P1.14

ERROR GROWTH AND DATA ASSIMILATION IN A PARAMETERIZED PBL

Joshua P. Hacker* **and Chris Snyder**
The National Center for Atmospheric Research†, Boulder, CO

1. Introduction

Surface-layer in-situ observations comprise a rich, accurate, and often dense data source, but they are generally under-utilized in current operational data assimilation (DA) systems because of the complex interactions between the surface and the atmosphere aloft. Representing the structure of the PBL accurately in a mesoscale NWP model initialization could lead to improved short-range forecasts of local thermally-driven flows such as the sea breeze and slope flows. Forecasts of larger-scale phenomena could also be improved. When a mesoscale NWP model is used as a tool to generate 3-D datasets for process studies, an accurate representation of the PBL would increase confidence in the results.

The EnKF (Burgers et al. 1995; Anderson and Anderson 1999) is attractive for assimilation of surface observations to specify the state of a parameterized PBL (hereafter PPBL) because it is easy to implement and is free from many of the limitations of the variational approach. An adjoint model is unnecessary, explicit balance constraints are not applied, and anisotropic estimates of the background error covariance structure are readily available. More importantly, the background error covariances contain information about the PPBL and the appropriate vertical extent of influence of a surface-layer observation.

Accounting for uncertainty in model parameters, which are ubiquitous in PPBL schemes, can also be easily accomplished with the EnKF. While model error in a dynamical model may have unknown characteristics, it is reasonable to expect that in a physical parameterization scheme a large part of the error is related to poorly chosen (empirically or otherwise) parameters that are static. With the EnKF, stochastic distributions of PPBL parameters ("constants" either physical or not) can be estimated through correlation with observations. Each parameter is treated stochastically, and carries a distribution in the ensemble that accounts for uncertainty in its value.

The efficacy of the EnKF in a PPBL, which has unknown error-growth properties and potentially significant model error, is unknown *a priori*. Analysis and experimentation are necessary to determine whether a useful DA system is realizable. A season of mesoscale forecasts is examined to understand summer error-growth characteristics in a PPBL. Then a 1-D PPBL is run offline from its mesoscale model parent, and observations from a full model solution are assimilated via the EnKF (Anderson and Anderson 1999) in an attempt to reconstruct the PPBL state.

2. Variance-covariance structures of real-time PBL forecasts over the Southern Great Plains

To understand the potential for a data assimilation system operating on a PPBL during summer over the Southern Great Plains, we examine Weather Research and Forecast (WRF) model forecast profiles for July-August 2002. The 48-h forecasts were run daily at NCAR on a continental U.S. (CONUS) domain with horizontal grid spacing 22 km and 28 vertical levels. Each was initialized at 00 UTC (7 PM LST). The WRF was configured with the MRF PPBL scheme (Hong and Pan 1996), which includes a five-layer slab soil model (Dudhia 1996) and a surface-layer similarity scheme (Louis 1979) coupled with the PPBL nonlocal diffusion scheme. A column located near 97.5°W, 35.2°N, in Oklahoma, was chosen for evaluation, and the profiles were interpolated to a uniform vertical grid with $\Delta z = 100$ m for comparison with the results presented later.

The set of 48-h forecasts comprise an ensemble of runs representing the summer climate of the WRF for the chosen column. Ensemble spread (denoted σ_ψ where ψ is any state variable), or the square-root of variance across the ensemble as a function of forecast lead time, is also a measure of the variability of model states during the entire period as a function of local time of day. Spatial covariances or correlations computed from the ensemble demonstrate the vertical coherence of structures in the profiles, and may be used to estimate the potential impact

**Corresponding author address:* Joshua Hacker, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307. Email: hacker@ucar.edu

†The National Center for Atmospheric Research is sponsored by the National Science Foundation

