRF model version 2.1.2. The physics used within the WRF model for both domains are the Mellor–Yamada–Janjic (MYJ) planetary boundary layer scheme (Janjic 1990, 1996, 2002), the Kain–Fritsch cumulus parameterization (Kain and Fritsch 1990, 1993), the Noah land surface model (Chen and Dudhia 2001), the WRF Single-Moment 3-class microphysics (Hong et al. 2004), the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al. 1997), and the Dudhia shortwave radiation scheme (Dudhia 1989). Cloud-track wind, Aircraft Communication, Addressing, and Reporting System (ACARS) aircraft (temperature and wind), radiosonde (temperature, wind, and relative humidity), and surface observations (temperature, wind, altimeter, etc.) are assimilated on both the 36- and 12-km domains. The surface observations used in TH08 included Automated Surface Observing System (ASOS) data, buoy data, and ocean vessel data. In this study, additional observational assets from independent federal, state, county, and city government networks, as well as private entities such as railroad companies, are used to provide a larger dataset for the experiments performed here. Figure 1 shows both the 36-km outer domain and the nested 12-km domain.

Three practical aspects of surface data assimilation using a WRF EnKF can significantly affect the quality of surface analyses. These issues are 1) terrain mismatches between the model and the actual terrain at and near observation locations that may cause representativeness error, 2) lack of variance in model surface wind and temperature fields, and 3) model bias. These issues and the methods used in this study to overcome them are described below.

a. Representativeness errors in terrain

Although using higher resolution, the nested grid still fails to resolve significant smaller-scale terrain features that can affect observations. Since the effects of such unresolved terrain do not affect the model’s estimate of the observations, the resulting error (representativeness error) can be a major problem during data assimilation. Representativeness error can be partially addressed in an EnKF by prescribing larger observation error variances for observations that are likely affected by poorly represented terrain. By assuming larger surface observation variances, less weight is given to potentially unrepresentative observations during assimilation. This was the case in TH08 (observation standard deviations of 2.5 m s\(^{-1}\) for 10-m winds, 1.8 K for 2-m temperature). These inflated observation variances allowed more weight to be placed on the background during assimilation since representativeness error can be extreme in mountainous terrain with the 45-km grid spacing used in TH08.

One problem with inflating surface observation variance is the loss of valuable surface observational information where representativeness error is not significant. In this
study we propose a method that excludes surface observations where terrain-induced representativeness error is likely and that assimilates surface observations where the model well represents the actual terrain. With this method we aim to reduce representativeness errors in places where they are likely dominated by a poor model characterization of terrain. This methodology is applied only to the 12-km domain, and the observation variance values from TH08 are used on the 36-km grid.

To estimate which observations are representative, we compare the model terrain to the actual terrain (represented by a 1.33-km terrain height grid) at and near observation locations. For each observation, the closest grid point on the 1.33-km real terrain grid is identified. The real terrain height is then compared to the interpolated model terrain height over a 5 gridpoint by 5 gridpoint box within the 1.33-km terrain height grid centered near the observation. An observation is not assimilated if the difference in terrain height exceeds 200 m at any of the 25 real terrain height grid points, or at the observation location itself. This extends the technique used in other studies (e.g., TH08) that discard observations based on observation and model terrain height differences at only the observation location. This terrain check is performed over roughly a 5 km by 5 km area and is meant to remove observations with any significant errors due to the model’s misrepresentation of the terrain near the observation, yet retains enough data to compose a significant sample size. Nonetheless, it should be noted that additional tuning of the horizontal size over which this terrain check occurs, as well as the height error threshold of 200 m, may improve this technique, and we have not tested these parameters in detail. We reduce the 10-m wind and 2-m temperature observation error standard deviations from 2.5 m s\(^{-1}\) and 1.8 K used in TH08 to 1.5 m s\(^{-1}\) and 1.0 K, similar to observation error standard deviations associated with lower atmosphere radiosonde observations.

