

PBL STATE ESTIMATION WITH SURFACE OBSERVATIONS, A COLUMN MODEL, AND AN ENSEMBLE FILTER

Joshua P. Hacker* and Dorita Rostkier-Edelstein¹

The National Center for Atmospheric Research[†], Boulder, CO

¹Permanent affiliation: *Israel Institute for Biological Research, Ness-Ziona, Israel*

Revised manuscript submitted to *Monthly Weather Review*

October 23, 2006

*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

Abstract

Following recent results showing the potential for using surface observations of temperature, water vapor mixing ratio, and winds to determine planetary boundary layer (PBL) profiles, this paper reports on experiments with real observations. A 1D column model with soil, surface-layer, and PBL parameterization schemes that are the same as in the Weather Research and Forecast (WRF) model is used to estimate PBL profiles with an ensemble filter. Surface observations over the southern great plains are assimilated during the spring and early summer period of 2003. To strictly quantify the utility of the observations for determining PBL profiles in the ensemble filter framework, only climatological information is provided for initialization and forcing. The analysis skill, measured against rawinsondes for an independent verification, is compared against climatology to quantify the influence of the observations. Sensitivity to changing parameterization schemes, and to prescribed values of observation error variance, is examined. Temporal propagation of skillful analyses is also assessed, separating the effects of good prior state estimates from the impact of assimilation at night when covariance is weak.

Results show that accurate profiles of temperature, mixing ratio, and winds are estimated with the column model and ensemble filter assimilating only surface observations. Results are largely insensitive to choice of parameterization scheme and specified observation error variance. The effects of using different parameterization schemes within the column model depend on whether assimilation is included, showing the importance of evaluating models within assimilation systems. At night, skillful estimates are possible because the influence of the observations from the previous day is temporally propagated, and atmospheric dynamics in the residual layer operate on

slow time scales. It is expected that these profiles will have applications for now-casting and secondary models (e.g. plume dispersion models) that rely on accurate specification of PBL structure.

1. Introduction

Accurate planetary boundary layer (PBL) analyses and very-short-range forecasts (nowcasts) can help with several practical forecasting and secondary-model applications. Convective weather nowcasting is perhaps the critical example, because convective initiation and forecasted precipitation has been shown to be sensitive to PBL structure (Crook 1996; McCaul and Cohen 2002; Martin and Xue 2006). Air-quality analysis and plume dispersion studies can also benefit from improved PBL analyses of stability and mixing depth. Yet deficiencies in NWP models and ineffective use of near-surface observations persist, leading to PBL analyses of dubious quality.

Several difficulties prevent the optimal use of surface (shelter and anemometer-height) observations in modern data assimilation systems. First, transient coupling with the Earth's surface and the free atmosphere produce intermittent, anisotropic, and nonstationary correlations of the observations with the model background state. Second, the error growth estimated with models providing the background state is largely unknown, highly variable, and likely not well-represented in current mesoscale models. Third, dynamic balances often exploited in large-scale data assimilation are inappropriate for PBL observations of temperature (T), component winds (U , V), and water vapor mixing ratio (Q).

Ensemble filters represent, theoretically, a method for overcoming many of these difficulties, but this approach to PBL assimilation is in its developing stages. The ensemble provides a means of estimating flow-dependent background error statistics, including the full covariance, and for-

mally handling model deficiencies. Adjoint methods are not necessary and construction of ensemble data assimilation systems is much simpler than the more complex variational schemes. Hacker and Snyder (2005) showed in perfect-model observation-system simulation experiments (OSSEs) that ensemble assimilation could prove fruitful for specifying overlying PBL profiles from surface observations. They also showed that land-surface parameters could be effectively estimated in the ensemble-filtering framework.

Here we take the step from observation-system simulation experiments (OSSEs) to observation-system experiments (OSEs) by assimilating real observations. A column over the Atmospheric Radiation Measurement (ARM) program Southern Great Plains Central Facility near Lamont, OK, is selected for analysis because of the robust data for both assimilation and verification. The lack of complex topography in that region reduces the likelihood that the climatology will be determined by mesoscale features, and the climatological distribution should be random above the PBL.

A column model containing soil, surface-layer, and PBL parameterization schemes that are the same as those in the Weather Research and Forecast (WRF) mesoscale modeling system propagates the column state between assimilation cycles. This results in a response to observations that is similar to what can be expected in the WRF.

Experiments are constructed to assess the vertical influence of surface observations on the model state, in the absence of additional information. All ensembles are initialized and forced with distributions drawn from climatology. Surface observations are assimilated and the resulting estimated profiles are verified against rawinsonde observations. The utility of surface observations in specifying the state of the PBL is assessed by comparing the skill of the estimated profiles with the skill of profiles generated by running an ensemble without any assimilation to generate the model climatology. Future work will combine background information from the most-recent WRF forecast and the assimilation of surface observations in the column model, providing a more

optimal and practically useful estimate.

The next section describes the model, the observation and verification data, and the assimilation system. Section 3 presents baseline results, and looks at sensitivity to the specification of observation error variance and parameterization schemes. Section 4 investigates the roles of vertical correlations and ensemble variance, and section 5 summarizes the key findings of this work.

2. Experiment description

a. *The 1D column model*

A column model containing a suite of physical parameterization schemes is useful for the experiments described here. We are interested in only vertical structures and relationships within and near the PBL, and a column model allows experimentation at a fraction of the cost associated with a 3D mesoscale model. Large ensembles are feasible, enabling convergence of results and experimentation with sensitivity to ensemble size.

The model can be thought of as a simpler cousin to the Weather Research and Forecast (WRF) mesoscale modeling system. The WRF modeling system includes two different cores for computing the resolved dynamics (Janjić 2003; Skamarock et al. 2005), and both have access to similar physical parameterization schemes. Here, simple dynamics are coupled to the same PBL, land-surface, and soil parameterization schemes as the full WRF modeling system. The dynamics are momentum, thermodynamic, and moisture conservation equations:

$$\frac{\partial U}{\partial t} = f(V - V_g) - \frac{\partial}{\partial z} \langle u'w' \rangle, \quad (1)$$

$$\frac{\partial V}{\partial t} = -f(U - U_g) - \frac{\partial}{\partial z} \langle v'w' \rangle, \quad (2)$$

$$\frac{\partial \theta}{\partial t} = -\frac{\partial}{\partial z} \langle w'\theta' \rangle, \quad (3)$$

$$\frac{\partial Q}{\partial t} = -\frac{\partial}{\partial z} \langle w'q' \rangle. \quad (4)$$

The usual Reynolds averaging has been applied to break the wind components (u, v) into mean and turbulent components, e.g. $u = U + u'$. Angle brackets denote an average over subgrid-scale eddies. The winds are relaxed to prescribed geostrophic wind components (U_g, V_g) , and the relaxation dominates when the turbulent fluxes are weak or zero. Net radiation at the land surface is also prescribed to force the system. A radiation scheme is not included in the column as used here, introducing a model deficiency in the radiative interactions between the atmosphere and the surface.

The model employs the same suite of physical parameterizations as the WRF for subgrid processes associated with the land surface, surface layer, and PBL. For the base case we chose the Mellor-Yamada-Janjić (MYJ) PBL scheme (Janjić 2001) as implemented in the WRF, which is similar to the implementation in the NCEP Eta model. The parameterization uses the prognostic equation for turbulent kinetic energy (TKE) with the assumption of down-gradient diffusion and pressure covariance, and diagnostic equations for potential temperature and moisture convergence. Surface layer parameterization follows Monin-Obukhov similarity theory, extended by Beljaars and Holtstg (1991) to the free-convection regime. The momentum roughness scale is constant; temperature and moisture roughness scales are assumed the same and calculated using the Zil-

itinkevich (1995) formulas. Land-surface parameterization is achieved with the Noah land-surface model (LSM) (Ek et al. 2003), which is a 4-layer soil temperature and moisture model with canopy moisture and snow-cover parameterizations. It provides sensible and latent heat fluxes to the atmospheric model taking into account atmospheric state, net radiation at the surface, and land use characteristics such as vegetation type and soil texture.

