# 8.6 PBL STATE ESTIMATES WITH SURFACE OBSERVATIONS, A COLUMN MODEL, AND AN ENSEMBLE FILTER: PROBABILISTIC EVALUATION UNDER VARIOUS MESOSCALE REGIMES

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

DORITA ROSTKIER-EDELSTEIN
*Israel Institute for Biological Research, Ness-Ziona, IL*

## 1. Introduction

A long-term goal of this work is to find an efficient system for probabilistic PBL nowcasting that can be employed wherever surface observations are present. One can also view it as a probabilistic PBL profile retrieval approach, given a background ensemble and surface observations. We are motivated to pursue this by the relative density of high-quality, inexpensive surface observations in local and regional networks worldwide, the fundamental difficulties associated with forecasting details of the PBL at lead times greater than a few hours, and the complexity and expense involved with current NWP models. One approach showing promise is the use of a single column model (SCM) and ensemble filter data assimilation techniques.

Hacker and Rostkier-Edelstein (2007) showed that surface observations can be an important source of information with an SCM and an ensemble filter. 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 that approach observation-error levels in the lowest few-hundred meters of the atmosphere.

Here we extend that work to quantify the probabilistic skill of the same SCM, and the SCM with added complexity. 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 model, with the expectation that it will make solutions more realistic, it is not immediately clear that additional complexity will improve the performance of a PBL nowcasting system based on a simple model. We address this question with regard to treatment of radiation in the column, and also 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. Thus it behooves us to quantify the role of the added complexity in a probabilistic sense.

## 2. Model and Data

The SCM used in this study contains vertical turbulence, atmospheric surface layer, and land-surface parameterizations identical to those in the Advanced Research WRF (ARW) version 2.2. Resolved dynamics are integrated in time with a Crank-Nicholson time step of 10 s on a vertically-stretched column with 61 levels and a model top at approximately 8 km. It is described in detail in Hacker et al. (2007), where the model was forced with prescribed surface radiation and geostrophic winds from a sample of 3D WRF solutions. Here we summarize the components considered for improving probabilistic analyses and nowcasts.

Advection is now an option in the SCM, to allow for the effects of 3D dynamics. Ghan et al. (1999) described 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 an option. These are included particularly to improve simulations during the night, when radiative cooling can be important in the PBL.

Ensemble-filter assimilation of surface observations is the third model component examined here. It is considered alone, as was described in Hacker and Rostkier-Edelstein (2007), and also in combination with the ad-

*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

vection and radiation described above.

The experiment period is 3 May – 15 July 2003, corresponding to the Bow Echo and Mesoscale Convective Vortices Experiment (BAMEX), and the SCM is configured to run over the Atmospheric Radiation Measurement (ARM) Central Facility near Lamont, Oklahoma, which 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. Independent soundings are used for verifying profiles.

Archived runs of the ARW version 2.1, coinciding with the experiment period are used for both ensemble initialization and to provide advection terms in the SCM. Each member of the ensemble is initialized by starting with the 12-h WRF forecast column closest to Lamont for a given day, then perturbing it with the scaled difference between that forecast and another forecast from the archive. The additional member to compute the difference is selected randomly from the experiment period, and the scaling of the difference is drawn randomly from $\mathcal{N}(0,1)$. With this approach, arbitrarily large ensembles can be generated, and we use 100 members in these experiments.

When ensemble filter assimilation is used, all simulations assimilate the surface observations every 30 minutes from 1200 – 1700 UTC, and then the 30-minute forecasts valid at 1730 UTC (1230 LDT) are verified against the surface observations valid at that time.

## 3. Analysis Methods

We seek a probabilistic PBL nowcasting system to use with available surface observations, and thus choose for initial evaluation metrics the mean absolute error (MAE), and the Brier Skill Score (BS; Wilks 1995). The MAE is a deterministic metric used on the ensemble mean, and quantifies systematic error in that mean. The BSS is easily decomposed into both a reliability and resolution term (Murphy 1973) to understand the trade-offs in different components of probabilistic skill. The metrics are examined for all possible combinations of advection, radiation physics, and assimilation of surface observations.