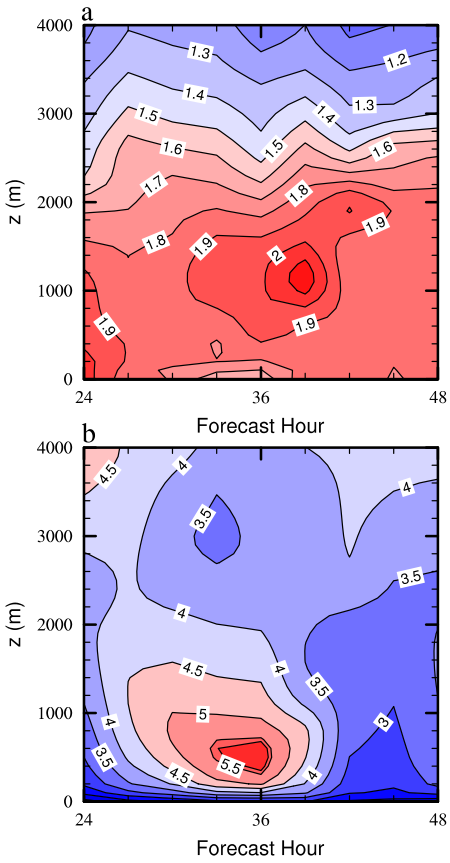

Figure 1: WRF ensemble standard deviation (spread) as a function of height AGL at forecast lead time. Contours of (a) σ_θ (0.1 K) and (b) σ_U (0.5 m s^{-1}) are shown up to 4000 m.

of an observation at a given height above ground. These calculations guide the interpretation of data assimilation experiments presented later.

Ensemble standard deviation (spread) plots at forecast lead times of 24 to 48 h (7 PM to 7 PM LST) are shown in Fig. 1 for the state variables θ and U-wind (σ_θ and σ_U respectively). The gridpoint at $z = 0$ m holds the screen-height diagnostic variables at $z = 2$ m for θ (denoted θ_s) and $z = 10$ m for U (denoted U_s). Only the last 24 h of the forecasts are shown so that the results are representative of the PPBL in the WRF instead of the Eta data-assimilation system (EDAS), from which these runs were initialized. The 24-h period also agrees with the DA experiments presented later.

Plots of σ_θ (Fig. 1a) show that below 4000 m the temperature varies more in the PPBL than it does in the free atmosphere aloft, which is synoptically inactive. The maximum at $t = 39$ h (10 AM LST), $z = 1000$ m, shows large daily variability in the growth of the PPBL with the morning onset of convection. Just before and during sunrise (30-36 h), the σ_θ is lowest near the surface. Figure 1b shows that σ_U in the PPBL demonstrates a strong diurnal dependence. During the day, when the PPBL is in a convective regime, it is coupled to the surface and the wind speeds are limited by the negative momentum flux associated with rising parameterized thermals. Conversely, the early-morning PPBL is decoupled from the surface because it is in a neutral or stable regime for most of the night, leading to a maximum in σ_U.

To improve forecast skill, a DA system needs to utilize observations to correct the regions with the fastest growing errors. The regions of high (low) spread in Fig. 1 are regions of fast (slow) error growth. Thus a DA system should be designed to correct the regions of high spread in the ensemble plots discussed above. This can be accomplished by directly observing the regions of high spread, or by observing regions that are well-correlated with the regions of high spread. Correlations with surface variables θ_s and U_s are presented here.

Temperature is correlated with $r = 0.8$ to a maximum height of approximately $z = 2000$ m in the late afternoon, with the depth decreasing to a minimum just before sunrise (Fig. 2a). This behavior corresponds to the growth and decay of the convective PPBL, but the correlation is surprisingly strong through the night when the residual layer is decoupled from the surface layer. From this we expect that assimilation of θ_s under quiet synoptic conditions can be productive nearly any time. But effective DA should be the most difficult when the spread at the top of the growing PPBL is maximized at $t = 39$ h (Fig. 1a).

The correlation of U with U_s shows a similar diurnal structure, but with greater deterioration at night (Fig. 2b). The maximum extent of correlation at $r = 0.8$ is approximately $z = 2400$ m for a short duration around $t = 45$ h (4 PM LST). The strong correlation reduces rapidly as the sun goes down and stays near the surface for most of the night. While the temperature correlation is positive all the way to $z = 4000$ m, cor(U_s,U) < 0 above $z = 3000$ m at night, and surface wind observations are not likely to help determine winds aloft.

