J5.4 PARAMETER ESTIMATION IN A LAND-SURFACE MODEL USING ATMOSPHERIC DATA ASSIMILATION: FINDING DISTRIBUTIONS FOR USE IN AN ENSEMBLE PREDICTION SYSTEM

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

Mariusz Pagowski
NOAA Research/Earth Systems Research Laboratory and
Colorado State University/Cooperative Institute for Research in the Atmosphere, Boulder, CO

Dorita Rostkier-Edelstein[1]
The National Center for Atmospheric Research
[1]Permanent affiliation: *Isreal Institute for Biological Research, Ness-Ziona, Israel*

March 30, 2006

1. Introduction

One of the difficulties with mesoscale prediction in the PBL is a fundamental lack of model fidelity. Errors in the mean behavior (biases) can be estimated and accounted for if a large number of cases and a good observing system are available. But probabilistic prediction using ensemble techniques relies on a model that effectively reproduces error growth. That is, a forecast from a perturbed initial state needs to diverge from an unperturbed forecast at approximately the same rate as either forecast diverges from the true PBL evolution. It is well known that mesoscale models, where the resolved-scale effect of PBL turbulence is usually parameterized, suffer from a lack of variability when compared to the real atmosphere. This results in a fundamental lack of error growth in the model, and difficulty producing a useful ensemble system for PBL forecasts.

Lack of internal variability in models has lead to a proliferation of the so-called "multi-model" ensembles for mesoscale prediction (e.g. Hou et al. 2001; Grimit and Mass 2002; Stensrud and Yussouf 2003). Multi-model ensembles can be formed by varying PBL parameterization schemes within a single dynamical modelling framework such as the Weather Research and Forecast (WRF) model or the Penn State/NCAR Mesoscale Model (MM5). They have demonstrated a skillful ensemble mean, relative to the individual members, and marginal ability to predict uncertainty. But no *a priori* reason exists to expect that such an ensemble will produce the best forecast probability distribution. It is possible that the skill improvements are the result of the different biases inherent in each parameterization scheme, and not the result of any improvement in error growth arising from internal variability.

Another approach to increasing the internal variability needed for a useful probabilistic forecast is to specify distributions of one or more of the "physical constants" in the parameterization schemes. This approach simply admits that the parameters, to which some schemes are highly sensitive, are uncertain quantities in themselves. Without assuming a non-trivial stochastic model for the parameters, the distributions can be objectively selected within an ensemble data assimilation system.

In this work, land surface parameters such as emissivity, albedo, thermal conductivity, roughness length, and moisture availability are allowed to vary within an ensemble data assimilation system. Covariance with observations modifies the parameters, which are treated as state variables, so that the model state can best fit the observations. After a few assimilation cycles, the resulting distributions are used to make an ensemble forecast. Experiments are completed with a 1D column model, containing land-surface, surface-layer, and PBL parameterization schemes available in the WRF model physics package.

The next section describes the state augmentation approach to parameter estimation. Section 3 describes the model and the experiment configuration. Section 4 presents some results, including both the effects of as-

**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

Table 1: Reference parameter values for the column experiments here.

Surface/Soil Property	Value
Land Use	USGS Category 2 (Ag.)
Emissivity (ε_{IR})	0.985
Albedo (a)	0.17
Roughness (z_0)	0.15 m
Soil Thermal Inertia (TI)	0.04 $Cal\ cm^{-1}K^{-1}s^{-0.5}$
Soil Moisture Availability (M)	Variable (Fig. 1)

similation on the parameter distributions and the effect on forecast skill. Section 5 summarized the key findings of this work.

2. State augmentation for parameter estimation

Ensemble filter data assimilation systems are designed to estimate the probability density of the atmospheric state, $\mathbf{X}$, at time t given all observations up to the time t, $\mathbf{Y}$. Mathematically, this is $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_t)$. By treating parameters, $\mathbf{x}$ as state variables, the theory supports parameter estimation. The probability density of the augmented state is $\mathbf{p}(\mathbf{Z}_t|\mathbf{Y}_t)$, where $\mathbf{Z}$ is the joint distribution of the state and the parameters, $\mathbf{Z} = (\mathbf{X}, \mathbf{x})$.