Direct examination in this instance is cumbersome, and we instead seek a framework for simpler interpretation. One framework is factor separation (Stein and Alpert 1993), where our factors are advection, radiation, and data assimilation. Factor separation conveniently quantifies the individual contribution of each model component to reducing the error in the simulations, as well as any beneficial or detrimental interactions between the different factors in the $2^3$ possible model configurations.

Each set of simulations over the BAMEX period can be summarized with an error or skill $e_i$, where $e$ can be one of mean absolute error (MAE), Brier Skill Score (BSS), reliability, or resolution, and the subscript $i$ denotes the binary state of each factor group. For example, $e_{111}$ is the error or skill when advection, radiation, and data assimilation are all switched on. The separated factor associated with each individual component or each combination of components is denoted $f_i$. For example, $f_{111}$ is the contribution to the error that is not simply the linear combination of all the components individually or the effects of combining two of the three factors. The possible combinations of advection, radiation, and data assimilation 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 |

Factor separation equations are derived by a Taylor expansion of the effect of each component, and as shown in Stein and Alpert (1993) the resulting equations for three components are:

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

Manipulation of the equations above shows clearly that the higher-order factors ($f_i$) are the pure interactions between the different possible combinations of factors at lower order, and describe the nonlinear and synergistic effect of putting the components together. Factor separation is helpful in the present experiments because the synergistic effects are not always obvious from direct examination of the errors.

In the next section, we plot factor separation results and interpret them. In each plot, the error itself is shown for the base case (000) and the case when all model components are included (total). On the same plots, the factors $f_i$ are shown for each level of synergism. The analy-

sis here is for surface variables only (observation space), and profiles will be considered in the future.

## 4. Results and Interpretation

### *4.1 Mean absolute error*

The MAE of a perfect forecast verified with error-free observations is zero, and a factor that reduces the MAE will be negative. Figure 1 shows the factor separation for the MAE of surface forecasts. The factor $f_{000}$ (red filled square) corresponds to an MAE that is the same as the factor value. When all the model components are included, the result is a reduced MAE denoted by the red square with a dot, which shows values near the assigned observation error variance. The factors for each level of synergism are shown by the colored **+** symbols.

When model components are invoked individually, the effects are mostly positive (negative contribution to MAE) but some exceptions are apparent. Radiation has a slight worsening effect on $Q_v$ and no effect, as expected, on $U$. In general the effect of a radiation parameterization is expected to be small during this daytime verification, because the 000 case has prescribed radiation at the surface, and direct heating of the column by the parameterized radiation is small. These effects may not be statistically significant (not yet tested).

Advection and assimilation individually result in clear error reduction equal to the single factors shown in the second column of the plots. Assimilation of the surface data has the most positive effect for all variables.

The combined factors require further interpretation. A value of zero for two factors means that the positive effect of each of the two factors combined linearly. A positive value indicates that combination of the two actually degrades the MAE. The black **+** shows that combining advection with assimilation does not continue to improve the forecasts, and rather that some of the benefit of the assimilation is lost as the advection sweeps out the influence of the observations during the 30-minute forecast period. Another way to see this is to consider that if the advection and assimilation reduced errors when they were combined in the same way that they did when they were alone, the total MAE reduction would be the sum of the values denoted by the + and +, or approximately -2.4°C for temperature. The synergistic effect of the two factors cancels approximately 1.2°C of that expected reduction, and the error would then be between that of the two factors used independently. The assimilation is clearly the most important factor, and relaxation of the ensemble mean to a 3D WRF solution via advection erases some of the benefit.

Radiation, when combined with the assimilation (+), appears to have an additional positive effect for $Q_v$ and a negligible effect for $U$. For $T$, the positive effect of the radiation in isolation is almost perfectly canceled by the assimilation, which apparently overwhelms it. Because the absolute value of the advection-radiation synergism (+) approximately the same as the absolute value of the improvement by radiation (+), the benefit of radiation is also entirely overwhelmed by advection of $T$. But the slight degradation in MAE of $Q_v$, caused by radiation, is partially eliminated by advection. These effects are generally small and may not be statistically significant.