The vertical grid is defined as 33 vertically stretched atmospheric levels, with the first layer extending to approximately 40 m above the surface and a top at approximately 4800 m. The time step is 10 s, and the model solution is insensitive to the time step length when it is this short. Further details of the column model are given in Pagowski (2004), Pagowski et al. (2005) and in Appendix A of Hacker et al. (2006). All of the usual atmospheric state variables (U , V , θ , Q , and diagnostic pressure P), and also the soil temperature profile, are included in the assimilation state vector and thus directly modified by the surface observations.

Initial conditions, large-scale forcing, and surface radiation are imposed by randomly sampling two forecasts from a (warm) season of WRF real-time¹ forecasts at a column located over Oklahoma, then combining them with a uniform random coefficient between zero and one ($\mathcal{U}[0, 1]$). The same weights and forecasts are used for both the initial conditions and the forcing, so that the forcing time series and initial conditions are consistent. WRF 36-h forecasts from the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX) observation period spanning 03 May through 14 July 2003 form the sample. Forecasts were launched at 00 UTC every day, on a $\Delta x = 4$ km grid. More details on the sampling approach are available in Hacker and Snyder (2005). Although the distribution of large-scale forcing is narrowed by the linear combination, the local effects on the column are isolated and ensembles larger than the WRF sample are available.

¹Forecasts were generated at NCAR with the Advanced-Research version of the WRF (ARW). Those used the Yonsei University (YSU) PBL scheme, which is a modification of the MRF scheme (Troen and Mahrt 1986; Hong and Pan 1996).

b. Assimilation and verification data

The Southern Great Plains (SGP) Central Facility of the Atmospheric Radiation Measurement (ARM) program is well-instrumented, providing observations for both assimilation and verification of the column analyses. Observations for assimilation are 30-minute averages of temperature and water vapor mixing ratio at $z = 2$ m (T_2, Q_2), and winds at $z = 10$ m (U_{10}, V_{10}). The reports include quality-control flags and estimates of uncertainty. Rawinsonde observations valid every 6 h are used for verification.

The location and experiment time period are not directly relevant to the results of this work, but the SGP ARM site is attractive because it is heavily instrumented and observations undergo extensive quality control. The BAMEX period is chosen because of the availability of the high-resolution WRF forecasts to provide initial conditions and forcing. This combination facilitates extensive research, beyond what is presented here.

Mean absolute error (MAE) in the ensemble mean is computed with reference to rawinsondes that observe the atmosphere overlying the assimilated surface observations. A total of 72 days of the BAMEX period produced approximately 45 useful soundings at any given verification time of day, after accounting for stringent quality control. Rawinsondes are launched from the same observing facility as the assimilated observations, but drift away with the mean winds. Horizontal drift, averaged over the verification data set, is nearly linear with height to approximately 3.5 km at $z = 2$ km AGL (not shown). The effect of drift on the quantitative verification is difficult to assess, but it is not central to understanding the impact of the surface observations.

The relevant comparison for verification is an ensemble of column model integrations over identical time periods, and with identical initial conditions and forcing. We refer to these simulations as “climatological” because the initialization and forcing are drawn from the sample clima-

tology and are not conditioned on a particular flow scenario, resulting in a sample with no skill. For each run during the experiment period, the same random distribution is drawn for initialization and forcing, and the same number of ensemble members goes into the computation of the ensemble mean. Thus the comparisons are fair in terms of initialization, forcing, and sampling. We do not consider the distribution of likely states conditioned on the synoptic regime, cloudiness, soil moisture, or anything else besides the local time of day. The column model also lacks 3D dynamics, including the lack of advective tendencies, and the only 3D information comes from the climatological distribution of geostrophic winds.

Assimilating surface observations introduces information about the actual state of the atmosphere, and the reduction in error quantifies the impact of those observations in the absence of any additional information. At some height above the surface, the error in the assimilation experiments approaches the error in the climatological simulations, showing the maximum vertical extent of the assimilation impact for this model configuration. When verifying estimated profiles in section 3, we show both error curves for reference.

c. Specific ensemble filter implementation

Ensemble filters are by now well-documented in the literature (e.g. Evensen 1994; Burgers et al. 1998; Houtekamer and Mitchell 1998), and here the equations are not presented. Experiments are conducted within the Data Assimilation Research Testbed (DART), developed at NCAR, and use the serial least-squares implementation of the Ensemble Adjustment Kalman Filter (EAKF) as documented in Anderson (2001, 2003). Screen-height temperature and water vapor mixing ratio (T_2 and Q_2) and anemometer-height wind component (U_{10} and V_{10}) observations are assimilated hourly, with ensembles of $N = 100$ members. The assimilation cycling is restarted daily at 1300

UTC (1-h after reinitialization of the ensemble), rather than continuous cycling over the entire experiment period, to fit within constraints imposed by the sample of WRF forcing.

Base-case observation error variances are specified the same as in Hacker and Snyder (2005), which agree roughly with values estimated in Crook (1996). These are 1.0 K^2 , $1.0 \times 10^{-6} \text{ kg}^2 \text{ kg}^{-2}$, and $2.0 \text{ m}^2 \text{ s}^{-2}$, for T_2 , Q_2 , and (U_{10}, V_{10}) , respectively. No further tuning of these values is performed, but in the next section we report results from an experiment using much smaller values, estimated from the observation data itself.

Vertical covariance localization is accomplished with an element-wise multiplication of the fifth-order piecewise rational function (Gaspari and Cohn 1999, eq. 4.10) and the background error covariance estimates. The half-width is chosen to be half of the domain so that the model lid is not affected by the assimilation. Hacker et al. (2006) showed that this has little effect on sampling error for ensembles as large as $N = 100$ members, but we impose it to prevent analysis increments at upper levels where the model has no skill.

Computationally, the assimilation algorithm scales with the number of observations per assimilation cycle (four), and the majority of the cost is in the ensemble of model integrations. For typical atmospheric data assimilation problems, the cost can be roughly equivalent to four-dimensional variational assimilation (4DVAR), where the cost is in a 4D minimization algorithm. Ensemble filters provide the added benefit of an ensemble of analyses that can be used for probabilistic prediction, and statistics useful for understanding linear relationships in the model. In these experiments, ensembles with $J = 100$ members are used, and 24 hours of integration/assimilation takes a few minutes on a desktop computer.

3. Verification of estimated profiles

In this section, we present a basic verification of the analyzed profiles, and examine the sensitivity of the results to changes in parameterization schemes and specified observation error variance (σ_o^2). The relatively small sample of approximately 45 verification times opens the question of significance. Statistical significance tests (Z-test) performed at each verification time separately show that although some of the small differences between mean results are not significant, all of the notable differences are significant to at least the 90% level. We restrict our discussion to those differences.

a. *Base case verification*

In this subsection we present a straightforward verification of the basic model and assimilation configuration. Observations are assimilated beginning at 1300 UTC [0800 local time (LT)] on each day of the experiment period. Assimilation continues with hourly updates for 24 h. Rawinsondes are available for PBL verification at 0000, 0600, 1200, and 1800 UTC, corresponding to assimilation cycling over periods of 12, 18, 24, and 6 h prior to verification.

Observation-space diagnostics are first presented to quantify first-order ensemble filter performance. Then state-space verification is presented, showing the depth of influence of the surface observations in the cycling model/assimilation system. All statistics are aggregated over the experiment period. For a quadratic metric, we choose the mean-absolute error (MAE) rather than the root-mean-square error (RMSE), except when comparing spread to error, because the former is more resistant to outliers. For a state variable ψ , an observation ψ^o , J ensemble members, and I verification times, the MAE is:

$$MAE = \frac{1}{I} \sum_{i=1}^I \left| \left(\frac{1}{J} \sum_{j=1}^J \psi_j \right) - \psi^o \right|. \quad (5)$$

The relationship between the spread (standard deviation squared) of the ensemble and the ensemble-mean error is the canonical first-order assessment of filter performance. To avoid the need for normalization, the relevant quantity is the mean-squared error (MSE) of the ensemble mean, similar to equation (5) except the value in the absolute value operator is squared. When computed over many cases, the ratio of spread to error should be near one, indicating that the magnitudes of spread and error are similar on average. Sampling error can lead to spread that is somewhat less than error. Results for the near-surface assimilation of T_2 , Q_2 , and U_{10} are shown as a function of assimilation cycle in Fig. 1. A diurnal variation is evident with ratios near one during the day, and error exceeding spread during the night. Decreasing spread/error ratios for T_2 are caused primarily by error increases with nearly constant spread during the night, while for Q_2 the spread also decreases. Ratios for U_{10} remain higher because the geostrophic winds produce variability in the ensemble. Although not shown here, these effects are also observed when the ensemble is initialized at either 0600 UTC or 1800 UTC, rather than 1200 UTC, confirming the diurnal characteristic. It is expected that further tuning of σ_o^2 could produce better ratios at night. We avoid manual tuning of σ_o^2 to have error estimates that reflect typical values, and prefer giving an estimate of the effectiveness of surface observations in a sub-optimal setting. Experiments with the present values of σ_o^2 comprise the base case, and in section 3c below we examine the effects of fitting the observations of T_2 and Q_2 more closely.