Figure 1 depicts the 12-km WRF EnKF domain with the observation locations (for both assimilation and verification) that passed the terrain check for a single assimilation time; this distribution is typical of the number of observations used in the experiments of this study for each assimilation cycle. Approximately 3000–5000 surface observation locations are available over the 12-km domain for each assimilation cycle (0000, 0600, 1200, and 1800 UTC), although each location does not necessarily provide every observation type. Performing the terrain check described above eliminates around 80% of these observations. The majority (~75%) of the National Weather Service ASOS observations are retained, but many observations from the additional land-based networks are eliminated, especially near the highest and steepest terrain. Although the number of eliminated observations is large, hundreds of wind, temperature, and altimeter observations remain. This remaining set of observations is split in half for each experiment by choosing either odd or even number indices from the observation file, the first set (odd) is used for assimilation and the second set (even) is used for independent verification (see Fig. 1).

b. Lack of variance in the surface wind and temperature fields

In a numerical model, surface winds and temperature are highly influenced by the boundary layer and land surface schemes. In an ensemble, the boundary layer scheme and land surface model will strongly influence the variance of winds and temperature near the surface. For example, if the effects of surface friction are large, variance in the wind speed at the top of the boundary layer may be decreased significantly near the surface. Similarly, surface temperature can be highly constrained by the model’s land surface scheme. Since the ensemble surface wind and temperature variance plays an important role in surface observation assimilation, too little variance caused by the influence of surface physics parameterizations can result in the EnKF being overly confident in the first-guess fields. This in turn can reduce the impact of potentially important surface observations and cause filter divergence. In this situation, forecasts become significantly different from reality, and resulting EnKF analyses are poor since the EnKF has too much confidence in its unrealistic background state.

Figure 2 depicts the background wind speed and observation wind speed variance, as well as the background temperature and observation temperature variance, throughout the troposphere from the 12-km WRF EnKF. The background variance values were averaged over every model horizontal grid point over a week of data assimilation cycles from 0000 UTC 15 October to 1800 UTC 21 October 2007. The observation variance used for this plot is the same as that used for surface and radiosonde observations in TH08. On average, the background wind speed variance is nearly twice as large as the observation variance at all levels except the surface, where the background variance value drops to roughly half the value of the observation variance. In this situation, surface observations are given little weight during assimilation and beneficial observational information may be lost. In the temperature field, the difference between the background and observation variance is largest at the surface and decays toward the tropopause, where the variance values are comparable. The smallest background variance is at the surface, which suggests that there may be issues with the use of deterministic land surface characteristics
such as soil temperature or moisture as is surmised in Stensrud et al. (2000), Hacker and Snyder (2005), and Sutton et al. (2006). In any case, relatively small background temperature variance coupled with inflated observation variance at the surface creates a similar situation to that with the wind field, which is that little weight is given to observations.

To determine whether the relatively small variance at the surface in the 12-km WRF EnKF model wind and temperature background fields are associated with too little spread, error-to-variance ratios are examined. As described in Houtekamer et al. (2005) and used in TH08, error-to-variance ratios can be examined to determine whether an ensemble has appropriate spread (Murphy 1988). Error-to-variance ratios $E$ can be expressed in the context of a discrete ensemble as

$$E = \frac{[(Y - Hx_b)^T \times (Y - Hx_b)]}{\text{trace}(HBH^T + R)},$$  \hspace{1cm} (1)

where $Y$ is a column vector of observations, $H$ is a matrix that maps model variables from the model state vector $x_b$ to observation locations by interpolating from the neighboring four model grid points, $B$ is the background error covariance matrix estimated from the ensemble background field, and $R$ is the observation error covariance matrix, assumed to be diagonal with no correlation between observation errors. The column vector $(Y - Hx_b)$ is the innovation, a column vector representing the difference between the observations and the ensemble-mean estimates of the observations. The denominator represents the innovation variance, or background variance plus the observation variance summed over all observations. Qualitatively, if $E$ is very large, EnKF mean forecast errors are generally much larger than the ensemble innovation variance (too little spread), and if $E$ is very small, errors are much smaller than the innovation variance (too much spread).