The effect of combining all three factors is generally neutral, as shown by the + symbols. The damage caused by the advection remains. We conclude that advection has a negative affect on the the MAE at the surface when assimilating surface data. It is caused by advecting imperfect atmospheric conditions, from a 3D forecast, into the column. Ensemble filter assimilation of the surface observations corrects for some of the error. A 30-minute forecast with a simple model assimilating surface observations with an ensemble filter is superior to an 18-h WRF forecast, from which advection is derived. This result is not likely to generalize to shorter 3D forecasts.

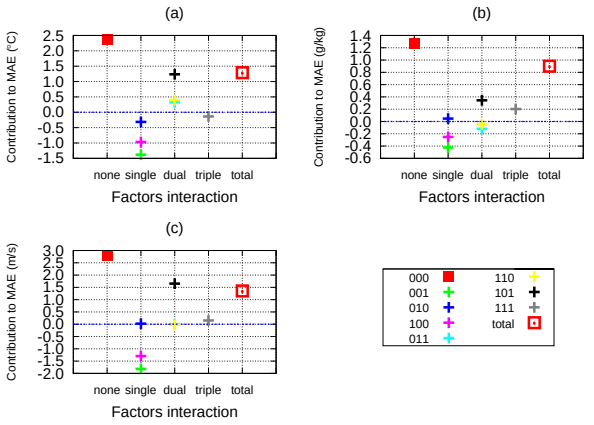


Figure 1: Factors contributing to MAE in surface forecasts of (a) $T$, (b) $Q_v$, and (c) $U$. The MAE of the basic model configuration (000) and the complete model configuration (total) are denoted by the red squares. The factors associated with each combination of model components, showing the synergistic contribution to the MAE corresponding to the model configurations in Table 1, are shown by the colored **+** symbols.

### *4.2 Brier skill score*

The BSS is a convenient metric for probabilistic verification because it can be easily decomposed into reliability and resolution terms, thereby enabling meaningful

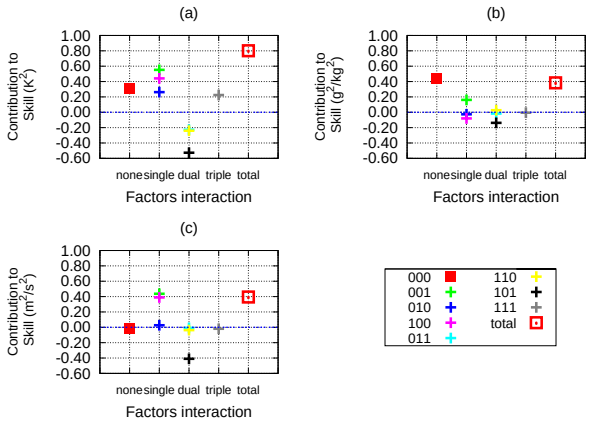


Figure 2: Factors contributing to the Brier Skill Score for the $75^{th}$ percentile of the observations.

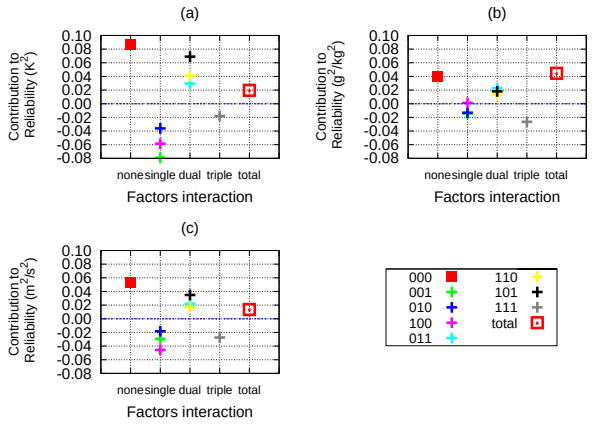


Figure 3: Factors contributing to the reliability term of the Brier Score for the $75^{th}$ percentile of the observations.

interpretation. It verifies events, and 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 BSS is positively oriented so that higher values indicate more skill, a perfect forecasting system gives BSS = 1, and BSS $\leq$ 0 indicates no skill or less skill than the reference. A reference forecast is needed for a skill score, and we use the mean of the sample of observations from the ARM site during the experiment period.