[Figure 1 about here.]

The MAE of estimated profiles are shown in Figs. 2, 3, and 4, respectively for θ , Q , and the U -wind component. The results show a time-dependent positive impact of assimilating the sur-

face observations, quantified by the difference between the error in the climatological simulations (dashed curves) and the assimilation experiments (solid curves). In each figure, the results are presented in order of increasing time from the experiment initialization at 1200 UTC (0700 LT). Thus the top two panels are at mid-day (1300 LT) and early evening (1900 LT); the bottom two panels are near midnight (0100 LT) and early morning (0700 LT). The early-evening verification is usually prior to the development of a stable surface layer, and the early-morning verification is prior to the onset of convective PBL growth.

In general the depth of influence of the surface observations increases slightly during the afternoon hours (from panels a to b). This results from the large ensemble covariance between the state in the surface layer, which is observed, and the state in the well-mixed PBL. The depth of the PBL also grows slightly after the 1300 LT verification, as observed qualitatively.

At night, covariance between the near-surface states and the profile states slowly decreases because both the correlation and the profile variance decrease. The lack of 3D dynamics results in little change to the profiles of θ and Q when mixing is absent. Nothing generates spread in the ensemble, and as long as a non-negligible correlation between the surface and the profile exists the spread reduces each time an observation is assimilated. Thus the variance in the column tends to decrease during nighttime hours. The same effect is not observed in U because the background climatological geostrophic wind tendencies are still active. At the same time, lower-boundary forcing causes a decorrelation between the near-surface state and the atmosphere aloft. The net effect is a covariance reduction over time, leading to little effect of the assimilation on the profiles. Both the decoupling and the assimilation itself act to diminish the effect of later observations on the atmosphere above the surface layer.

We note here that assigning an unambiguous physical interpretation to the nighttime behavior is difficult. Decoupling of the surface with the overlying atmosphere in cases of weak wind and strong

stability is nearly complete, except for the residual layer that was created based on conditions a few hours prior, and the radiative interaction between the surface and the overlying column. The latter is not present in this model, resulting in weaker correlations than may otherwise be present. Further, shelter and anemometer-level variables are diagnosed assuming that the surface layer extends to the lowest model layer and that Monin-Obukhov similarity theory holds at night, neither of which may be the case. We focus our analysis primarily on correlation and ensemble spread, which can be measured directly, without making strong statements about the physics in the model.

Skill in the analyses, as quantified by improvements over the climatological simulations, depends on the relative time scales of the decorrelation and variance reduction. The error in the assimilation experiments is much lower at the end of the day, and the skill is retained. This behavior can be seen in Figs. 2c,d and 3c,d. In regions aloft where the error when assimilating exceeds the error in the climatological simulations (e.g. Fig. 3d), the differences are not statistically significant. Those differences can be considered first-order error bars, which are narrower in the PBL where affected by the assimilation and lower boundary forcing.

Despite the relative lack of nocturnal dynamics aloft in θ and Q , some tendency of the error toward climatological error values can be observed in Figs. 2c,d and 3c,d. This may be caused by some weak dynamics in the residual layer, or a weak effect of small covariance values interacting with the deficient model to produce poor covariance estimates. As discussed in section 4, this slight error increase does not destroy the generally positive effects overnight.

In U , the decoupling leads to an increase in variance as the columns are forced by random climatological geostrophic winds. As with θ and Q the correlations decrease, but in this case result in a more rapid loss of skill, which approaches the values of the climatological simulations near sunrise (Fig. 4c, d). The relative contributions of variance, decorrelation, and the lack of dynamics

is further investigated in section 4.

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

b. Sensitivity to changes in the parameterization schemes

As discussed in section 2, the 1D model contains the full suite of PBL and land-surface parameterization schemes available in the WRF modeling system. The base case experiments presented above use the MYJ PBL scheme. Different schemes will result in different estimates of error covariance, and also different prior state estimates. Here we compare the base results with results from experiments using the Yonsei University (YSU) scheme (Skamarock et al. 2005), which is an update of the well-used MRF scheme (Hong and Pan 1996) to include different handling of entrainment, countergradient fluxes, and PBL height estimation for the convective PBL. The surface-layer formulations are both based on Monin-Obukhov similarity, but are slightly different in their handling of surface roughness and also the classification of stability regimes.

The spread/error ratios in observation space (T_2, Q_2, U_{10}, V_{10}) are similar when either the MYJ or YSU scheme is used, but the underlying values of error and spread do differ. Figure 5 shows the errors in observation space for the climatological case, which does not include data assimilation. Error differences are larger during the night, with the advantage of each scheme dependent upon the variable. MYJ appears more skillful in T_2 , YSU appears more skillful in Q_2 , and neither holds a clear advantage in U_{10} . This behavior is limited to the surface layer. Climatological error profiles that result when using the YSU scheme are nearly the same as the dashed curves in Figs. 2-3.

[Figure 5 about here.]

Assimilating observations changes the nature of the error, and underscores the importance of model evaluation within a relevant data assimilation system. Figure 6 shows the posterior error during the assimilation experiments, and can be compared to Fig. 5. The effect of assimilation is obvious in the error magnitude, but the relative error when using the different PBL schemes also changes significantly. In T_2 , the difference between the MYJ and YSU schemes is smaller in both relative and absolute terms. Results are similar for Q_2 , with intermittent differences appearing. Just after sunset the MYJ scheme shows a short-lived advantage over the YSU scheme, opposite from the climatological case. In U_{10} , the differences are amplified, and the YSU scheme shows a clear advantage during the night when used in conjunction with the assimilation of surface observations.

[Figure 6 about here.]

The MAE in the T and Q profiles is similar for the MYJ and YSU schemes, and the comparison is omitted for brevity, but the U_{10} improvements seen in Fig. 6c extend to the profiles of the U -wind at night (Fig. 7). Given that the climatological errors in the profiles are similar, this error reduction results from improved interaction with the data assimilation system. One possible cause is that the forward operator, essentially the diagnosis of U_{10} from the profile in the YSU scheme, is superior. This could result from better specification of coefficients in the surface-layer similarity relationships at the vertical grid spacing in this model implementation. Other possibilities are more directly related to dynamics. First, it is possible that the YSU scheme displays slower error growth in the surface layer during the one-hour period between observations. This effect would not show up in the climatological statistics, where the night begins 12 hours after initialization and the error may be saturated. Second, well-mixed profiles may persist longer into the night in the YSU scheme, leading to correlations that persist longer. Finally, bounds on internal variability may be

different for the MYJ and YSU schemes, leading to different error growth characteristics depending on error magnitude. Such an analysis is beyond the scope of this work, but does not diminish the main point that model evaluation can be strongly dependent upon the data assimilation scheme. Despite some differences, these results in general suggest that the effect of surface observations is not strongly dependent on the choice of parameterization scheme.

[Figure 7 about here.]

c. Sensitivity to specified observation error variance

The closeness of fit to assimilated observations is partially controlled by the specified σ_o^2 . Smaller values of σ_o^2 usually improve analysis skill at the observing locations, but can lead to over-fitting and poorer verification where observations are not assimilated. This would be manifested by errors that are closer to the climatological errors or in some cases worse, and also poor performance in the states prior to assimilation. Here we show evidence of weak over-fitting when observation σ_o^2 is specified much smaller, but the sensitivity is not as strong as might be expected.

