17A.3 Probabilistic nowcasting of PBL profiles with surface observations and an ensemble filter

DORITA ROSTKIER-EDELSTEIN *
IIBR, Israel

JOSHUA P. HACKER
The National Center for Atmospheric Research†, Boulder, CO

July 15, 2009

1. Introduction

Successful nowcasting and forecasting of the state of the planetary boundary layer (PBL) is of value for a wide range of practical forecasting applications. The aim of this work is to find an efficient method for probabilistic nowcasting (0-3 h forecasting) of the state of the PBL wherever surface observations are available under a vast range of weather scenarios. One can also view it as a probabilistic PBL profile retrieval approach, given a background ensemble and surface observations. Surface-layer in-situ observations (2-m shelter and 10-m anemometer height) are a dense, inexpensive, accurate and reliable data source. However, as discussed by Hacker and Snyder (2005) and Hacker and Rostkier-Edelstein (2007) (hereafter referred to as HR07), several difficulties limit their optimal utilization in present operational numerical weather prediction and data assimilation systems. HR07 showed that surface observations can be an important source of information with a single column model (SCM) and an ensemble filter (EF). Comparisons in that work were against free-running simulations, representing a"climatological" distribution, to obtain a direct measure of the impact for surface observations. The results showed that without additional sources of information, it is possible to obtain error levels in the lowest few-hundred meters in the atmosphere that approach observation-error levels.

We extended this work to quantify the probabilistic skill of the same SCM, and the SCM with added complexity [Hacker and Rostkier-Edelstein (2008), hereafter referred to as HR08; and the forthcoming paper Rostkier-Edelstein and Hacker (2009), hereafter referred to as RH09]. We also use a relevant 3D forecast to center the ensemble on each day, rather than using a climatological ensemble as before, to generate an ensemble that is valid for a particular time.

Although it is appealing to add additional physics and dynamics to the SCM, expecting that these will make solutions more realistic and reduce errors that arise from the imperfect model, it is not immediately clear that additional complexity will improve the performance of a PBL nowcasting system based on a simple model and an EF. This question is investigated with regard to two model components and their importance relative to surface assimilation: parameterized radiation in the column and horizontal advection to account for realistic 3D dynamics. The cost of these, when running with tens to hundreds of ensemble members (and possibly at many surface observation sites simultaneously) can be significant (at least a 40% increase in the computational effort). Thus it behooves us to quantify the role of the added complexity in a probabilistic sense.

Direct examination of the probabilistic skill of the system as a function of the model components may be cumbersome, and we instead seek a framework for simpler interpretation. One framework is factor separation (FS) analysis (Stein and Alpert, 1993), which is useful in this case as it quantifies the individual contribution of each model component to the probabilistic performance of the system, as well as any beneficial or detrimental interactions between them. The present analysis is the first to apply the FS technique in the framework of deterministic and probabilistic verification of EF assimilation and forecasts. Primary results of the FS analysis in observation space have been presented in HR08.

To assess the real utility of the flow-dependent error covariances estimates we compare the skill of the SCM/EF system to that of a reference system based on climatological error covariances and surface forecast errors.

The complete analysis and set of results is presented in RH09. In this paper we quote representative results and survey the main findings.

* *Corresponding author address:* Dorita Rostkier-Edelstein, Israel Institute for Biological Research, P.O.Box 19, Ness-Ziona, 74100, Israel. Email: doritar@iibr.gov.il

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

2. PBL prediction methods

2.1 Single column model

The SCM used in this study has been described in detail in HR07, HR08 and RH09. We briefly summarize the components considered for improving probabilistic analyses and nowcasts. It contains vertical turbulence, atmospheric surface layer, and land-surface parameterizations identical to those in the Advanced Research WRF version 2.2 (Skamarock et al., 2005). Resolved dynamics are integrated in time with a Crank-Nicholson time step of 10 seconds, on a vertically-stretched column with 81 levels and the model top at approximately 16 km.

Horizontal advection to allow for the effects of 3D dynamics is implemented following (Ghan et al., 1999) which describe an approach for upstream advection of temperature (T), water vapor mixing ratio (Q_v), and wind (U and V components) in SCMs. It acts to relax the SCM state toward a prescribed 3D state (which may be time dependent) on the advective time scale. In the absence of any other forcing terms, the SCM would track the prescribed time-dependent state.

The RRTM long-wave (Mlawer et al., 1997) and Dudhia short-wave Dudhia (1989) radiation schemes are also options. The long-wave scheme is included particularly to improve simulations during the night, when radiative cooling can be important in the PBL.

2.2 Ensemble filter implementation

EF assimilation of surface observations is the third model component examined here. It is considered alone, as was described in HR07, and also in combination with the advection and radiation described above.

The specific SCM and EF implementations are summarized in detail in HR07, HR08 and RH09 and we present here the most relevant information.

The SCM is coupled to the Data Assimilation Research Testbed (DART), a comprehensive ensemble filter system designed for research and education on ensemble filters with a variety of dynamical systems, and developed at the National Center for Atmospheric Research (NCAR). Many filter algorithms are available in DART and we chose the default ensemble adjustment Kalman filter algorithm (EAKF; Anderson 2001), which is a square-root filter and here implemented in serial (Anderson, 2003). Because of the extensive documentation in the literature, we do not include a description of that filter, and refer the interested reader to the citations above. Predictions are simply the 30-minute ensemble forecasts from the EF analyses.