At any one time, the optimal estimate of an individual discrete representation of the augmented state, z, is given by the usual statistical analysis equation,

$$\mathsf{z}_t^{\mathsf{a}} = \mathsf{z}_t^{\mathsf{f}} + \mathsf{K}_t\left(\mathsf{y}_t^{\mathsf{o}} - \mathsf{H}\mathsf{z}_t^{\mathsf{f}}\right) \tag{1}$$

$$\mathsf{K}_t = \mathsf{P}_t^{\mathsf{f}}\mathsf{H}^T\left(\mathsf{H}\mathsf{P}_t^{\mathsf{f}}\mathsf{H}^T + \mathsf{R}_t\right)^{-1} \tag{2}$$

where superscripts a and f represent analysis and forecast (or background), respectively, and T represents transpose. The vector $\mathsf{y}_t^{\mathsf{o}}$ contains the most recent observations, and matrices $\mathsf{P}_t^{\mathsf{f}}$ and R_t are the forecast and observation error covariances, respectively. The above set of equations assume normal statistics, a perfect model to propagate z, and a linear operator relating the discrete state z to the observations y^{o}.

Many approaches to solving equation (1) have been reported in the literature, and here we use the ensemble adjustment Kalman filter in a regression framework (as reported in Anderson 2003).

3. Model and experiment

The 1D column model used in these experiments is further described in paper J5.8 in these proceedings, and with additional details in Pagowski (2004); Pagowski et al. (2005), and Appendix A in Hacker et al. (2006).

Table 2: Parameterization schemes in the 1D column model used in this experiment.

Component	Parameterization
Soil	Force-Restore (Dickenson 1988)
Soil Wetness	Simple Bucket Model (Dudhia 1996)
Surface Layer	M-O Similarity
PBL	MRF Scheme (Hong and Pan 1996)
Radiative Forcing	WRF Forecast
Geostrophic Forcing	WRF Forecast

A summary of the parameterization schemes is provided in Table 2.

Observation system simulation experiments (OSSEs) are constructed as follows:

1. Choose any number of the parameters in Table 1 for estimation.

2. Produce a “true” evolution for a particular day in the BAMEX period by running the column model with randomly perturbed parameter values.

3. Generate an initial ensemble by randomly drawing from the climatological distribution.

4. Generate initial parameter distributions by specifying a distribution class, mean, and standard deviation.

5. Assimilate half-hourly surface observations from the true evolution for six hours. Estimate the parameter distributions at the same time.

6. Repeat 5, with the initial specified distribution, unaffected by the assimilation.

It should be emphasized that these are not perfect-model experiments because the mean of the parameter distribution does not agree with the parameter values determining the evolution of the truth.

Beta distributions are chosen for the parameters in Table 1 (except land use, which is not varied). These are bounded [0,1] to agree with the expected bounds on the parameters. It is easily transformed to a normal distribution with the logistic transformation, to agree with the assumptions of normality underlying the the ensemble filter algorithm. The mean and standard deviations of the initial distribution is enforced in the transformed space. After the initial specification, the distribution is allowed to vary in the assimilation algorithm according to covariances with the observation.