BSS factors show qualitatively similar overall impacts (reversed in sign because of the score orientation) as they do on the MAE (Fig. 2) for $T$ and $U$, but different impacts for $Q_v$. Probabilistic skill as measured by the BSS is hardly affected by any factor except the assimilation individually (+). The advection destroys the positive impact, resulting in an overall worse BSS when all components are included (red square with dot). Again statistical significance tests are needed.

Examination of the reliability and resolution terms individually enable further interpretation. The result is:

BSS = (resolution-reliability)/uncertainty.

Uncertainty here is purely a function of the distribution of observations and is therefore a constant. A system with no skill can occur if the resolution is less than or equal to the reliability, and the resolution is equal to the sum of the reliability and the uncertainty in a perfect system. In general, the BSS is improved by maximizing the resolution (upper bound reliability plus uncertainty) and minimizing the reliability term (lower bound 0). It is important to remember that resolution and reliability are independent because they result from different distributions.

Differences between the factors for reliability (Fig. 3) and resolution (Fig. 4) show the benefit of BSS decomposition. In several instances the reliability and resolution are both improved, with trends qualitatively similar to the MAE and BSS. In particular, individual components lead to improvement in both $T$ resolution and reliability, but the synergistic effects of two factors are all detrimental. This result is consistent with both the MAE and obviously the BSS factors. The $Q_v$ factors are small, but also broadly consistent with the earlier analysis.

Factors for $U$ resolution and reliability show opposing affects in some cases. Reliability is improved for all single components (Fig. 3c), but resolution deteriorates slightly when parameterized radiation is used (+ in Fig. 4c). The BSS and MAE show negligible change, which is perhaps the expected result, but the BSS does not change because the improvement in reliability is offset by a reduction in resolution. One possible explanation is that bias is reduced, which improves the reliability but also shifts the forecast distribution toward climatology; climatology has by definition no resolution.

Dual factors show that synergistic effects are all detrimental to reliability, but can improve resolution. Radiation combined with assimilation (+ in Fig. 4c) provides added benefit. At this point we can only hypothesize an explanation, but one possibility is that the parameterized radiation improves forward operators in the assimilation of, say, $T$, which then has a positive impact on the $U$ state through multivariate covariances.

The results of this section are summarized in Tables 2 ($T$), 3 ($Q_v$), and 4 ($U$), where a '+' indicates an improvement and a '-' indicates a deterioration. A missing entry denotes a factor that is less than 10% of the metric before additional components are added (red filled squares in figures). A 10% threshold is a proxy for real signifi-

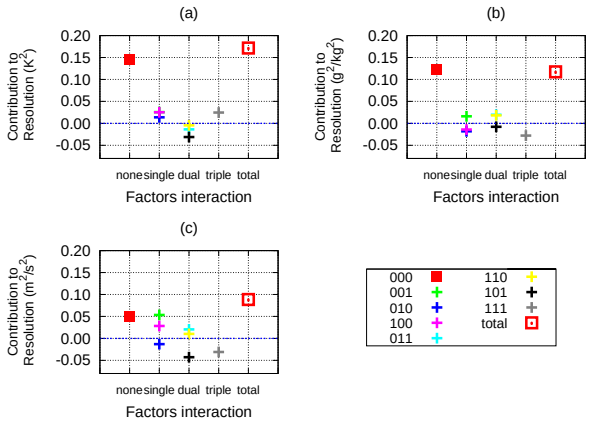


Figure 4: Factors contributing to the resolution term of the Brier Score for the $75^{th}$ percentile of the observations.