The ensemble mean of initial conditions, large-scale forcing, advective tendencies and surface radiation is specified equal to a WRF forecast valid for a given day and hour. Perturbations are created by adding to it the scaled difference between that forecast and another forecast from the archive. This additional member is selected randomly from the experiment period, and the scaling of the difference is drawn randomly from a normal distribution $\mathcal{N}(0,1)$. This approach enables the creation of arbitrarily large ensembles, and we use 100 members in our experiments. The same weights and WRF forecasts are used for both the initial conditions and the forcing, so that forcing time series and initial conditions are consistent.

Ensembles are derived from a distribution of WRF forecasts with wider variance than the real analysis and forecast error, potentially introducing too much spread in the ensembles relative to ensemble-mean error. Although the assimilation of surface observations will contract the distribution quickly (c.f. HR07) the advective tendencies can lead to unrealistically rapid growth in ensemble spread. In an effort to avoid this we adopt a state-augmentation approach by which the advection speed is dynamically tuned with the surface observations simulating the effect of assimilating data in three spatial dimensions. Additional modeling details, such as the vertical covariance localization are described in HR07.

2.3 Climatological dressing (CD) technique

The reference system, based on climatological covariances from the 3D WRF, is in essence a dressing technique, whereby a deterministic 3D WRF mesoscale forecast is dressed with a normal distribution according to a sample of forecasts and surface forecast errors. In this implementation, the sample is the same sample of WRF forecasts that provide forcing for the SCM, and the most-recent valid WRF forecast is dressed using statistics derived from the complete sample. The results should be unrealistically good because the forecast and dressing samples are the same, and thus the dressed forecast represents a difficult reference for the SCM ensemble system to beat. We are exploiting WRF-derived covariances and a recent observation to generate a probabilistic PBL prediction without performing data assimilation.

The approach is to first correct the error in the profile using covariances computed from the distributions of available 3D WRF forecasts, and then dress a timely deterministic forecast with an uncertainty distribution scaled by the most recent observed error. A forecaster might have access to a recent mesoscale forecast such as the WRF forecast initialized at e.g. 00 UTC. Given a surface observation o_k at time k, the error in the corresponding forecast f_k is $d_k = o_k - f_k$. If we are still addressing a true forecast and the observation time has not yet passed, o_k is unavailable. Because we are interested in very-short range forecasting (30 min.), we assume that persistence over the 30-minute prediction time is the best estimate

of the error: $d_k \approx d_{k-1}$. This error is then used, with climatological model covariances, to produce a probabilistic short-range forecast from a recent deterministic forecast.

WRF forecast covariance in the column are computed from the sample of all available forecasts during the experiment period that are valid at the verification time. For example, the collection of forecast lead times of 5.5 h and 17.5 h from 00 UTC are used to correct and dress each WRF forecast at 0530 and 1730 UTC, respectively. Values in the column can be adjusted along the regression line that relates an arbitrary column variable x_k and the model prediction f_k:

$$\Delta x = \frac{\sigma_{xf}^2}{\sigma_f^2} d \ , \tag{1}$$

where the time subscripts have been dropped for clarity. In (1), Δx is the adjustment applied to a WRF-column state variable, σ_{xf}^2 is the covariance between the state variable and the forecast in observation space, and σ_f^2 is the variance of the forecast in observation space.

To estimate uncertainty we dress the adjusted profile by scaling the variance of the climatological distribution to give a reliable system in observation space:

$$\begin{aligned} \langle d^2 \rangle &= \alpha \sigma_f^2 + \sigma_o^2 \\ \alpha &= \frac{\langle d^2 \rangle - \sigma_o^2}{\sigma_f^2} \end{aligned} \tag{2}$$

In (2), $\langle \circ \rangle$ is the expectation operator, α is the derived scaling factor. Because the individual error d_k is not useful here (i.e. only the climatological mean is used), all quantities in (2) are valid at the dressing time of day. This is universally applied in state space such that:

$$\begin{aligned} \Delta\sigma_x^2 &= (\alpha - 1)\,\sigma_x^2 \\ \Delta\sigma_x^2 &= \left(\frac{\langle d^2 \rangle - \sigma_o^2}{\sigma_f^2} - 1 \right) \sigma_x^2 \ , \end{aligned} \tag{3}$$

where $\Delta\sigma_x^2$ is the difference between scaled and unscaled variance of a state variable, and σ_o^2 is the observation error variance assigned as in the assimilation system. Given distributions of f, o, and observation error valid for a single forecast lead time (and corresponding time of day), the scaling does not depend on a particular forecast. In the present application, the RHS of (3) is negative so the scaling acts to shrink the climatological variance.

2.4 Experiment design

The experiment period is 3 May – 15 July 2003, and the SCM is configured to run over the Atmospheric Radiation Measurement (ARM) Central Facility near Lamont, Oklahoma. It offers high-quality surface data and balloon-borne soundings. We use 30-minute average surface observations of winds, T, and Q_v for both assimilation and verification. Soundings are used for verifying profiles aloft. Archived runs of the ARW version 2.1 ($\Delta_x = 4$ km), coinciding with the experiment period are used for ensemble initialization and to provide advective terms tendencies, large scale forcing and surface radiation.

SCM simulations are initialized at 0300 and 1500 UTC. Assimilation of surface observations is accomplished every 30 minutes, starting at 0400 and 1600 UTC respectively. Observations error variances are specified the same as in HR07, which agree roughly with values estimated in Crook (1996).