3. Experiment

To test the efficacy of a surface-observation DA system, the MRF PPBL scheme (hereafter "the model") is extracted from the WRF model and run off-line with a perturbed-observation EnKF system. This approach facilitates an inexpensive investigation and a large statistical sample with a model that approximates the variance-covariance relationships described in the last section. The model is initialized with a random linear combination of 24-h forecasts from the set of WRF forecasts. Initialization at 24-h allows the PPBL to be a product of the WRF forecast, rather than residual PPBL structure as

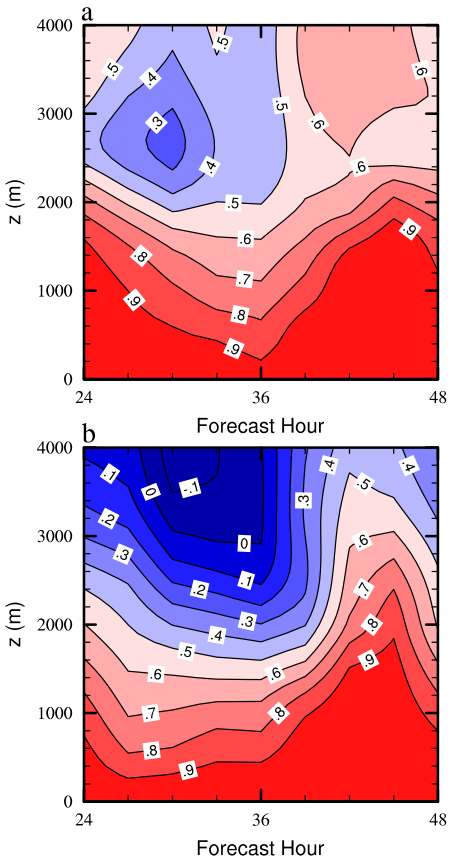

Figure 2: Correlation coefficients relating screen-height forecasts to profiles of the same variable, as a function of height AGL and forecast lead time. Contours of (a) cor(θ_s,θ) and (b) cor(U_s,U) are shown.

represented in its own cold-start initialization. The state vector is propagated forward in time by the MRF PPBL. External forcing, designed to account for 3-D dynamics present in the WRF integration, but not in the stand-alone PPBL, is provided by a time series that is consistent with the initialization. The time series includes all state variables and radiative forcing, and is applied as a tendency term over each 3-h period for which they are available.

The model is configured to run on 120 vertical grid points with $\Delta z = 100$ m and $\Delta t = 120$ s. Each run is integrated for 24 h corresponding to the 24-48 h forecast period of the WRF.

The "true" state and its evolution, and the ensemble used for the EnKF system, are also generated from the WRF forecasts. Observations are extracted from a single randomly-chosen forecast. Because the true state and its evolution are not derived from the MRF PPBL as configured off-line here, this is not a perfect-model experiment. All the results presented are for $N = 1000$ member ensembles and averaged over 20 random cases.

This model is not perfect, and its climatology will not be exactly the same as the WRF climatology. Furthermore, all of the initial conditions and forcing time series are within the extrema set by the WRF runs, but the ensemble is a more complete distribution. The climatological mean of the model will be the same initially, and the spread will be slightly smaller. As the forecast progresses, the differences will slowly grow as both the mean and the spread could be different. This behavior was confirmed by comparing Figs. 1 and 2 to similar plots generated from the ensembles of off-line model integrations. Error growth in this model is unconstrained by anything except its climatology.

4. Application of the EnKF

We refer the reader to Burgers et al. (1995) or Anderson and Anderson (1999) for a treatment of the EnKF algorithm, and examine the results of of assimilating 2-m θ and specific humidity Q (θ_s and Q_s), and 10-m wind components (U_s and V_s). Observations are extracted from the truth run. EnKF updates occur every 3 h, and the first update occurs at $t = 3$ h. The observations errors are assumed to be uncorrelated with error variances 1.0 K^2, 2.0 m^2 s^{-2}, and 1.0×10^{-9} g^2 kg^{-2}, for θ_s, (U_s, V_s), and Q_s respectively, which are one order of magnitude smaller than the ensemble spreads.

The EnKF can also be used to estimate the moisture availability (M) by including it in the state vector $\mathbf{x}$, and allowing the correlation between the parameter and the surface observations to update the distribution. The sensitivity of PPBL schemes to M has been documented in Troen and Mahrt (1986).

An easy way to evaluate the performance of the EnKF DA system is to compare error growth of the constrained and unconstrained ensembles. Constrained ensembles are generated with the same initial conditions and tendencies as were applied in the last section documenting unconstrained error growth. The results are averaged over the same 20 cases. A reduced error indicates that the observations have a positive impact on the skill of the forecast by spreading upward. In evaluating the performance of the parameter estimation, we focus on the PPBL and ignore the impact aloft.