Table 2: Summary of BSS factor separation results for temperature. The '+' denotes improvement and the '-' denotes deterioration. A blank entry means the factors is less than 10% of the total metric before improvement.

| Label | Skill | Resolution | Reliability |
|---|---|---|---|
| 100 | + | + | + |
| 010 | + | | + |
| 001 | + | + | + |
| 110 | - | | - |
| 101 | - | - | - |
| 011 | - | | - |
| 111 | + | + | + |

Table 3: Same as table 2 but for mixing ratio.

| Label | Skill | Resolution | Reliability |
|---|---|---|---|
| 100 | - | - | |
| 010 | | - | + |
| 001 | + | + | + |
| 110 | | + | - |
| 101 | - | | - |
| 011 | | + | - |
| 111 | | - | + |

Table 4: Same as table 2 but for $U$-wind.

| Label | Skill | Resolution | Reliability |
|---|---|---|---|
| 100 | + | + | + |
| 010 | | - | + |
| 001 | + | + | + |
| 110 | | + | - |
| 101 | + | + | - |
| 011 | + | - | |
| 111 | | - | + |

cance tests, which we have not yet completed.

## 5. Summary

A primary goal of this work is to develop an efficient and skillful probabilistic nowcast system for local PBL profiles. Efficiency includes not only computational cost, but also the cost of maintenance and re-tuning when deployed in new environments. Thus efficiency can often be maximized with a simple system, and we seek the simplest system that can provide a high level of probabilistic skill. A simple system is also easier to understand, may be used in process studies, and exposes errors more easily.

An ensemble of simple single column model (SCM) realizations is the basic system. The basic model uses land-surface, surface-layer, and PBL parameterizations from the WRF model, and it is forced with prescribed surface radiation and geostrophic winds drawn from a history of 3D WRF forecasts during May-June 2003. We consider the change in deterministic and probabilistic skill from the addition of surface data assimilation, parameterized radiation, and advection with tendencies from the same 3D WRF forecasts. Factor separation is a convenient way to evaluate the $2^3$ possible configurations with these three additions. It quantifies the synergistic effects of multiple model components, which are non-zero when the effect of the components do not combine linearly. Metrics include mean absolute error (MAE), the Brier Skill Score (BSS), and the reliability and resolution terms from the decomposed BSS. Here we examine 30-minute forecasts of surface temperature, water vapor mixing ratio, and winds valid at 1730 UTC (1230 LDT) during the experiment period.

Results show that the added complexity of including all three new model components improves the forecasts under all skill metrics. The factor separation shows similar synergistic effects for many of the combinations and metrics, suggesting few situations where one skill metric improves and another deteriorates under a specific model configuration.

Factor separation shows that assimilation of surface observations is the most important contributor to improved skill. Advection can also substantially improve skill when assimilation is not included, but when it is combined with other components it dominates to cancel gains provided by the other individual factors. Most importantly, it eliminates a lot of the improvement from the assimilation by rapidly sweeping it out of the column.

Advection relaxes the SCM toward the WRF solution for the same time period, leading to behavior that is fundamentally different from a 3D forecast and assimilation system. In a 3D forecast and assimilation system the surrounding grid points are updated by an observation. For

a time period, defined by the advective time scale and the spatial scale of the analysis increment, the advection from those surrounding grid points accounts for that observation. In our case that time is only the time it takes for the atmosphere to advect one grid cell in the forcing model (4 km here). Ongoing work addresses this deficiency.

Including parameterized radiation, which adds significant complexity to the model, has little effect on the skill. This may change during the night, when radiative cooling near the surface can be important.

Some of the results show a clear and repeatable trend, but the smaller effects should be viewed with the caveat that they may not be statistically significant. That work is ongoing.

**Acknowledgement** We are grateful to the Data Assimilation Research Testbed development group at NCAR for providing useful software and supporting the SCM within the infrastructure. Matt Pocernich was a great help in providing code for the skill assessment. Observation 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

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.

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., J. L. Anderson, and M. Pagowski, 2007: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Wea. Rev.*, **135**, 1021–1036.

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

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 correleted-k 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.

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).