The 30-minute surface and PBL forecasts are verified at 0530 and 1730 UTC (0030 and 1230 LT) against the surface observations and the available soundings respectively, corresponding to assimilation cycling over periods of 2.5 hours prior to verification. Skill scores, described in the next section, use 71 verification times for the surface variables and 57 and 65 verification times for the atmospheric profiles during the experiment at day and night time respectively. The total number of verification times is determined simply by the number of available surface observations and the times that both the soundings and the WRF forecast were available during the experimental period. The limited number of verification events raises the question of statistical significance of our results. Confidence intervals (CIs) are calculated for each verification score following a bootstrapping technique as detailed in section 3.3.

3. Analysis Methods

3.1 Factor separation analysis

Implementation of FS analysis requires the examination of the system performance for all 2^3 possible model configurations associated with the three studied model components: assimilation, radiation, and advection. The possible combinations are listed in Table 1 for reference.

Table 1: Key to all possible combinations of the model components considered here.

Label	Advection	Radiation	Assimilation
000	NO	NO	NO
100	YES	NO	NO
010	NO	YES	NO
001	NO	NO	YES
110	YES	YES	NO
101	YES	NO	YES
011	NO	YES	YES
111	YES	YES	YES

FS equations are derived by a Taylor expansion of the effect of each component. Stein and Alpert (1993) derived the resulting equations for the factors f_{ijk} for three model components [eqs. (17)-(24) in their manuscript]. Rearranging those equations allows us to directly quantify the effect of adding new system components or sets of components to the base system. We then define the variables ef_{ijk} as the "effect" of a given factor, f_{ijk}, on the skill of the base system configuration, e_{000}, as follows:

$$
\begin{aligned}
ef_{000} &= f_{000} = e_{000} \\
ef_{100} &= f_{000} + f_{100} = e_{100} \\
ef_{010} &= f_{000} + f_{010} = e_{010} \\
ef_{001} &= f_{000} + f_{001} = e_{001} \\
ef_{110} &= f_{000} + f_{110} = e_{110} - (e_{100} + e_{010}) + 2e_{000} \\
ef_{101} &= f_{000} + f_{101} = e_{101} - (e_{100} + e_{001}) + 2e_{000} \\
ef_{011} &= f_{000} + f_{011} = e_{011} - (e_{010} + e_{001}) + 2e_{000} \\
ef_{111} &= f_{000} + f_{111} = e_{111} - (e_{110} + e_{101} + e_{011}) \\
&\qquad + (e_{100} + e_{010} + e_{001}) \; .
\end{aligned}
$$

ef_{000} represents the joint effect of all model components that are not analyzed in the present study on the system performance (and it also represents the system skill itself). ef_{001}, ef_{010} and ef_{100} show the pure effect of each single model component evaluated in this study. In these cases the values of ef_{ijk} represent also the measured performance of the system (e_{ijk}) when each individual component is included in the system, and in the absence of second or third components. In contrast, ef_{011}, ef_{110}, ef_{101} and ef_{111} show the effect of synergistic factors on the system performance.

Values of ef_{ijk} do not correspond to the measured performance of any configuration and they cannot be directly computed from the output of individual simulations. Rather, they show the change in performance of the base configuration due to non-linear interaction between the studied model components. Dual and triple interactions are obtained when the system is run in a configuration that includes more than one of the studied components. Synergistic effects are not always obvious from direct examination of the e_{ijk} values. This representation is also convenient as CIs are calculated for the variables ef_{ijk}, thus they precisely show the statistical significance of the change induced by each factor to the base configuration. Moreover, the values of some of the chosen verification metrics (see section 3.2) are characterized by bounded skillful ranges. The calculation of CIs for ef_{ijk} directly shows whether a given factor improves the system and brings the scores to a skillful range, or on the contrary it pushes the scores out of a skillful range.

3.2 Verification metrics

The mean absolute error (MAE), rather than the root-mean square error (RMSE), is chosen as a deterministic metric to quantify systematic error in the ensemble mean because it is more resistant to outliers.

The value of the probabilistic information provided by the SCM/EF system is measured by three main attributes of the probabilistic forecasts: reliability, resolution and discrimination.

Reliability is a measure of the statistical consistency between the predicted probability distributions and the verifying observations. In a reliable system the verifying observation is statistically identical to a random realization drawn from the predicted distribution. A reliable system provides unbiased estimates of the observed frequencies associated with different forecast probability values. Because reliability can be achieved by predicting the climatological probability distribution, it is necessary but not sufficient for a valuable probabilistic forecast.

Resolution measures the capability of the forecast to distinguish between separate groups of observed events when these have a frequency different from the climatological frequency. Resolution and the closely related attribute of discrimination are not independent attributes but they are not the same. Resolution refers to the conditional distribution of the observations given the forecasts, while discrimination evaluates for the conditional distribution of the forecasts given the observations. Both of them are independent of reliability and represent a measure of the potential value of the system, since resolution is insensitive to forecast bias.

The probabilistic skill of the system is evaluated through the Brier Skill Score (BSS; Wilks 1995) and the area under the relative operating characteristic (ROC) curve (AUR; Mason and Graham 1999). Both of these metrics evaluate the probabilistic prediction of the occurrence of a binary event and measure the system performance relative to a reference system. The BSS is easily decomposed into a reliability and resolution term (Murphy, 1973) to understand the trade-offs in different com-

ponents of probabilistic skill. The ROC curve has been widely used in the field of signal detection to distinguish between two alternative results (Mason, 1982), thus, the AUR quantifies the ability of the forecast to discriminate between events. We define an "event" here to be a forecast value exceeding the 75^{th} percentile of the observations at the ARM central facility during the experiment. The climatology derived from the observations during the experiment period is used as reference in these calculations.