ARM reports include the variance over each 30-minute averaging interval, and may be used to estimate σ_o^2 . The observation error variance typically includes contributions from instrument sampling rates, the disparity between spatio-temporal truncation in the model and instrument, and mean and random components of instrument error. These quantities are difficult to estimate, but the use of a 1D column model potentially simplifies the problem. Horizontal truncation caused by discretization is not a factor in the internal model dynamics. It is present in the forcing, but the effects should be small because the forcing is geostrophic. The time step of 10 s is small relative to the time scale of the averaged observations, and much closer to the sampling rate of the instrument. The primary remaining source of error is then the instrument representativeness,

which is quantified by the variance over the 30-minute averaging period. The experiment-period average of those variances gives 0.08 K^2 , $1.1 \text{ m}^2 \text{ s}^{-2}$, and $7.7 \times 10^{-8} \text{ kg}^2 \text{ kg}^{-2}$, for T_2 , (U_{10}, V_{10}) , and Q_2 respectively. Values for T_2 and Q_2 are small compared to what are typically assigned in data assimilation experiments (c.f. Hacker and Snyder 2005, and references therein), and what we use for our base case above, but values for winds are of the same order. These values do not include bias, or random errors attributable to the instrument itself, which can increase the total observation error variance by a factor of two (in winds) to an order of magnitude (in T and Q) (c.f. ARM 2006).

In observation space, smaller values of σ_o^2 lead to reductions in both analysis error and spread, and some effect on the relationship between them. A comparison of spread/error ratios using the small values of σ_o^2 to the base case is summarized in Fig. 8. The changes observed in Fig. 8 can be explained as follows. During the day, error in U_{10} is reduced by a slightly larger factor than spread, and during the night the spread is reduced but the error is affected little. The result is an amplified diurnal variation in the spread/error ratio (Fig. 8b), although σ_o^2 in U_{10} is reduced by only a factor of two. Error variance in T_2 and Q_2 are reduced by two and three orders of magnitude respectively, and the weak effects on U_{10} emphasize the weak cross-variable correlations between winds and other variables.

Results for T_2 and Q_2 are more notable, with further amplification in the diurnal cycle. The error in T_2 is always smaller than the base case, and much smaller at night. The error in Q_2 is also smaller than the spread in the early morning hours (0500-0700 LT), but the overall cycle is more in phase with the base case. The error (not shown) is reduced by factors of approximately three (seven) during the day (night) for both T_2 and Q_2 , fitting the observations much closer.

Although the spread/error ratios behave differently for small values of σ_o^2 , the ratios appear only moderately sensitive. We conclude that the first-order probabilistic aspects of filter performance in observation space are reasonable in either case. Looking at similar diagnostics for the error/spread

ratio in the ensemble prior to assimilation (not shown) leads to the same conclusion. Using either set of σ_o^2 values, filter divergence does not appear to present a problem within the 24-h period of each run.

[Figure 8 about here.]

Profile estimates suggest a weak over-fitting tendency when using small σ_o^2 for surface observations. In general, the skill improves during the day, but results at night are variable-dependent. Figure 9 shows MAE of θ at 1300 LT and 0100 LT. In the well-mixed daytime PBL (panel a), a closer fit to the surface observations leads to less error in the lowest 500 m, but slightly more error in the upper part of the PBL and the capping inversion. Small values of σ_o^2 work synergistically with the high correlation between the profile and the surface observations, resulting in less ensemble spread in the profile by the end of the day, as verified by comparison to the base case (not shown). Consequently, the analysis increments are smaller than in the base case during the transition to night, and the state does not accommodate the changing stability in the surface layer as well (Fig. 9b). The adverse affect suggests that the small errors in T_2 , documented by the large spread/error ratio in Fig. 8 are in fact too small to facilitate the best assimilation.

[Figure 9 about here.]

Profiles of Q and the U -wind do not appear to be adversely affected, and smaller values of σ_o^2 result in less error (Figs. 10–11). During both day and night, the estimate is improved, suggesting that over-fitting is not a problem in these variables.

[Figure 10 about here.]

[Figure 11 about here.]

Overall, the profile estimates with small values of σ_o^2 suggest a weak over-fitting is only observable in θ . Combined with the spread/error ratios presented above, ensemble assimilation of surface observations with an imperfect 1D model appears to be only moderately sensitive to the specification of σ_o^2 at or below the base case levels.

4. Background state estimates and observation influence

Section 3 demonstrates estimation of skillful profiles with a simple column model and ensemble-filter data assimilation, suggesting the potential for use of surface observations in a 3D ensemble data assimilation system. One characteristic of cycling data assimilation systems is the temporal propagation of observational information. In this section we examine the relative roles of skillful prior state estimates and observations by looking at ensemble-mean analysis increments and the prior ensemble spread. The results elucidate differences between the vector U -wind component and scalars (θ , Q) in terms of temporal propagation of the influence of observations.

Ensemble-mean absolute analysis increments partially summarize the effect of the assimilation. Increments are computed $\left| \left\langle \psi^a(z) \right\rangle - \left\langle \psi^b(z) \right\rangle \right|$ for any variable ψ , where the superscript a denotes the analysis (posterior), the superscript b denotes the background (prior), the brackets denote the ensemble mean. The experiment-mean of this quantity is plotted, corresponding to the mean over I in equation (5).

Time-height cross sections of absolute increments show different behavior in scalar quantities than in vector wind components (Fig. 12), agreeing with many of the results presented in section 3. Prior to assimilation at 0800 LT, the ensemble distribution is climatological, and the first assimilation time produces large increments in all variables. In θ (panel a), assimilation leads to very little additional change to the profile during the day, a brief period of deep effects during the tran-

sition to night, and then large increments within a shallow layer during the night. Increments in Q (panel b) show greater magnitude throughout the afternoon, with the transition to the nighttime regime occurring slightly earlier than for θ . Increments are small at night in the whole column. Increments in U (panel c) are large throughout the entire diurnal period, with a local maximum during a transition to night that occurs between the transitions in Q and θ . Nighttime increments are possible in the column because the relaxation to climatological geostrophic winds maintains variance in the column, and the correlation between 10-m winds and the overlying column winds remains stronger at night than the correlation between 2-m observed scalars and the corresponding overlying column.

[Figure 12 about here.]

Interplay between the model dynamics (and forcing) and the assimilation can be evaluated with the temporal evolution of ensemble spread in the profile. During periods of increasing spread, internal error growth (the divergence of model trajectories independent of any observations) in the column is occurring faster than the assimilation of surface observations can damp it. Conversely, decreasing spread over time indicates that the internal error damping by the assimilation is occurring faster than the error growth. Vertical correlation, which can be inferred from the mean increments and spread, also plays a role in harnessing the internal error growth for effective use in the assimilation of surface observations.

Time-height cross sections of prior ensemble spread also show different behavior among the different variables (Fig. 13). Spreads in both θ and Q behave similarly (Fig. 13a, b). Prior ensemble spread at 0800 LT is simply the climatological spread, which significantly decreases up to 2000 m from assimilation of the 0800 LT surface observations. Subsequently, a slower decrease is evident through a depth modulated by the evolution of the mixed layer. After 1900 LT the PBL collapses,

but the assimilation continues to modify the residual layer, which also has residual spread. The assimilation reduces the spread more quickly near the surface, where the correlation with observations remains large, but the lack of both correlation and variance within the profile leads to a slower reduction in spread.

[Figure 13 about here.]

Continually decreasing prior spread in θ and Q profiles show that the internal error growth in the model is generally slow compared to the effect of assimilation on the column. One exception is during the phase of rapid PBL growth in the bottom 500 m of the domain, when the spread in Q increases between approximately 1200 – 1300 LT. The dynamics of the model restrict significant internal error growth to the daytime hours, when mixing is occurring. Internal error growth is often identically zero (infinitely slow) at night because no tendency on the scalar column is present to generate ensemble variability. Including advective tendencies in the model, sampled from a climatology or another relevant distribution, would generate internal error growth in the scalars.