The plots in Fig. 3 show the difference in ensemble-mean error between the unconstrained (and with estimated parameter M allowed to vary) and constrained ensembles, where negative values (shaded) show an error reduction. The results demonstrate that the observations constrain the ensemble and reduce the error through much of depth of the PPBL. For the state variables θ and U the error at screen height is steadily reduced after $t = 3$ h, as expected, because it should converge to the square root of the observation error variance. But

the error is also reduced on model layers above that and within the PPBL.

The ensemble-mean error in U is also reduced, but through a deeper layer than the θ error reduction. In particular, the variance maximum that could be seen just before sunrise in Fig. 1b is coincident with the maximum error reduction in Fig. 3b. Eliminating the pre-sunrise error maximum could have implications for larger-scale moisture and pollutant transport associated with inertial low-level jets.

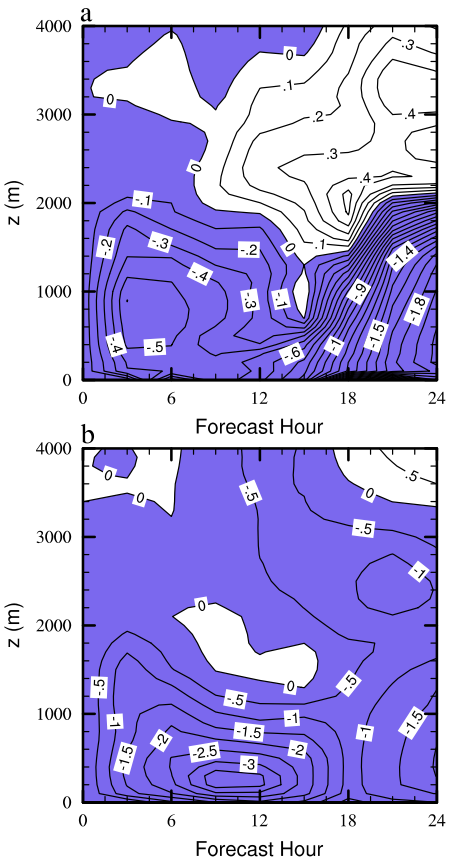

Figure 3: Ensemble-mean error difference, between the EnKF constrained ensemble and the unconstrained ensemble, as a function of height AGL at forecast lead time. The EnKF constrained ensemble includes stochastic treatment of the moisture-availability parameter M. Contours of (a) Potential temperature (θ, 0.1 K) and (b) U-wind are shown up to 4000 m are shown.

5. Summary and conclusions

This paper documents experiments on the efficacy of the EnKF approach to data assimilation and parameter estimation in a parameterized PBL. An analysis of variance-covariance structures in the PPBL of WRF model real-time forecasts over the Southern Great Plains indicates regions of expected maximum error growth, and correlation of PPBL state with near-surface observation locations suggest the potential for correcting them. An imperfect off-line model was constructed with the MRF PPBL scheme and random forcing derived from the WRF forecasts. The model allows efficient experimentation on data assimilation of synthetic surface observations and parameter estimation.

Results show that the surface observations, extracted from a random WRF forecast, are effective at constraining the state of the PPBL and reducing ensemble-mean forecast error. The regions of maximum error growth in θ, which occur during transitions and the growth or collapse of the PPBL, are almost eradicated. Similarly, the early-morning U maximum, associated with low-level jet structures, disappears when observations are introduced. The improvement in Q (not shown) is not as prominent, as was expected because of the weaker correlation between surface observations and the moisture in the PPBL.

Further error reduction results when the moisture availability parameter is treated stochastically, with the distribution updated at assimilation intervals with the rest of the model state vector. These experiments are an attempt to mitigate model error in the forecast, and the positive impacts suggest that it could be a viable approach.

REFERENCES

Anderson, J. and S. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.

Burgers, G., P. VanLeeuwen, and G. Evensen, 1995: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Dudhia, J., 1996: A multi-layer soil temperature model for MM5. *Sixth Annual PSU/NCAR Mesoscale Model Users' Workshop, Boulder CO, July 1996*, 49–51.

Hong, S.-Y. and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.

Louis, J.-F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Bound.-Layer Meteor.*, **17**, 187–202.

Troen, I. and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **37**, 129–148.