3.3 Confidence interval calculation

Since the number of realizations, I, is finite, an estimate of the uncertainty in the verification scores is required to allow meaningful statistical conclusions. We quantify uncertainty in the scores through the estimation of CIs computed via a bootstrapping resampling procedure. It consists of recalculation of the scores a number of times, N_b, with a sample of I realizations randomly extracted from the original dataset. We define CIs derived from the bootstrapped distributions of the scores using the bias-correction and acceleration (BCa) technique as described by Efron and Tibshirani (1993). The BCa intervals are corrections to the standard percentile intervals. For example, for a 90% CI the interval would be (0.05, 0.95). The BCa technique adjusts this interval so that the mean of the bootstrapped distribution fits the original estimate of the score (from the original data set) and the width gives a more accurate estimate of the CI.

CIs are calculated for the variables ef_{ijk}, thus, they precisely represent the statistical significance of the effects of the factors relative to the base configuration. Since all experiments are run over the same season and verified against the same observations, the different magnitude of the CIs among the various system configurations at a given verification time reflect the sensitivity of the skill of each system configuration to flow variability from day to day. We seek a robust system that performs consistently skillfully over the range of flow scenarios present during our experiment. This situation is represented by skillful values of a verification score with narrow CIs. Wide CIs indicate skill more sensitive to flow variability.

4. Selected results and interpretation

4.1 FS analysis of SCM/EF forecast profiles

The complete FS analysis is presented in RH09, and here we focus on some of the most representative findings. Along the present paper we chose to compare between e_{000} and ef_{ijk} values for the several verification metrics [1].

4.1.1 Deterministic verification

Fig. 1 compares the MAE of 30-min forecasts of U profiles valid at 1230 LT for the base (ef_{000}, black lines) with the full system (ef_{111}, red lines) configurations. Bold solid lines represent the original results obtained from our experiments and thin dashed lines denote 90% CIs calculated using 1000 bootstraps. The full configuration performs clearly better than the base configuration at all vertical levels.

Detailed FS analysis reveals that pure surface assimilation and advection are the most important factors for improving the system performance. This might be expected because of the strong covariance between the surface and the column in the SCM. Radiation, as expected, has no significant effect on U. Fig. 2(a)-(c) shows the effect of three factors on the MAE of 30-min forecasts of U profile valid at 1230 LT: (a) pure surface assimilation (black lines for the base configuration, ef_{000}; red lines for ef_{001}), (b) pure advection (line types as before but red stands for ef_{100}), and (c) assimilation-advection synergism (line types as before but red stands for ef_{101}).

Surface assimilation and advection play similar roles in the first hundreds of meters AGL and advection dominates aloft. This is caused by decreasing surface-atmosphere coupling (which determines the vertical extent of the effect of assimilation) with height. The assimilation-advection synergistic interaction gives rise to a slight detrimental effect everywhere, indicated by the red curve lying to the right of the black curve, which is most noticeable between 400-600 m AGL. At these intermediate levels the advection dominates the assimilation because the coupling between the observation and the column is weaker. Thus, our state-augmentation strategy aiming to tune the advection with the assimilation of observations is less effective at these levels.

Near 1000 m AGL and above, assimilation does little and the synergism is small; the total benefit is simply from the advection. Comparison between Figures 1 and 2 proves that the detrimental synergism results in less overall improvement than would be attained from the sum of pure assimilation and advection terms. However, 90% CIs partially overlap revealing that in many cases the synergistic detrimental effect is not significant [Fig. 2(c)].

The magnitude of the CIs is an indication of the sensitivity of the calculated scores to flow variability. CIs narrow in the full configuration showing less dependence of the system skill on weather scenarios. Both improving

[1] Jolliffe (2007) pointed out that although extensively used, the comparison between CIs may be misleading in some situations and advises computing CIs on the differences between the scores to draw significant conclusions. This issue is discussed in RH09.

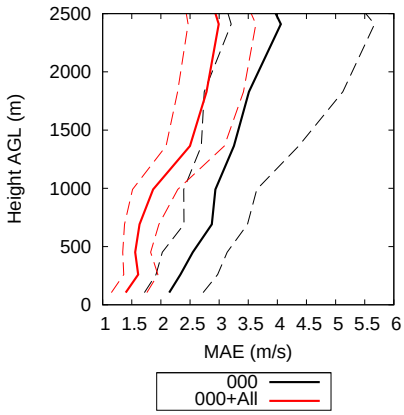


Figure 1: MAE of 30-min forecasts of U profiles valid at 1230 LT for the base configuration (black lines, e_{000}) and for the full configuration including the three studied model components (red lines, e_{111}). Bold solid lines represent the original scores from our experiments, thin dashed lines are 90% confidence intervals calculated using the BCa bootstrapping technique.

factors, i.e., surface assimilation and horizontal advection, act to contract CIs. However, the synergistic effect between them is characterized by wider CIs, which can be expected as it depends on the variability of each of them.

Figures 3, 4 and 5 present a similar analysis to that in Figures 1 and 2 but for 30-min forecasts of θ profiles valid at 0030 LT. The improvement achieved in the full configuration is clearly seen in Fig. 3, and once more the FS analysis facilitates simple interpretation.