Evolution of the prior spread in U profiles is somewhat different after 1300 LT (Fig. 13c). The collapse of the mixed layer is more notable between 1500 – 2000 LT, marked by a rapid decrease in prior ensemble spread followed by nearly constant spread later at night. The winds in the column are prone to error growth from the relaxation to geostrophic winds where mixing is weak or absent (eqs. 1 and 2). Internal error will grow because each ensemble member is forced with different geostrophic winds. Above the effect of assimilation in the column, the internal error grows nearly monotonically. In this case, total error will also grow because the distribution of geostrophic winds is biased compared to any single run, and the model is imperfect.

Together, the ensemble-mean increments and the evolution of the spread explain why skillful PBL estimates are possible when assimilating surface observations with an ensemble filter and

this simple column model. In all three variables (θ , Q , and U), the daytime skill is a direct function of covariance built through the parameterized convection in the mixed layer, which ties the model state in the surface layer to the column in the PBL via the turbulent fluxes. This result is expected, and the primary advantage of the ensemble filter approach is the temporal evolution of the covariance that results from the turbulent fluxes.

At the end of the day, the covariance structure in the profile is robust within the limits of model and sampling error. This covariance is used effectively by the assimilation system to cross the transition to night without a significant loss of skill. Thereafter, the skill in scalar profile estimates depends almost entirely on the rate of change of the real atmosphere. Ensemble-mean values of θ and Q are almost stationary, while the spread decreases slowly. The column above the surface layer does not experience the effects of any dynamics or external forcing, and changes on very long time scales. Thus the skill at night depends on the time scale of atmospheric dynamics. These results suggest that the error-growth time scale in the residual layer is slow enough such that the error does not saturate to the climatological value before sunrise, when the vertical flux profiles can again dominate.

After the transition to night, the winds continue to be relaxed toward the climatological geostrophic wind distribution. The ensemble-mean increments continue to show the effect of the covariance, which may be climatological, but the spread does not change much after midnight (LT), suggesting that the time scale of the internal error growth is similar to the time scale of error damping. Thus the error saturates at some level slightly below climatology (Fig. 4d). The level of saturation does depend on the climatological error level, and would be much lower if the forcing were drawn from a distribution conditioned on the flow of a particular analysis scenario.

The analysis presented in this section demonstrates the benefit of a cycling data assimilation system. Namely, a skillful analysis obtained while covariance structures are most useful results in

the propagation of skill in time, regardless of model dynamics.

5. Summary

This work quantifies the influence of real observations, assimilated via an ensemble filter with a 1D column model, on the skill of estimated PBL profiles. During assimilation, the only two sources of information are surface observations and the background ensemble of states propagated from the previous assimilation time, one hour prior. Skill is quantified against rawinsonde observations and computed across approximately 45 cases over Oklahoma during May–July 2003. Comparison with climatological ensembles quantifies the time-integrated positive influence of assimilating the surface observations. Sensitivity to the chosen parameterization schemes, sensitivity to specified observation error variance, and the relative importance of state propagation and the observations are quantified.

The primary conclusions of this work are as follows:

- Analyzed (estimated) PBL profiles show a significant reduction in the error (e.g. up to 85% in the lower 900 m for the potential temperature) relative to error in climatological simulations (Figs. 2 – 4). This reflects the information content of the surface observations assimilated with the ensemble filter.
- Details of the results depend on the choice of parameterization schemes and the specified observation error variance, but skillful PBL estimates are possible under a variety of specifications (Figs. 6 – 11). The effects of using different parameterization schemes are different, depending on whether assimilation is included.
- Skillful estimates are possible at night even without relevant dynamics or forcing condi-

tioned on a particular flow scenario. Information from the daytime observations are temporally propagated in the nighttime residual layer, and the dynamics time scale there is slow (Figs. 12, 13).

The experiments reported here were designed to quantify the effects of real surface observations on a column, with little other information conditioned on a particular flow scenario. Providing a more realistic distribution for initialization and forcing is likely to improve overall skill through the entire column. Mesoscale NWP forecasts of the PBL state do not typically saturate to climatological error levels in quadratic metrics within 24 h, and could be used to construct a distribution for initialization and forcing. Such a distribution would be centered on a timely NWP forecast, with spread structured to reflect initial-condition uncertainty. Because 3D forecasts are limited by truncation and bias, among other things, assimilation of surface observations in the column is expected to reduce error in the PBL structure. Bias in the PBL would be reduced, and the resulting column would be effectively down-scaled from the NWP grid. Such improved PBL columns may be useful for nowcasting applications or downstream applications such as plume dispersion models. Additionally, probabilistic information in the column would be available.

The results with this simple model suggest that surface observations could be valuable in an ensemble filter assimilation system for a 3D mesoscale model. The column model lacks important dynamics, and model error is undoubtedly an important factor. Current NWP models may show less error and more favorable error growth properties. The near-surface horizontal covariance structures from the model require investigation. Impacts on gravitational modes in the assimilating model may also warrant attention.

Some deficiencies in the column model are easily addressed, but others are less tractable. A long history of research in column models is evident in the literature, presenting several possi-

ble approaches for including advection. Fundamental problems with the PBL parameterization schemes are more difficult, but the present model may prove useful for identifying first-order errors. Because it is coupled to a data assimilation system, the evolution of some parts can be tightly constrained, particularly when robust observing systems are available. Selective observations can be withheld to observe the behavior of unconstrained components of the system.

Successful estimation of PBL profiles opens several avenues of research to understand the deficiencies of parameterization, the column response to different forcing (including stochastic terms), and the behavior of other types of data assimilation systems with surface and PBL observations. Work continues in these areas, with the eventual goal of useful mesonet assimilation in operational data assimilation for mesoscale NWP.

Acknowledgement The authors are grateful to the developers of the NCAR Data Assimilation Research Testbed (DART) for providing a useful platform for experimentation. We acknowledge C. Snyder for early discussions motivating the potential usefulness of a column estimate, and A. Aksoy and J. Anderson for thoughtful internal reviews strengthening the manuscript. Two anonymous reviews improved the accuracy of the discussion. J. Hacker was supported by the Army Test and Evaluation Command while performing this research. D. Rostkier-Edelstein acknowledges the NCAR Mesoscale and Microscale Meteorology and Research Applications Laboratory visitor funds. Data were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Environmental Sciences Division.

References

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.
- ARM, 2006: Surface meteorological observation system handbook. Technical Report ARM TR-031, Atmospheric Radiation Measurement Climate Research Facility, U.S. Department of Energy, available from <http://www.arm.gov>.
- Beljaars, A. C. M. and A. A. M. Holtstag, 1991: Flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteor.*, **30**, 327–341.
- Burgers, G., P. VanLeeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767–1785.
- Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land-surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108D**, 8851–8867.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.
- Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

- Hacker, J. P., J. L. Anderson, and M. Pagowski, 2006: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Wea. Rev.*, accepted.
- Hacker, J. P. and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.*, **133**, 3260–3275.
- 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.
- Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Janjić, Z. I., 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. Technical Report 437, National Centers for Environmental Prediction Office Note.
- Janjić, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285.
- Martin, W. J. and M. Xue, 2006: Sensitivity analysis of convection of the 24 May 2002 IHOP case using very large ensembles. *Mon. Wea. Rev.*, **134**, 192–207.
- McCaul, E. W. and C. Cohen, 2002: The impact on simulated storm structure and intensity of variations in the mixed layer and moist layer depths. *Mon. Wea. Rev.*, **130**, 1722–1748.
- 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>.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF Version 2. Technical Report TN-468, National Center for Atmospheric Research.
- Troen, I. and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **37**, 129–148.
- Zilitinkevich, S. S., 1995: Non-local turbulent transport: pollution dispersion aspects of coherent structure of convective flows. Power, H., N. Moussiopoulos, and C. A. Brebbia, eds., *Air Pollution III - Volume I. Air Pollution Theory and Simulation*, 53–60, Computational Mechanics Publications, Southampton Boston.