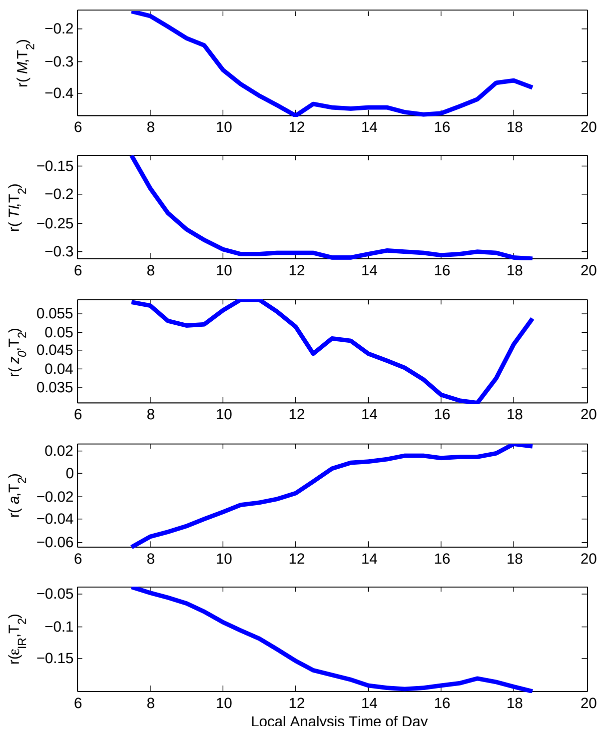

Figure 1: Cross-correlation of T_2 with the various parameters estimated, as a function of time of day.

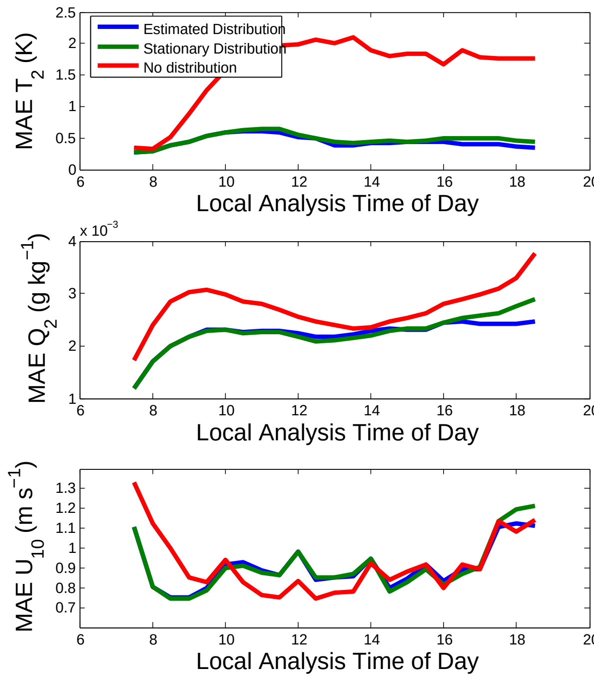

Figure 2: Mean absolute error in observation space: T_2 (top), Q_2 (middle), and U_{10} (bottom). The colors denote different experiments: when parameters are estimated (blue), when the initial distribution is not modified by the assimilation process (green), and when the parameter values are considered a constant single value (red). Statistics are compiled over 54 case days.

The extra assimilation cycle in step 6 provides a metric for quantifying the effects of assimilation on the distribution. Analysis and forecast skill with the distribution initially specified is the null case of varying the parameter without any information from the observations in the assimilation cycle. Comparison of analysis and forecast skill against this quantifies the impact of the observations.

4. Results

Results are presented for a 12-h assimilation period beginning at 1200UTC (0700 local time). A total of 54 days in the BAMEX period are used to aggregate statistics. First, the cross-correlation of each of the estimated parameters with the ensemble of model values in observation space (e.g. diagnostic screen and anemometer height values, in this case T_2) are shown in Fig. 1. This shows that the moisture availability (M) and the thermal inertia (TI) are reasonably correlated with T_2, and suggests that assimilating T_2 will affect those parameters more than the others. The correlations also quantify the sensitivity of T_2 to the parameters. The high correlation between M and T_2 is generally expected. The similar character of TI may result from a similar dependence on soil moisture.

The ability of the model to fit the assimilated observations is shown in Fig. 2. Mean absolute error (MAE) is computed over all 54 assimilation periods. The results show a generally better fit when a distribution of parameters is used (blue,green) rather than a single fixed value (red), with a small temporary reduction in skill for U_{10} during the middle of the day. The difference between the skill when the distribution is estimated (blue) and when it is specified and stationary (green) are rather small according to this metric. A slight divergence near the end of the time period (1800) suggests that the assimilation cycling may need to continue longer to see real differences.

Some dependence on time of day is evident, and other times of day need to be examined before conclusions can be drawn. These results only suggest that this is a promising approach.