Fig. 4(a) reveals that the effect of surface assimilation dominates in the first 200 m AGL and Fig. 4(c) that advection dominates higher. The vertical extent of the effect of assimilation is shallower than that observed in Fig. 1, illustrating its dependence on the flow characteristics (convective unstable vs. nocturnal stable regime), which produce shallow covariance structures. Fig. 4(b) shows that pure radiation plays a minor positive role in the SCM compared to pure assimilation and pure advection and its vertical extent is limited to the first 150 m AGL, above it CIs show that the effect is not statistically significant.

The assimilation-advection non-linear interaction [Fig. 5(a)] leads to a 90% statistically confident negative effect at levels between 50-400 m AGL. It should be noted that the local flow is characterized by shorter advection time scales above the PBL, which is shallow at night because of reduced vertical momentum exchange. In contrast, very close to the surface (up to 50 m) the synergistic effect is null, which results from longer advection time scales and successful assimilation of

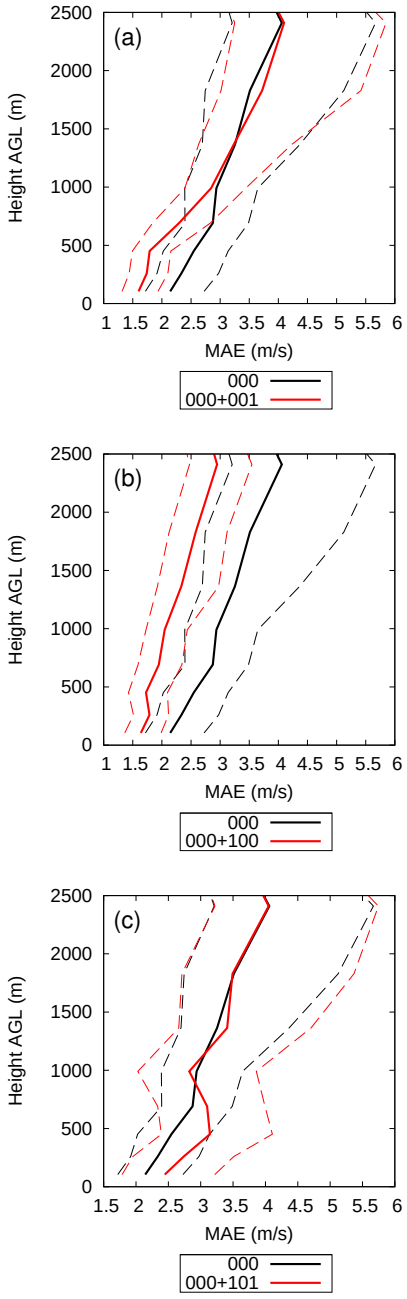


Figure 2: Same as Figure 1 but dashed red lines denote the resulting MAE when the effect of each of the following factors is added to the MAE of the base configuration: (a) pure assimilation (ef_{001}); (b) pure advection (ef_{100}); (c) assimilation-advection synergism (ef_{101}).

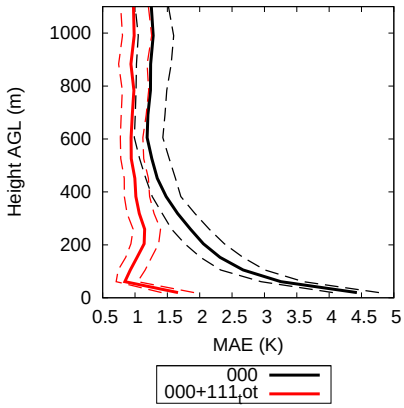


Figure 3: Same as Figure 1 but for 30 min forecasts of θ profiles valid at 0030 LT.

surface observations.

The synergistic interactions of surface radiation with advection and assimilation [Figures 5(b) and 5(c) respectively] lead to slightly negative effects and the short vertical range is consistent with that observed for pure radiation.

Similar to (Alpert P. and Stein, 1995) a general observation of the multi-factor results show that when the number of relevant factors increases, the role of any given factor is reduced because the synergistic interactions more often lead to some cancellation than positive synergism. One interpretation of this is that the system is near a lower bound in MAE, so that multiple components that improve MAE to a similar degree are not needed.

4.1.2 Probabilistic verification

One of the most important findings is the importance of EF surface assimilation to the probabilistic skill of the forecast profiles. Synergistic factors show little effect. Fig. 6(a),(b) and (c) show the BSS and its decomposition into reliability and resolution terms respectively for 30-min forecast of T profiles valid at 0030 LT for the base configuration (ef_{000}, black lines) and for the configuration including assimilation of surface observations only (ef_{001}, red lines). Surface assimilation improves the BSS up to 600 m AGL. It should be noted that the base configuration shows less skill (negative value of BSS) than the reference at levels closest to the ground (up to 50 m). Assimilation provides skill and CIs indicate that in some cases it achieves its upper bound value (BSS = 1). Inspection of the Brier terms reveals that the improvement due to assimilation results mostly from the enhanced Brier reliability term shown in Fig. 6(b), while no significant effect on the Brier resolution term [Fig. 6(c)] is seen.