TH08 investigated surface wind and temperature error-to-variance ratios with a 45-km WRF EnKF over a 2-yr period, and found $E$ values close to 1. We have calculated the surface wind and temperature error-to-variance ratios over a 5-week period with the 12-km WRF EnKF using the same observation variance values used in TH08 (referred to as the 12-km BENCHMARK case). In contrast to TH08, the 12-km BENCHMARK case WRF EnKF surface wind error-to-variance ratio is 3.4. Error-to-variance ratios aloft range from 1.2 to 1.8, which more closely resemble the results of TH08. The error-to-variance ratio for surface wind is about twice that for wind aloft, indicating that too little variance exists in the 12-km EnKF background surface wind field. For temperature, the surface error-to-variance ratio for the 5-week period is 1.8, and values at and above 850 hPa range from 0.9 to 1.2, again indicating a lack of surface spread relative to that aloft. Since we reduce surface wind and temperature observation variance for the 12-km WRF EnKF, which will cause the denominator in Eq. (1) to become smaller, the lack of surface variance is likely to become a more pronounced issue.

One method to combat ensemble background variance deficiencies at the surface involves using various land surface characteristics (e.g., surface roughness, moisture content, or conductivity) or different physics schemes within the ensemble, which can increase variance and improve ensemble forecasts (Stensrud et al. 2000; Sutton et al. 2006; Fujita et al. 2007). However, in this study, we
propose inflating surface variance using the larger values found aloft, a technique motivated by the results shown in Fig. 2. This method preserves the spatial distribution of background variance, can be applied generally, and may be an effective solution to increase surface variance without running a multimodel ensemble. It should be noted that the results achieved using this method apply only to the specific modeling/physics configuration used here, and may not extend to other EnKF applications that utilize different physics schemes than those used in this study and in TH08. Specifically, the $u$ and $v$ surface wind and temperature background variance at each grid point is set to be the maximum variance in these variables directly aloft between model levels 11 and 16 (generally between about 900 and 800 hPa, a zone intended to capture the variance at the top of the boundary layer). Only an increase in surface variance is allowed, so if the variance aloft is smaller than the surface variance, no change is made.

c. Model bias

Model bias can be communicated three-dimensionally during data assimilation through EnKF covariances, and can subsequently contaminate analyses. Dee and DaSilva (1998) note that, in general, data assimilation systems assume that forecast errors are strictly random and have zero mean. They further explain that deviations from this assumption will adversely affect analysis quality, and that the only way to properly account for forecast bias in an operational, statistical-analysis scheme is to do so explicitly by estimating the bias and removing it prior to the assimilation cycle. More recently, Dee (2005) describes the positive effects of a data assimilation system in which biases are estimated offline and bias removal is applied to the background prior to the analysis cycle. In this study, we attempt to remove forecast bias in this way prior to data assimilation in an effort to avoid degrading the analysis. Figure 3 shows the bias in the mean 6-h WRF EnKF 12-km forecast (background) surface temperature field valid at 0000, 0600, 1200, and 1800 UTC for the period 0000 UTC 17 September–1800 UTC 21 October 2007. These plots reflect the bias averaged over all observations that have passed the terrain check at each assimilation cycle. Although significant day-to-day fluctuations occur, a cold bias is apparent in the background field for virtually all assimilation cycles. This cold bias is most significant in the 1800 and 0000 UTC background fields (average of $-1.85$ and $-1.87$ K over all 1800 and 0000 UTC cycles, respectively), and is slightly smaller in the 0600 and 1200 UTC cycles ($-1.59$ and $-1.07$ K).