An example of the effect of the assimilation on the distribution of TI is shown in Fig. 3. We choose TI rather than M, despite its lower distribution, because the bucket model also modifies M dynamically, making it more difficult to interpret. Figure 3 shows that the mean peaks in the middle of the day to better fit the observations, suggesting that the values need to increase when the temper-

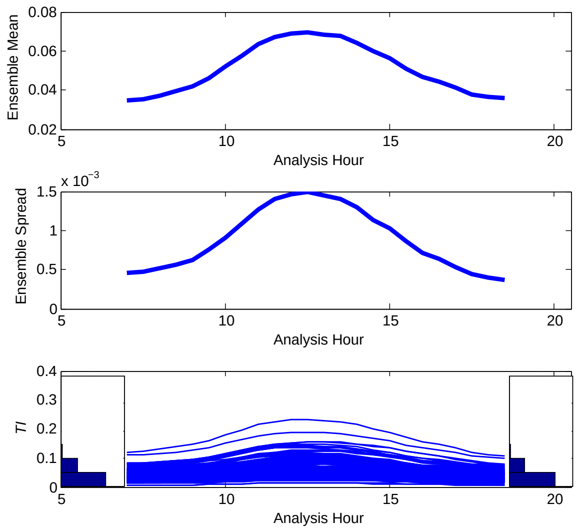

Figure 3: Example of the evolving thermal inertia mean (top), standard deviation (middle), and ensemble (bottom). Initial and final histograms are also shown inset bottom.

atures are high. The ensemble spread also peaks, suggesting a lack of confidence in that estimate. In this case, the values decrease later in the day, suggesting greater confidence in lower values. This type of evolution is not representative of typical behavior, which can take many other forms.

5. Summary and conclusions

Parameter estimation experiments with an ensemble filter are run with a column model containing WRF land-surface, surface-layer, and PBL physical parameterization schemes. Shelter and anemometer-height observations are assimilated half-hourly. The overall goal is to estimate distributions of parameters that will improve error growth properties in the PBL and improve skill and estimates of uncertainty in both analyses and forecasts.

A subset of the experiments were presented here, and general conclusions cannot be drawn, but the results appear promising. Parameters such as moisture availability and thermal inertia are correlated with the model state at observation locations, suggesting that the parameters could be estimated. The model fits the observations better in the assimilation when the parameters are treated as a distribution rather than a fixed value. A longer assimilation period may be necessary to truly understand the effects of the estimation procedure compared to using a fixed distribution. Assimilation over different parts of the diurnal cycle will elucidate the effects of parameter estimation in the appropriate diurnal regime.

The talk will address these issues, and additional issues not discussed here. These include the effects on short-range forecast skill and internal error growth in the model, an examination of the behavior near transition times, and comments on how to improve ensemble forecast systems. Both observation-space verification, and a verification of the profiles, will be shown.

REFERENCES

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.

Dickenson, R., 1988: The force-restore model for surface temperature and its generalization. *J. Climate*, **1**, 1086–1097.

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.

Grimit, E. P. and C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. and Forecast.*, **17**, 192–205.

Hacker, J. P., J. L. Anderson, and M. Pagowski, 2006: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations, submitted.

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.

Hou, D., E. Kalnay, and K. Drogemeier, 2001: Objective verification of the SAMEX '98 ensemble experiments. *Mon. Wea. Rev.*, **129**, 73–91.

Pagowski, M., 2004: Some comments on PBL parameterizations in WRF. *Joint WRF/MM5 Users' Workshop, Boulder CO, July 2004*, 43–46, available at http://www.mmm.ucar.edu/mm5/workshop/workshop-papers_ws04.html.

Pagowski, M., J. Hacker, and J.-W. Bao, 2005: Behavior of WRF PBL schemes and land-surface models in 1D simulations during BAMEX. *Joint WRF/MM5 Users' Workshop, Boulder CO, July 2005*, available at http://www.mmm.ucar.edu/wrf/users/workshops/WS2005/WorkshopPapers.htm.

Stensrud, D. J. and N. Yussouf, 2003: Short-range ensemble predictions of 2-m temperature and dewpoint temperature over new england. *Mon. Wea. Rev.*, **131**, 2510–2524.