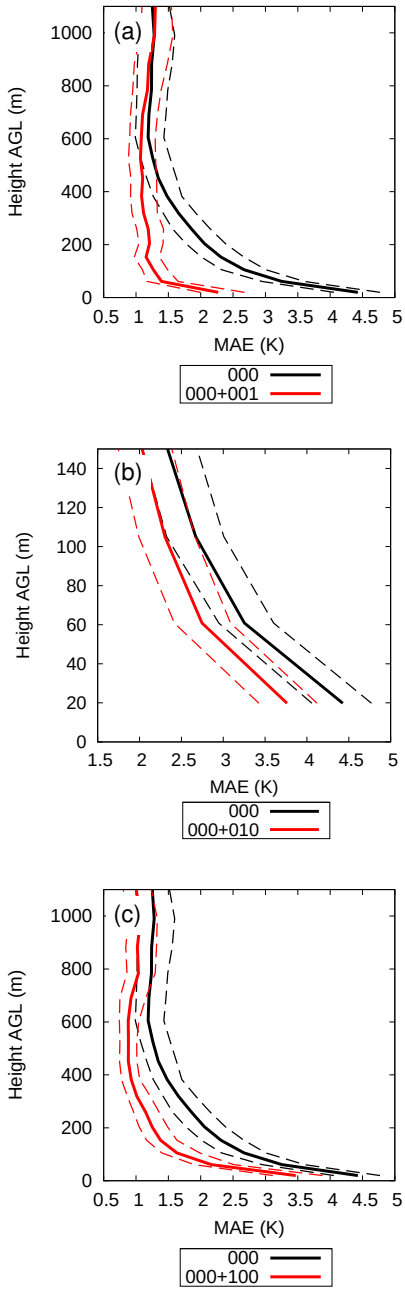


Figure 4: Same as Figure 3 but red lines denote the resulting MAE when the effect of each of the following factors is added to the MAE of the base configuration: (a) pure assimilation (ef_{001}); (b) pure radiation (ef_{100}); (c) pure advection (ef_{101}).

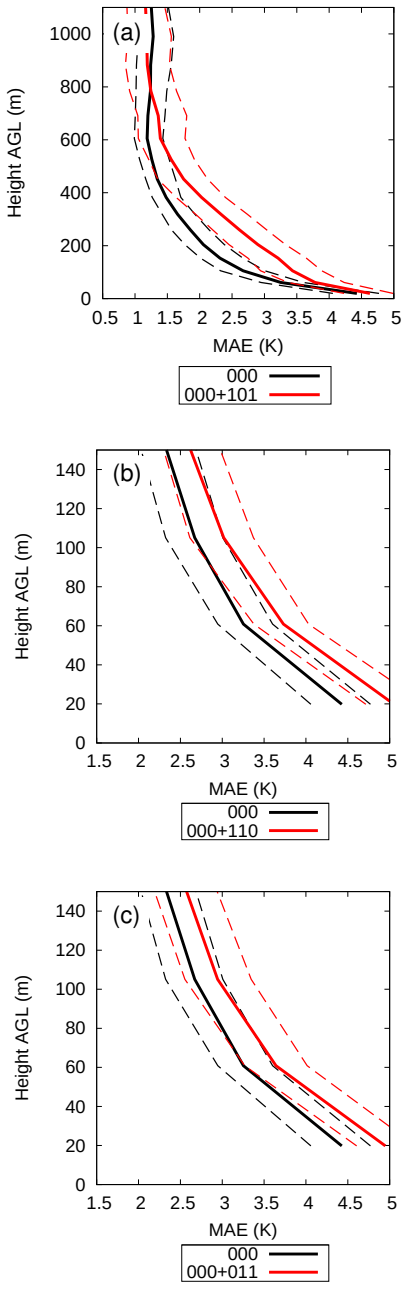


Figure 5: Same as Figure 3 but red lines denote the resulting MAE when the effect of each of the following factors is added to the MAE of the base configuration: (a) assimilation-advection synergism (ef_{101}); (b) advection-radiation synergism (ef_{110}); (c) assimilation-radiation synergism (ef_{011}).

We further inspect the effect of assimilation on the capability of the system to discriminate between groups of events through the calculation of the AUR. The AUR is a positively oriented metric, and values in the range [0.5, 1] implies discrimination. Fig. 7 shows the AUR for 30 min forecast of T profiles valid at 0030 LT for the base configuration (ef_{000}, black lines), and the configuration including only assimilation of surface observations (ef_{001}, red lines). Both the base system, and the system including assimilation, produce discriminating forecasts. This agrees with the Brier resolution term shown in Fig. 6(c), and confirms that the base system provides a useful probabilistic prediction system compared to climatology. Because the base system is nearly optimal in discrimination (AUR=1) the assimilation cannot improve upon it.

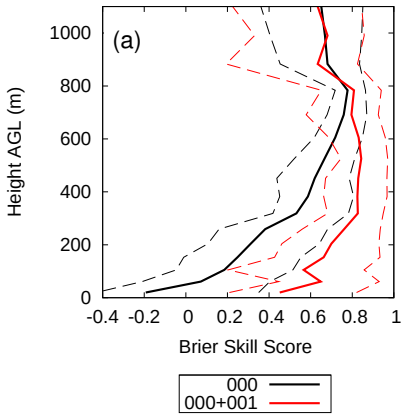
4.2 SCM/EF and climatological dressing (CD) forecasts

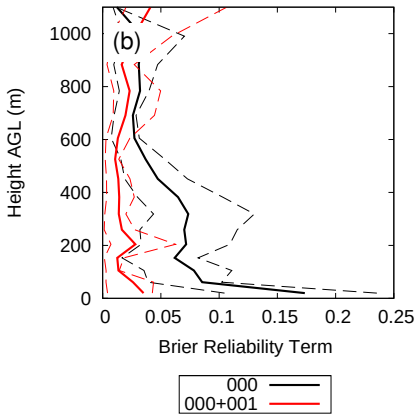
Fig. 8(a) and (b) display MAE of 30-min forecast profiles of θ at 1230 and 0030 LT respectively obtained by: (1) ensemble mean of the SCM/EF in its full configuration (red lines), (2) the CD procedure (blue lines), and (3) the 3D WRF forecasts (4-km grid size) used in the SCM/EF and CD method (black lines). Solid lines denote the original results from our experiments, and dotted lines 90% show CIs.