In terms of wind, Fig. 4 shows the background surface wind speed and direction biases averaged over all representative surface observations for all assimilation cycles over the same period as in Fig. 3. Wind direction bias is only shown for observation wind speeds greater than $1.54$ m s$^{-1}$, as observed wind direction is generally inaccurate for low wind speeds. No significant directional bias is apparent since the average mean error over all cycles is approximately zero. In contrast, wind speed biases are found to be large. When the observed wind speed is less than $1.54$ m s$^{-1}$, the model bias is $2.47$ m s$^{-1}$. Two other categorical model wind speed biases were noticed when observation wind speeds were greater than $1.54$ m s$^{-1}$: a slow bias ($-1.76$ m s$^{-1}$) occurred when model wind speeds were less than $2.57$ m s$^{-1}$, and a fast bias ($2.35$ m s$^{-1}$) occurred when model wind speeds were greater than $5.14$ m s$^{-1}$.

To reduce the effects of the biases in the surface wind and temperature fields, a crude bias removal technique is used for each data assimilation cycle. Model surface temperature and wind biases are calculated at each representative observation location using the mean background (6-h forecast) field. These biases are averaged over the entire domain for a given assimilation cycle, and then averaged over an entire week of cycles. This averaging is first performed over the week of cycles prior to the five weeks of data assimilation experiments tested here in order to provide a representative bias for the first week of assimilation experiments, and the average bias over each of the first four weeks of experiments is used as a representative bias for each subsequent week of experiments. The representative biases (calculated over each prior week) are removed from every model grid point in each background member prior to assimilation, and are held constant over each week of the experiments. This bias-averaging technique is meant to roughly capture the recent biases experienced in the data assimilation/forecasting system. For temperature, these average values are calculated at each of the 0000, 0600, 1200, and 1800 UTC cycles, whereas for wind speed, bias is averaged over all cycles for the three categories discussed above and shown in Fig. 4. The difference between the average wind speed from the closest five observations to each model grid point, and the wind speed at the model grid point itself, are used to determine that model grid point’s speed bias category.

d. Experimental design

The quality of surface analyses and short-term forecasts are measured by the root-mean-square (RMS) differences, and in some cases mean absolute differences, of a set of independent, unassimilated surface wind and temperature observations. These unassimilated observations are the same for each experiment, and within each experiment a set of identical observations are assimilated, including surface wind, temperature, and altimeter observations; ACARS aircraft wind and temperature data; cloud-track wind data; and radiosonde wind, temperature, and relative
humidity data. All of the surface wind and temperature observations used in these experiments, both assimilated and unassimilated, have passed the terrain check discussed above. The experiments are performed over a 5-week period from 0000 UTC 17 September to 1800 UTC 21 October 2007, with the exception of the experiment that tests background variance inflation, which is performed over a 1-week period from 0000 UTC 15 October to 1800 UTC 21 October 2007.

The first experiment is designed to examine the effects of increased resolution alone. Two cases are run, one is performed on the 36-km outer domain and the other on a 12-km nested domain (BENCHMARK) using the relatively large surface wind and temperature observation variances from TH08. A second case (REDVAR) explores using reduced surface wind and temperature observation error variances on the 12-km WRF EnKF domain. Using these reduced observation error variances, two more experiments are performed independently testing the effects of background surface wind and temperature inflation (BGINF), as well as background surface wind and temperature bias removal (BIASREM). Last, a fifth experiment (ALL) uses observation variance reduction, background surface inflation, and background surface bias removal simultaneously. The WRF EnKF configuration with the smallest errors is then compared to the GFS, NAM, and RUC analyses and 6-h forecasts over the same 5-week period. The same bias removal scheme used in the BIASREM experiment is applied to each of these operational models in order to make a fair comparison.

FIG. 3. Bias (K) in the background surface temperature field averaged over all representative surface observations at (a) 0000, (b) 0600, (c) 1200, and (d) 1800 UTC 17 Sep–21 Oct 2007.
3. Results and discussion

a. 36- versus 12-km horizontal resolution

The results of the 36- and 12-km experiments, which assimilate identical observations and use the same observation variance values, are shown in Fig. 5, and the average RMS values for this and other experiments are summarized in Table 1. The 12-km domain produces a closer fit to observations than the 36-km domain for both the 6-h forecast and the analysis. The resolution-dependent improvement in the 6-h forecast of the surface wind is 11%, and for temperature it is 6%. Improvements in the 12-km analysis (15% for wind, 14% for temperature) are larger than for the 6-h forecast. The analysis RMS fit to altimeter observations (not shown), which were assimilated on both the 36- and 12-km domains, are improved by about 14% with the 12-km EnKF, whereas the forecast improvement is about 6%. All of these improvements are significant at the 95% confidence level using a one-sided Student’s $t$ test.