List of Figures

1	Ratio of ensemble spread (standard deviation squared) to MSE in the posterior ensemble mean (after assimilation) for T_2 (long-dashed), Q_2 (short-dashed), and U_{10} (solid).	31
2	MAE profiles of θ for the assimilation experiments (solid) and climatological simulations (dashed), presented in order of increasing time from initialization (0700 LT). Verification times are (a) 1300 LT, (b) 1900 LT, (c) 0100 LT, and (d) 0700 LT.	32
3	Same as Fig. 2 but for Q	33
4	Same as Fig. 2 but for the U -wind component.	34
5	Ensemble-mean absolute error of (a) T_2 , (b) Q_2 , and (c) U_{10} for the climatological simulations (no data assimilation). The solid curve shows results when the MYJ PBL scheme is used, and the dashed curve shows results when the YSU scheme is used.	35
6	Posterior ensemble-mean absolute error of (a) T_2 , (b) Q_2 , and (c) U_{10} for data-assimilation experiments. The solid curve shows results when the MYJ PBL scheme is used, and the dashed curve shows results when the YSU scheme is used.	36
7	MAE profiles of U -wind for the base case assimilation experiments using the MYJ PBL scheme (solid), and experiments with the YSU PBL scheme (dashed), presented in order of increasing time from initialization (0700 LT). Verification times are (a) 1300 LT, (b) 0100 LT.	37

8	In (a) percent difference of ratios of ensemble spread to MSE in the ensemble mean in the posterior (after assimilation) between the base case (Fig. 1) and panel (b) here, the case when observation error variances are small. Results are shown for T_2 (long-dashed), Q_2 (short-dashed), and U_{10} (solid).	38
9	MAE profiles of θ for the assimilation base case experiments (solid), and experiments with small σ_o^2 (dashed), presented in order of increasing time from initialization (0700 LT). Verification times are (a) 1300 LT, (b) 0100 LT.	39
10	Same as Fig. 9 but for Q	40
11	Same as Fig. 9 but for the U -wind component.	41
12	Ensemble-mean absolute analysis increment in the profiles of (a) θ ($^{\circ}\text{C}$), (b) Q (g kg^{-1}), and (c) U -wind (m s^{-1}) as a function of analysis time.	42
13	Prior ensemble spread in the profiles of (a) θ ($^{\circ}\text{C}$), (b) Q (g kg^{-1}), and (c) U -wind (m s^{-1}) as a function of analysis time.	43

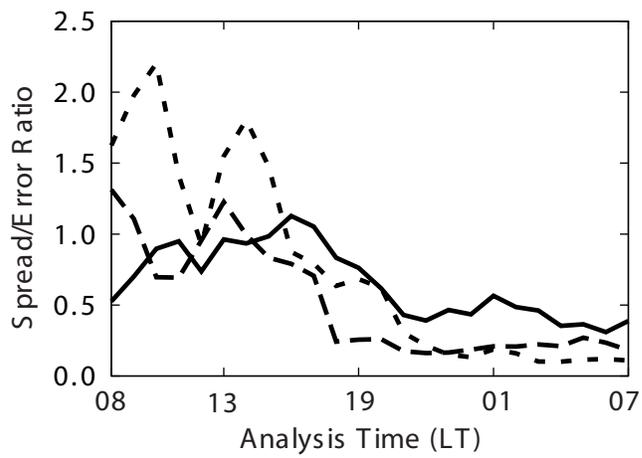


Figure 1: Ratio of ensemble spread (standard deviation squared) to MSE in the posterior ensemble mean (after assimilation) for T_2 (long-dashed), Q_2 (short-dashed), and U_{10} (solid).

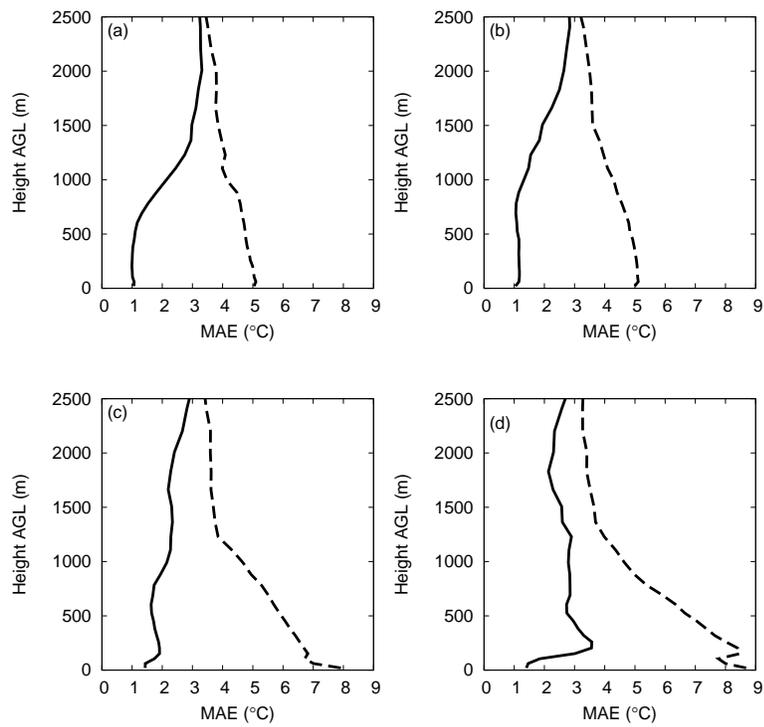


Figure 2: MAE profiles of θ for the assimilation experiments (solid) and climatological simulations (dashed), presented in order of increasing time from initialization (0700 LT). Verification times are (a) 1300 LT, (b) 1900 LT, (c) 0100 LT, and (d) 0700 LT.

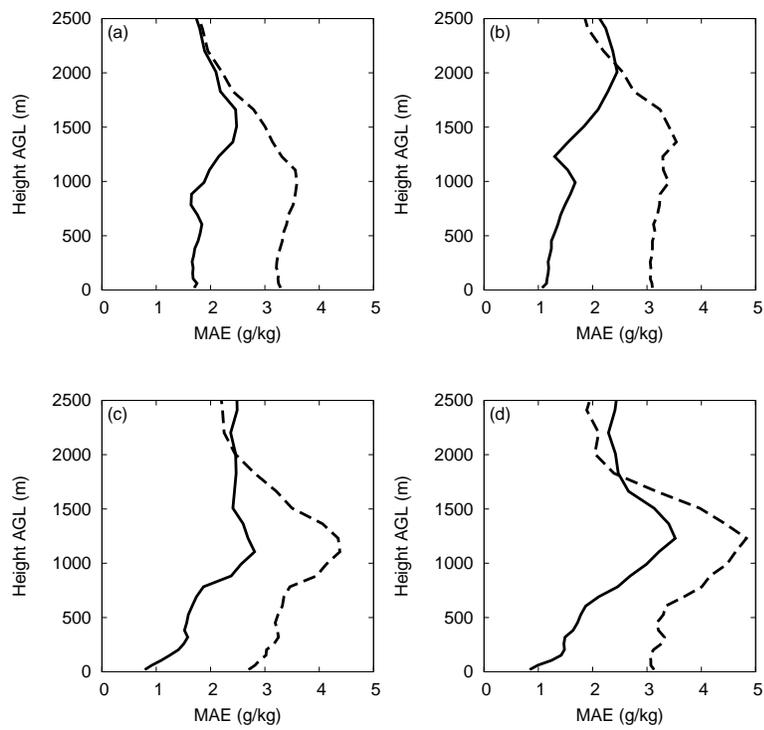


Figure 3: Same as Fig. 2 but for Q .