The performance of the CD procedure is basically dictated by two factors: the climatological surface-atmosphere covariance and the validity of the linear surface-column error relationship assumption (see Section 2.3). Assuming that the linear relationship holds, we expect to obtain improvement in the CD forecasts compared to WRF if the observed climatologies are narrow and WRF climatological covariances reliably represent the observed ones. In such cases the flow-dependent SCM/EF covariances will show little advantage over the climatological ones. The opposite is expected when the flow is characterized by large variability from day to day.

Fig. 8(a) shows that θ forecast profiles calculated using the SCM/EF and CD perform better than WRF at 1230 LT up to 1000 m AGL. The CD forecasts show a slight advantage relative to the SCM/EF. The present CD forecasts benefit from the fact that climatological covariances are calculated over the same period for which the numerical experiments are performed, thus they contain information about the mean flow during the experiment period, a situation not possible in real time forecasts.

At 0030 LT (Fig. 8(b) the SCM/EF profile shows an advantage relative to WRF in the first 200 m AGL but skill becomes poorer above. This detrimental behavior is likely to be a result of small spurious covariances and can possibly be corrected through vertical localization. At 0030 LT the CD forecasts perform significantly poorer relative





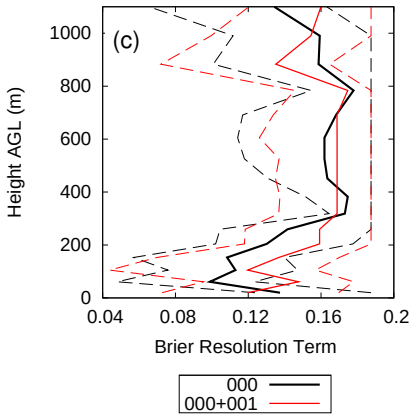


Figure 6: Brier skill score [BSS, panel (a)], Brier reliability term [panel (b)] and Brier resolution term [panel (c)] (for the 75^{th} percentile of the observations) of 30-min forecasts of T profiles valid at 0030 LT for the base configuration (black lines, e_{000}) and for the configuration including assimilation only (red lines, e_{001}). Bold solid lines represent the original scores from our experiments, thin dashed lines are 90% confidence intervals calculated using the BCa bootstrapping technique.

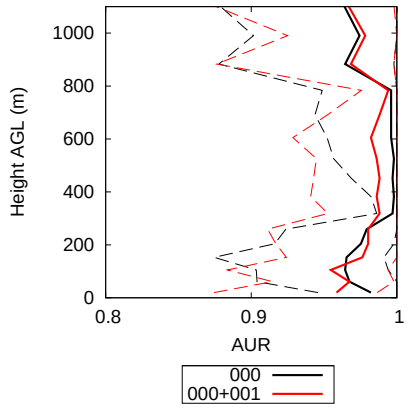


Figure 7: Same as Fig. 6 but for the area under the relative characteristic curve (AUR).

to the SCM/EF and WRF forecasts at all vertical levels. Hacker and Snyder (2005) calculated the standard deviation of the WRF climatology used in the present work (Fig. 1 in their paper). They found that θ is characterized by small variability in the first 1000 m AGL at 1230 LT. However, their results show that at 0030 LT standard deviations are considerably wider than those at 1230 LT, which leads to the advantage of flow-dependent covariances at 0030 LT.

The vertical extent of the superior performance of the SCM/EF and CD θ forecasts as compared to WRF is determined by the magnitude of the surface-atmosphere covariances, which decreases with height. Here we do not verify either the WRF or the CD forecasts probabilistically. As shown, the SCM/EF system provides the advantage of skillful probabilistic forecasts. The deterministic WRF is incapable of this without dressing, and probabilistic verification of the CD predictions is ongoing.

5. Summary

A system based on an SCM of the PBL and on assimilation of surface observations with an EF is used to produce probabilistic nowcasts of PBL profiles wherever surface observations are available. The SCM encompasses several components and we seek to assess the importance of some selected components on the system performance, with the aim of implementing the simplest efficient system with good skill that can be easily deployed, tuned and maintained. The computational cost of these components when running many ensemble members and eventually over many locations can be significant. Three system components were investigated: assimilation of surface observations, horizontal advective tendencies, and parameterized radiation.