Figure 6 shows both the background sea level pressure field as well as the analysis increments made to the surface zonal wind component in the 36- and 12-km domains for a single assimilation cycle. These analysis increments illustrate the effects of finer-scale covariances from EnKF data assimilation at higher resolution. The finer structural detail on the 12-km grid is most noticeable near prominent features in complex terrain, such as in western...
Washington, the Strait of Georgia north of Vancouver Island, and in Utah. This suggests the greatest benefits of increasing the resolution within an EnKF may occur at and near complex terrain, and future studies will investigate whether the results found here are reproduced over more heterogeneous terrain.

b. 12-km reduced observation variance

The remaining EnKF experiments were performed on the 12-km grid. The first examines the effects of reduced surface wind and temperature observation variances that are made possible by the terrain check. Figure 7 shows the RMS analysis differences for assimilated and unassimilated surface wind and temperature observations for both the BENCHMARK case (large observation variances as in the 36-km WRF EnKF), as well as the case with reduced observation variance values (REDVAR). Examining the fit to both assimilated and unassimilated observations can provide insight on how altering observation variances at assimilated observation locations spreads information to unassimilated observation locations. In REDVAR the $u$-wind and $v$-wind

<table>
<thead>
<tr>
<th>Case name</th>
<th>Analysis Assimilated</th>
<th>Unassimilated</th>
<th>Background (6-h forecast)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wind (m s$^{-1}$)</td>
<td>Temperature (K)</td>
<td>Altimeter (hPa)</td>
</tr>
<tr>
<td>36 km</td>
<td>2.72</td>
<td>2.22</td>
<td>1.68</td>
</tr>
<tr>
<td>12-km BENCHMARK</td>
<td>1.89</td>
<td>1.83</td>
<td>1.44</td>
</tr>
<tr>
<td>REDVAR</td>
<td>1.63</td>
<td>1.55</td>
<td>1.36</td>
</tr>
<tr>
<td>BIASREM</td>
<td>1.65</td>
<td>1.44</td>
<td>1.40</td>
</tr>
<tr>
<td>ALL</td>
<td>1.29</td>
<td>1.36</td>
<td>1.41</td>
</tr>
<tr>
<td>GFS</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>NAM</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RUC</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>REDVAR (1 week)</td>
<td>1.80</td>
<td>1.41</td>
<td>0.55</td>
</tr>
<tr>
<td>BGINF (1 week)</td>
<td>1.48</td>
<td>1.34</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Observation standard deviations were reduced from 2.5 to 1.5 m s$^{-1}$, and temperature observation standard deviation were dropped from 1.8 to 1.0 K.

Large reductions in RMS analysis error for assimilated observations (14% for wind, 15% for temperature) occur when reduced observation variance is applied at the surface. Such improvement occurs because reducing observation variance puts more weight on the observations. The analysis RMS fit to assimilated altimeter observations is also improved (6%, not shown) with reduced surface wind and temperature observation variance.

Improvements in analysis RMS errors are also realized when verifying against independent, unassimilated observations using reduced surface wind and temperature observation variances. These improvements, as shown in Fig. 7, are smaller than those associated with assimilated observations, likely a consequence of the reduced weight the analysis increments are given farther from the assimilated observation locations. Nonetheless, a better analysis RMS fit to unassimilated observations is achieved for both wind (4%) and temperature (9%). All of the improvements in analysis RMS differences for this experiment are statistically significant at the 95% confidence level. These results suggest that observations passed through the terrain check did not introduce significant representativeness error.