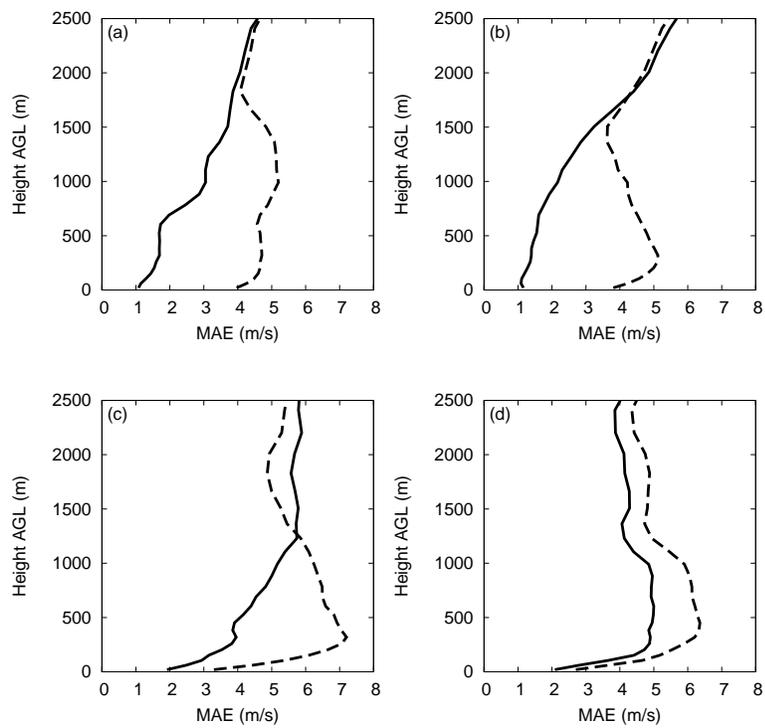


Figure 4: Same as Fig. 2 but for the U -wind component.

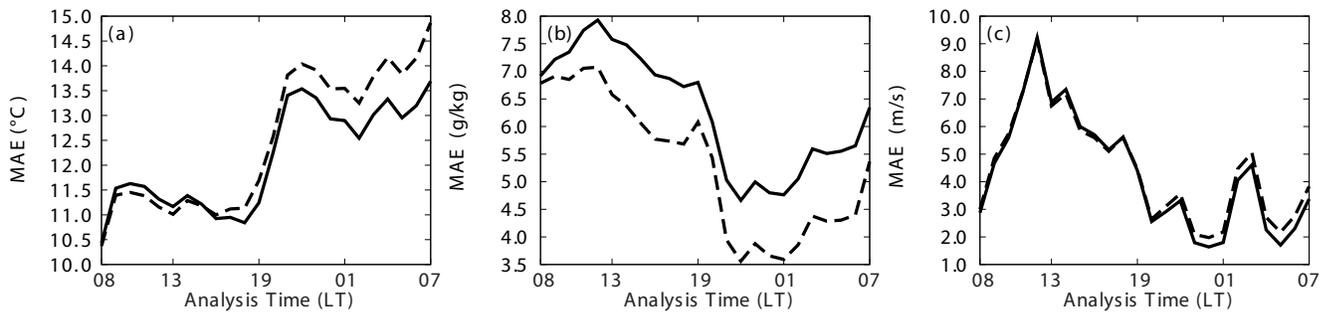


Figure 5: Ensemble-mean absolute error of (a) T_2 , (b) Q_2 , and (c) U_{10} for the climatological simulations (no data assimilation). The solid curve shows results when the MYJ PBL scheme is used, and the dashed curve shows results when the YSU scheme is used.

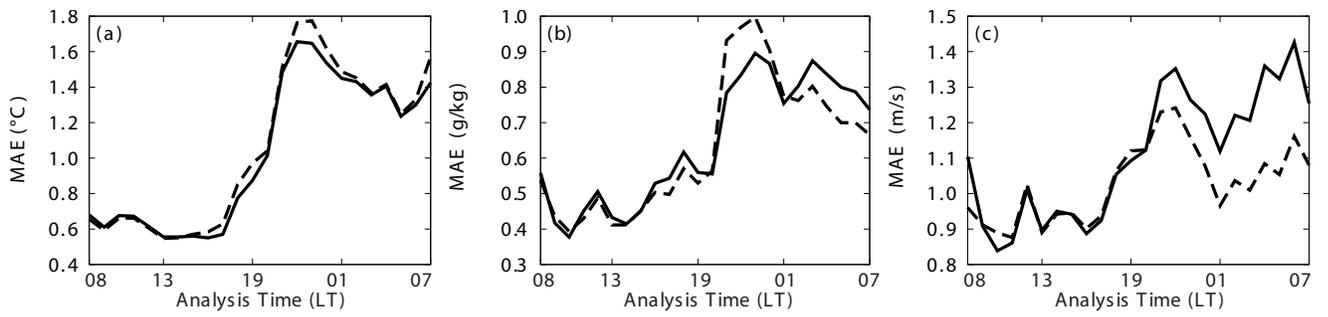


Figure 6: Posterior ensemble-mean absolute error of (a) T_2 , (b) Q_2 , and (c) U_{10} for data-assimilation experiments. The solid curve shows results when the MYJ PBL scheme is used, and the dashed curve shows results when the YSU scheme is used.

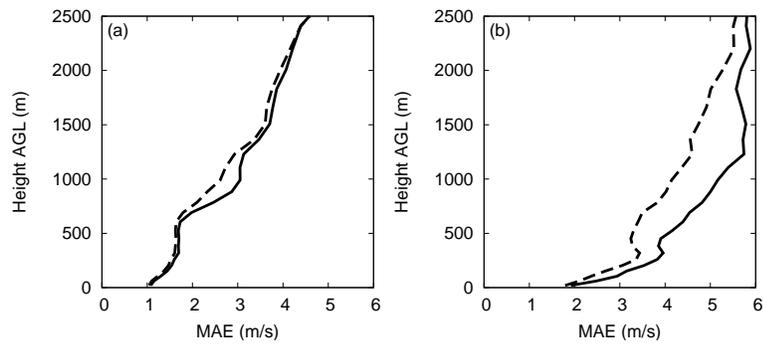


Figure 7: MAE profiles of U -wind for the base case assimilation experiments using the MYJ PBL scheme (solid), and experiments with the YSU PBL scheme (dashed), presented in order of increasing time from initialization (0700 LT). Verification times are (a) 1300 LT, (b) 0100 LT.

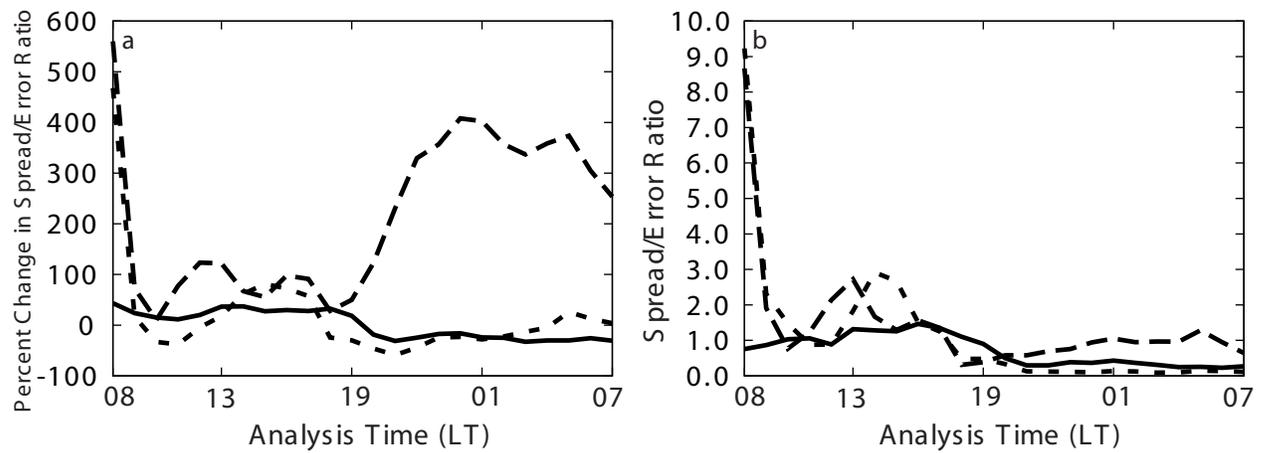


Figure 8: In (a) percent difference of ratios of ensemble spread to MSE in the ensemble mean in the posterior (after assimilation) between the base case (Fig. 1) and panel (b) here, the case when observation error variances are small. Results are shown for T_2 (long-dashed), Q_2 (short-dashed), and U_{10} (solid).