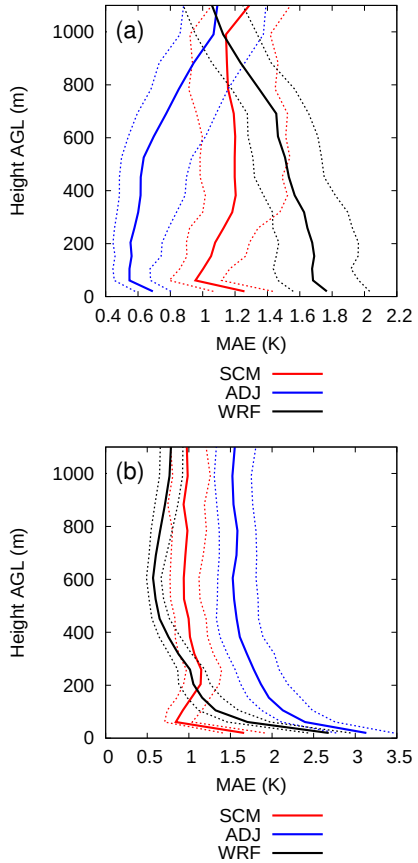


Figure 8: MAE of 30-min forecasts of θ profiles at (a) 1230 LT and (b) 0030 LT. Red lines represent the ensemble-mean forecasts of the SCM/EF system in its full configuration; blue lines represent the forecasts using the climatological dressing (CD) technique; and black lines represent the 3D WRF forecasts (4-km grid size) used in the SCM/EF and CD calculations. Bold lines represent the scores observed in our experiments and thin lines 90% confidence intervals using the bootstrap technique.

The performance of the system was investigated through deterministic and probabilistic verification metrics, inferring statistical significance and flow dependence of the verification scores through the calculation of CIs based on a bootstrapping technique. To the best of our knowledge this is the first work to systematically verify the deterministic and probabilistic skill of SCM-predicted profiles at sites where real surface observations are assimilated with an EF, thus assessing the value of the EF surface assimilation on the forecast of the vertical structure of the PBL.

One of the most important conclusions out of this study is the fact that surface assimilation plays a more significant role in consistently improving ensemble mean and probabilistic skill, over a vast range of weather conditions, than potentially meaningful model enhancements (e.g. parameterized radiation or advection). The effect of the advection tendencies depends on the advective time scales, which are dictated by the characteristic flow at the various atmospheric levels. Advection generally leads to a positive effect as it introduces a one-way forcing from 3D dynamics, but when simultaneously acting with assimilation it may cancel part of the improvement achieved through assimilation. The role of the SCM radiation scheme in improving forecasts performance is minor and in the near vicinity of the surface only.

Flow-dependent covariances estimated with the SCM/EF show a clear advantage over the use of climatological covariances (CD forecasts) when the flow is characterized by wide variability from day to day. In this instance the CD procedure fails to improve the WRF forecasts. The specific implementation of the CD technique led to enhanced results relative to the SCM/EF under a convective PBL regime, which is characterized by narrow variance in the specific geographic location. We recall that the climatological covariances were derived from the same set of WRF forecasts that was used to calculate the CD adjusted profiles, thus providing additional mean flow information not available in a real forecast system (and not available in the SCM/EF calculations).

The improvement observed in the SCM/EF forecast profiles relative to WRF proves that assimilation of surface observations with an EF may be more useful in forecasting the vertical structure of the PBL than full 3D dynamics. Moreover, simple adjustment of the WRF profile by linear regression of the surface forecast errors onto the WRF profile with climatological covariances, i.e., the CD technique, is useful whenever climatology is narrow and a linear surface-column error relationship is valid.

The SCM/EF shows probabilistic skill, thus it provides additional information not available from full deterministic 3D WRF forecasts: an estimate of the flow-dependent uncertainty in the forecast profile at the cost of a few minutes calculation on a personal computer. We expect that

several results will extend to a 3D WRF/EF system assimilating surface observations.

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 M. Pocernich and E. Gilleland for guidance on the R-packages ("verification" and "bootstrap", http://www.r-project.org), which were used for our calculations. We acknowledge T. Hopson and Z. Klausner for insightful discussions on the bootstrap technique. D. Rostkier-Edelstein acknowledges the NCAR 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

Alpert P., M. T. and U. Stein, 1995: Can sensitivity studies yield absolute comparisons for the effect of several processes? *J. Atmos. Sci.*, **52**, 597–601.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

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

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767–1785.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.

Efron, B. and R. J. Tibshirani, 1993: *An introduction to the Bootstrap*. Chapman and Hall, 436 pp.

Ghan, S. J., L. R. Leung, and J. McCaa, 1999: A comparison of three different modeling strategies for evaluating cloud and radiation parameterizations. *Mon. Wea. Rev*, **127**, 1967–1984.

Hacker, J. P. and D. Rostkier-Edelstein, 2007: PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon. Wea. Rev.*, **135**, 2958–2972.

———, 2008: PBL state estimation with surface observations, a column model, and an ensemble filter: probabilistic evaluation under various mesoscale regimes. *Eighty eight Annual Meeting of the American Meteorological Society, New Orleans, LA, January 2008*, available from the AMS web site.

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.

Jolliffe, I. T., 2007: Uncertainty and inference for verification measures. *Wea. Forecasting*, **22**, 637–650.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteorol. Mag.*, **30**, 291–303.

Mason, S. J. and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operative levels. *Wea. Forecasting*, **14**, 713–725.

Mlawer, E. J., S. J. Toubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correletedk model for the long-wave. *J. Geophys, Res.*, **102D**, 16 663–16 682.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.

Rostkier-Edelstein, D. and J. P. Hacker, 2009: On the relative effect of surface observations assimilation in probabilistic nowcasting of PBL profiles (part I and part II). To be submitted.

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. Tech. Rep. TN-468, National Center for Atmospheric Research.

Stein, U. and P. Alpert, 1993: Factor separation in numerical simulations. *J. Atmos. Sci.*, **50**, 2107–2115.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press (San Diego).