In contrast to analyses, the RMS fit to surface observations of the 6-h forecasts hardly changes when reducing the surface wind and temperature observation variance, with the difference in the background RMS fit to surface temperature (2%), wind (0%), and altimeter (1%) for the cases with and without observation variance reduction. We speculate that this insensitivity in the 6-h forecasts is a result, at least in part, of the identical boundary conditions, supplied from the 36-km WRF EnKF. Since very few observations exist west of the North American coastline, the initial conditions over the Pacific Ocean are very similar for both experiments. In 6 h, these similar oceanic initial conditions can advect into a portion of the 12-km domain, resulting in similar 6-h forecasts. Since the lower boundary conditions are not updated during data assimilation, the same lower boundary forcing used in both experiments may also contribute to the insensitivity of forecasts to changes in observation error variance.

Furthermore, the analysis increments of both experiments, while different at the surface, are not sufficiently different aloft to produce significant changes in the 6-h forecast. Figure 8 depicts the analysis increment with respect to 700-hPa geopotential height (GPH) for the BENCHMARK and REDVAR cases. Very similar structures are evident in both cases, with only small differences in magnitude. Inspection of wind and temperature analysis increments at 700 hPa reveals a similar situation (not shown). The resulting differences in the analyzed 700-hPa GPH field between the cases are also shown in Fig. 8, as are the differences in the 6-h forecast. It is clear that the already small 700-hPa analysis differences between the two runs decay over the 6-h forecast to the point where the two cases are indistinguishable. This situation occurs frequently throughout the 5 weeks of assimilation cycles with this and the other experiments in this study, and provides strong evidence that on the average, short-term forecasts are insensitive to varying the surface data assimilation parameters tested here. However, it is possible that more noticeable reductions in forecast errors associated with specific flow configurations may be realized as in Fujita et al. (2007).
be particularly true in cases that exhibit strong initial condition sensitivity, since localized initial condition differences do exist as shown in Fig. 8c.

c. Background surface temperature and wind variance inflation

The variance in the background surface wind and temperature fields in the 12-km WRF EnKF was shown to be too small through examination of error-to-variance ratios, and an experiment was designed to test the effects of inflating surface variances with values aloft. Two cases are considered in this experiment, those with (BGINF) and without (REDVAR) background surface wind and temperature inflation. Both of these cases use reduced observation variance as described earlier. Larger variances in the background allow a closer fit to assimilated observations, and indeed, the RMS fit of the analysis to assimilated surface wind and temperature observations is improved (18% for wind, 5% for temperature) using surface background inflation in the BGINF experiment (Fig. 9). The smaller impact on the temperature field is due to the fact that background temperature variance does not increase with height as significantly as the wind field variance, and thus less inflation is applied to surface background temperature. The inflated surface wind and temperature background variances had little effect on the RMS fit of the analysis to assimilated altimeter observations.

Fig. 7. Analysis RMS differences for both the 12-km BENCHMARK experiment (solid) and the REDVAR experiment (dashed) with respect to (a) assimilated surface wind observations, (b) unassimilated surface wind observations, (c) assimilated surface temperature observations, and (d) unassimilated surface temperature observations from 0000 UTC 17 Sep to 1800 UTC 21 Oct 2007.
Figure 9 also shows the RMS fit of the analysis to unassimilated surface wind and temperature observations for this experiment. The relatively large improvements in the fit to assimilated observations do not extend to unassimilated observations: improvements are quite small (2% for both wind and temperature), and are not statistically significant at the 95% confidence level. However, the mean absolute difference between the REDVAR and BGINF analyses at unassimilated observation locations (averaged over the week of assimilation) is relatively large (0.20 K for temperature, 0.47 m s\(^{-1}\) for wind). This indicates that although the REDVAR and BGINF analyses are practically indistinguishable with regard to their RMS analysis errors, they are actually quite different. In turn, inflation of surface background variance with values aloft has a large effect on the analysis, but a small effect on RMS errors. Future efforts will focus on tuning this variance inflation method to optimize error reduction since the method has been shown here to impact the analysis field. It would also be interesting to compare the variance inflation used here to multiphysics methods that have also been shown to increase ensemble variance (Stensrud et al. 2000; Sutton et al. 2006; Fujita et al. 2007). Such a comparison may reveal the relative merits of an EnKF with a single set of physics schemes using the variance inflation proposed here over that of a multiphysics EnKF configuration that may be difficult to properly configure if different physics schemes are not equally skillful as discussed in Stensrud et al. (2000). As with the REDVAR experiment, little to no forecast improvement occurred by inflating surface background wind and temperature variance.