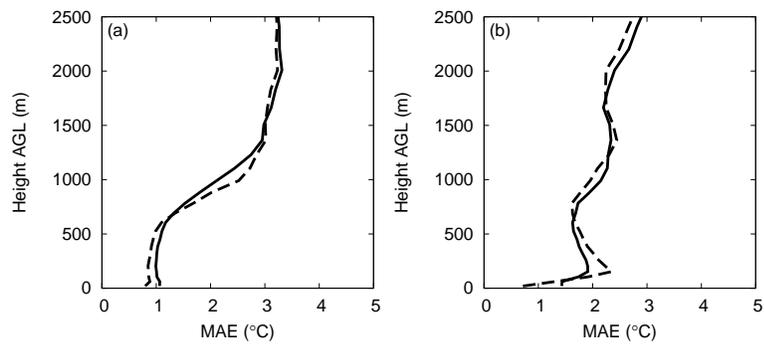


Figure 9: MAE profiles of θ for the assimilation base case experiments (solid), and experiments with small σ_o^2 (dashed), presented in order of increasing time from initialization (0700 LT). Verification times are (a) 1300 LT, (b) 0100 LT.

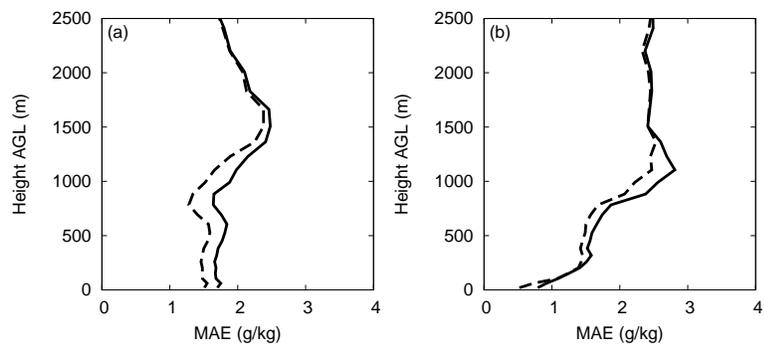


Figure 10: Same as Fig. 9 but for Q .

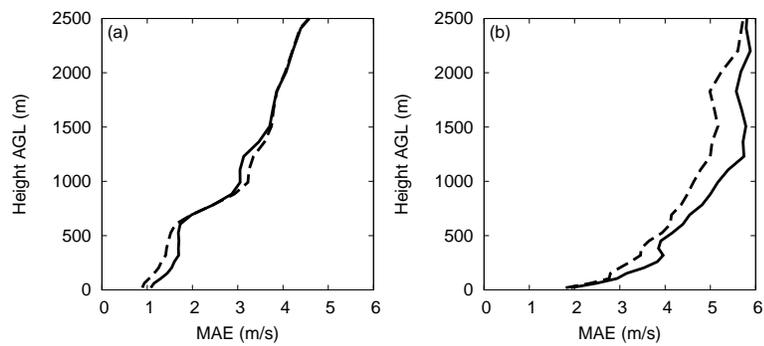


Figure 11: Same as Fig. 9 but for the U -wind component.

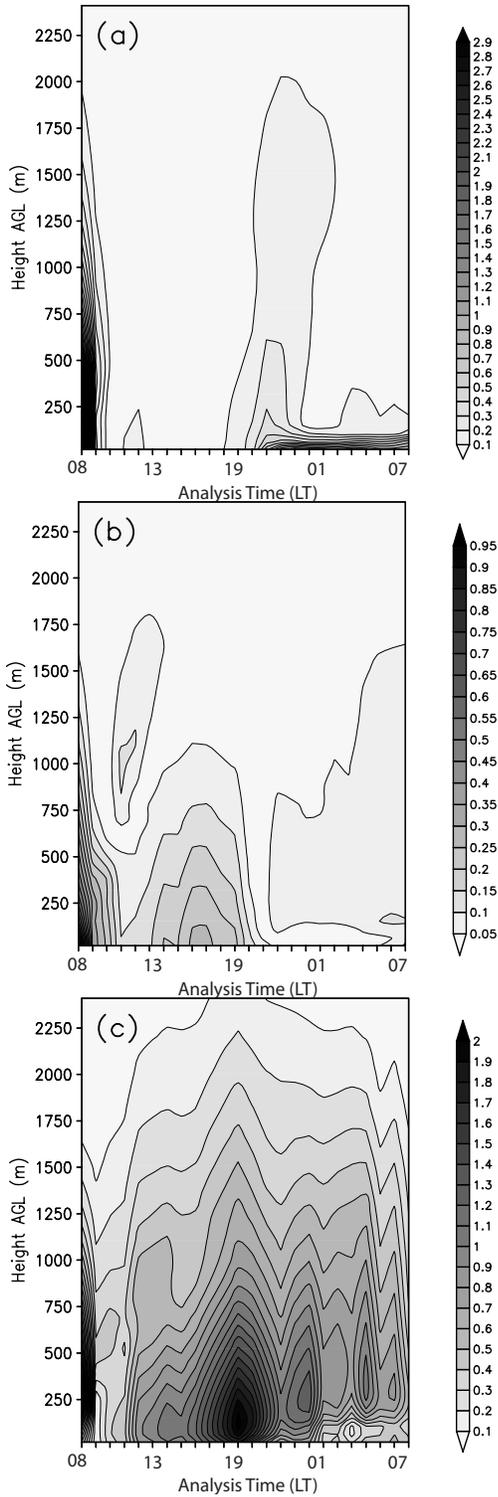


Figure 12: Ensemble-mean absolute analysis increment in the profiles of (a) θ ($^{\circ}\text{C}$), (b) Q (g kg^{-1}), and (c) U -wind (m s^{-1}) as a function of analysis time.

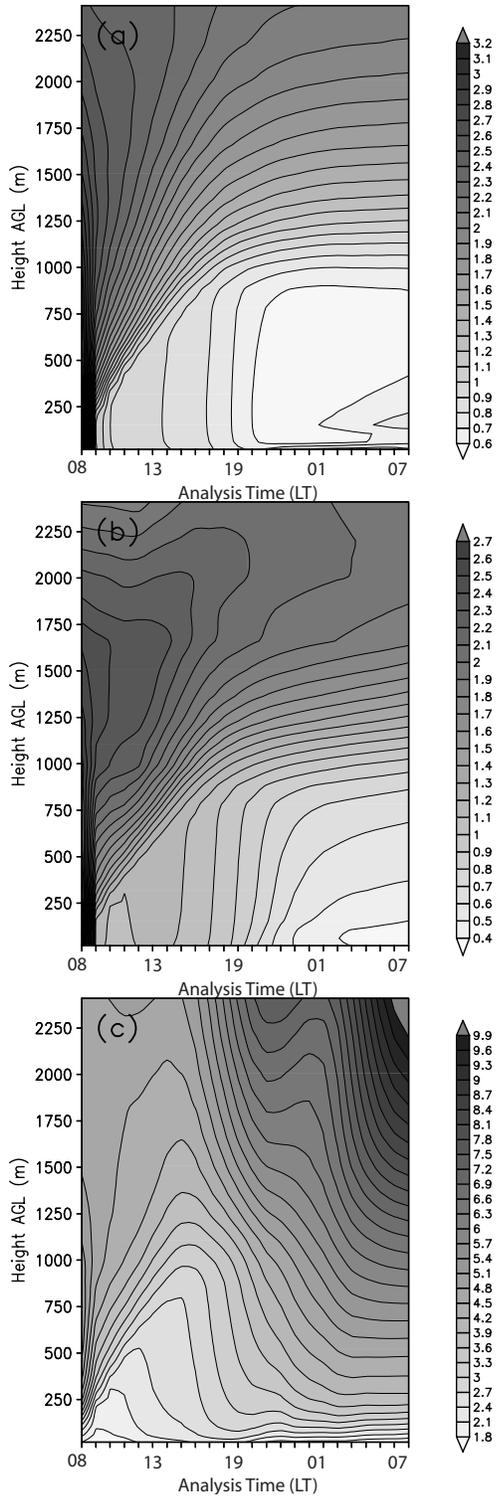


Figure 13: Prior ensemble spread in the profiles of (a) θ ($^{\circ}\text{C}$), (b) Q (g kg^{-1}), and (c) U -wind (m s^{-1}) as a function of analysis time.