d. Bias removal

Model surface wind and temperature biases were shown to be large, and a simple bias removal technique was described in section 2c. Two cases were run to test
the effects of bias removal, one with no bias removal (REDVAR) and the other with both surface background wind and temperature bias removal (BIASREM). Both cases use reduced observation variances, and neither uses background surface variance inflation. Figure 10 depicts the background RMS errors with respect to unassimilated surface wind and temperature observations with and without bias removal. For wind, there is a reduction of the background (6-h forecast) RMS error by 18%. Slightly less improvement in background error occurs with the temperature field (13%). Both these improvements are significant at the 95% confidence level.

Figure 10 also shows the analysis RMS errors with respect to unassimilated surface wind and temperature observations for REDVAR and BIASREM cases. A 6% reduction in the RMS fit to temperature observations, and a 9% improvement in the RMS fit to wind observations is observed. Like the reduced errors in the 6-h forecast field, both analysis wind and temperature RMS reductions are statistically significant at the 95% confidence level. Although the bias removal scheme used in this study is crude, these results indicate that bias removal from the background can improve the analysis. It is likely that surface temperature and wind biases vary spatially with the land surface and terrain height inhomogeneity of the Pacific Northwest, and more sophisticated bias removal schemes should be explored.
e. Combined surface variance inflation and bias removal

A final experiment was performed using both background variance inflation and bias removal (ALL). Reductions in the RMS errors of the analyses with unassimilated observations compared to the REDVAR experiment (neither variance inflation nor bias removal) are 9% for wind and 7% for temperature. These reductions are statistically significant at the 95% confidence interval, and are essentially the same as that for bias removal only, indicating the significant influence of bias removal over surface variance inflation in this experiment. The 6-h forecasts within the ALL and BIASREM experiments were also nearly identical. Thus, for the experiments performed here, the ALL and BIASREM cases are essentially indistinguishable, and provide the best analysis estimates over the range of assimilation cycles tested. It should be noted that both variance inflation and bias removal are only performed at the surface in this study, and to extend the impacts shown here above the surface, these issues must be addressed in the near-surface environment as well.
f. 12-km WRF EnKF compared to operational systems

The ensemble mean from the 12-km WRF EnKF configuration using reduced observation variance, background surface variance inflation, and bias removal (ALL) is compared to the 6-h forecasts and analyses from the National Weather Service GFS, NAM, and RUC modeling systems. These models were interpolated to the WRF 12-km grid prior to verification. Biases were removed from each of the GFS, NAM, and RUC 6-h forecasts and analyses as in the 12-km WRF EnKF experiments, allowing for a fair comparison. Figure 11 depicts the RMS errors for unassimilated surface temperature and wind observations from 0000 UTC 17 Sep to 1800 UTC 21 Oct 2007 for the best 12-km WRF EnKF configuration (red), GFS model (blue), NAM model (green), and RUC model (black).

![Figure 11](https://example.com/figure11.png)

**FIG. 11.** Background RMS differences from (a) unassimilated surface wind observations and (b) unassimilated surface temperature observations, as well as analysis RMS differences also from (c) unassimilated surface wind observations, and (d) unassimilated surface temperature observations from 0000 UTC 17 Sep to 1800 UTC 21 Oct 2007 for the best 12-km WRF EnKF configuration (red), GFS model (blue), NAM model (green), and RUC model (black).
Summary and conclusions

The quality of the 12-km WRF EnKF aloft compared with the GFS, NAM, and RUC is difficult to determine due to the limited amount of upper-air observations on the 12-km grid. However, the limited temperature, wind, and GPH observations aloft (all of which are assimilated) indicate that the 12-km WRF EnKF is inferior to all 3 operational models for both the 6-h forecast and the analysis. Thus, it appears the 12-km WRF EnKF is able to overcome deficiencies in the governing flow aloft to produce better surface analyses and 6-h forecasts over the complex terrain of the Pacific Northwest. The reason the NWS models are better aloft is probably the assimilation of significantly more data than for the WRF EnKF. In turn, further improvements to finescale EnKF nests such as those used in this study will likely come from tuning the methods developed here for bias correction and variance adjustments, and improving the synoptic-scale flow aloft and offshore by assimilating more data in the coarser domain, leading to improved atmospheric structure within the downscaled nests.

4. Summary and conclusions

This study examined surface analyses and 6-h forecasts using a 36/12-km WRF EnKF in a region of complex terrain. Surface wind, temperature, and altimeter observations were assimilated over a 5-week period using a 6-h analysis/forecast cycle, and subsequent surface analyses and forecasts were verified against an equally large, independent, unassimilated set of observations. To reduce representativeness error, observations were only considered for locations where the actual terrain height closely matched the model terrain at and near the observation site. It was found that both the analyses and 6-h forecasts on the 12-km grid were significantly better than at 36 km. Forecast improvements at finer grid spacing were likely due to a better characterization of the terrain and surface characteristics, with analyses also improved by higher-resolution covariances provided by EnKF data assimilation.

Three issues that affect how observations are assimilated were examined in 12-km EnKF experiments. The first was reducing surface wind and temperature observation variance relative to that on the 36-km grid, which is made possible by the terrain check. Reducing observation variances places more weight on assimilated observations, and in our experiments produced more accurate surface wind and temperature analyses as verified against independent observations. The second issue was insufficient surface background wind and temperature variance, which overweights model surface fields relative to observations. To remedy this problem, surface temperature and wind background variances were inflated using larger values aloft. This caused large changes in the surface analyses at independent observation locations, but resulted in only a small overall improvement in the analyses. Last, since model surface wind and temperature biases were substantial, these biases were removed prior to assimilation. Both analyses and 6-h forecasts benefited from surface bias correction. Surface wind and temperature analyses using all of the above methods were better than National Weather Service GFS, NAM, and RUC analyses, which also had biases removed. Future work will compare high-resolution EnKF surface analyses with operational, high-resolution NWS approaches: the Real-Time Mesoscale Analysis (Caldwell 2010) and MatchObsAll (Foisy 2003).

Short-term forecasts were essentially the same for all of the 12-km EnKF experiments at all levels. This insensitivity appeared to be linked to the relatively small differences aloft among the experiments. Although some differences were observed in the analysis increments aloft among the different cases, these differences generally disappeared by the 6-h forecast time. Furthermore, analyses aloft on the 12-km grid were inferior to that of the GFS, NAM, and RUC because of the use of less upper-air observation types in the UW system. Interestingly, surface EnKF analyses are of similar or better quality than the operational analyses, apparently overcoming deficiencies in the synoptic flow aloft. Since the synoptic flow aloft on the 12-km grid is highly influenced by the quality of the outer domain 36-km EnKF, significant improvements to high-resolution surface analyses and short-term forecasts may come from improving data assimilation within the 36-km EnKF. Finally, additional improvements may result by enhancements to the surface variance inflation scheme and the simple bias removal technique.

Acknowledgments. This work was supported by the CSTAR program through Award NA17RJ1232 and the U.S. Forest Service through Grant PNW07-CR-11261987-170. The third author acknowledges support by the NSF and ONR through Awards ATM-0842384 and N00014-09-1-0436, respectively. The authors also
wish to thank two anonymous reviewers for their helpful and thorough comments on an earlier version of the manuscript.

REFERENCES


——, 2002: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp.


