# Assimilation of lidar planetary boundary layer height observations.

**Andrew Tangborn[1], Belay Demoz[1,2], Brian J. Carroll[2], Joseph Santanello[3]and Jeffrey L. Anderson[4]**

[1]Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD, USA

[2]Dept. of Physics, University of Maryland Baltimore County, Baltimore, MD, USA

[3]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

[4]National Center for Atmospheric Research, Boulder, CO, USA

Corresponding author: Andrew Tangborn, tangborn@umbc.edu

## Abstract

Lidar backscatter and wind retrievals of the planetary boundary layer height (PBLH) are assimilated into 22 hourly forecasts from the NASA Unified - Weather and Research Forecast (NU-WRF) model during the Plains Elevated Convection Convection at Night (PECAN) campaign on July 11, 2015 in Greensburg, Kansas, using error statistics collected from the model profiles to compute the necessary covariance matrices. Two separate forecast runs using different PBL physics schemes were employed, and comparisons with 6 independent radiosonde profiles were made for each run. Both of the forecast runs accurately predicted the PBLH and the state variable profiles within the planetary boundary layer during the early morning, and the assimilation had a small impact during this time. In the late afternoon, the forecast runs showed decreased accuracy as the convective boundary layer developed. However, assimilation of the doppler lidar PBLH observations were found to improve the temperature and V velocity profiles relative to independent radiosonde profiles. Water vapor was over corrected, leading to an increased differences with independent data. Errors in the U velocity were made slightly larger. The computed forecast error covariances between the PBLH and state variables were found to rise in the late afternoon, leading to the larger improvements in the afternoon. This work represents the first effort to assimilate PBLH into forecast states using ensemble methods.


## 1 Introduction

The planetary boundary layer (PBL) plays an important role in both weather and climate. This layer is where the Earth's surface interacts with the atmosphere, exchanging heat, moisture and pollutants. The PBL height (PBLH) is central to these interactions and is controlled by the energy flux from the surface. Under certain conditions during daytime it defines the convective boundary layer (CBL) and during nighttime it is the stable (non-convective) boundary layer (SBL). Trace gases and aerosols emitted from the surface are rapidly transported within this layer by turbulent atmospheric motion, and transfer of energy and mass into the free troposphere occurs across an interfacial layer at the top of the PBL. The PBLH is fundamental to weather, climate, atmospheric turbulence and pollution through its role in land-atmosphere interactions and mediation of Earth's water and energy cycles (Santanello et al. 2018). It affects convection in the troposphere, which is generally initiated within the boundary layer and then penetrates

the top (Hong and Pan, 1998; Browning, et al. 2007). Thus, accurate knowledge of the PBLH is essential for both weather, pollution and climate forecasting.

The PBLH is defined by thermodynamic properties such as a temperature inversion or hydrolapse which can be measured by radiosonde. Alternatively, the drop off in aerosol concentration that occurs across the top of the PBL is used, since aerosols are well mixed throughout the PBL (Hicks, et al., 2019). Atmospheric models rely on parameterization schemes to define the structure of the PBL and compute PBLH. These are generally either local mixing schemes that use local turbulent kinetic energy (TKE, Janjic, 1994) or non-local flux schemes (Hong and Pan, 1996). Generally, these PBL parameterizations have systematically higher PBLH relative to observed values (Hegarty et al., 2018), and also have difficulties modeling the growth of the convective layer during the morning. The variety of definitions of PBLH make it difficult to effectively evaluate existing models or develop new ones.

Observations of PBLH are traditionally made by radiosonde measurements, which have high vertical resolution but are expensive to launch frequently and are thus limited to special experiments and/or ill-timed launches (*e.g.* 00/12 UTC National Weather Service launches) with respect to convective and stable PBL development. Likewise, spaceborne measurements of the lower troposphere from passive and active instruments are severely limited in vertical, spatial, and/or temporal resolution (Wulfmeyer et al. 2015). Ground based measurement of PBLH has been proposed for an extensive network of ceilometers by adding to the functionality of instruments that were designed for measuring cloud heights (Hicks et al., 2016). The ceilometer measures the time required for a laser pulse to return to a receiver, from which the height of the scattering is determined. The intensity of the backscatter is correlated with the density of aerosols at a given height and the PBLH is inferred from the location of the maximum negative gradient of the backscatter intensity. Several algorithms employ wavelet transforms to identify the location of the negative gradient (e.g. Brooks, 2003; Knepp, *et al.*, 2017). This existing network of ceilometers could be used to create a relatively dense network of frequent PBLH observations, as was recommended by the 2009 study from the National Research Council (NRC, 2009) and the Thermodynamic Profiling Technologies Workshop (NCAR, 2012).

Since the ceilometer PBLH observations were not yet available for the campaign we are using, we employ doppler lidar observations made at the PECAN site in Greens-

burg, Kansas, to demonstrate the methodology. The data is from a Leosphere WINDCUBE-200S Doppler lidar owned and operated by the University of Maryland, Baltimore County (Delgado et al., 2016). This lidar operates at an infrared wavelength, and hence receives its strongest backscattered signal within the aerosol-laden PBL and is often below the measurement noise floor above the PBL. The Doppler shift of the backscattered signal is used to calculate wind speed as a function of range, which can then be used to produce a multitude of wind and turbulence variables useful for PBL characterization (e.g. vertical velocity variance and signal-to-noise ratio variance). While Doppler lidars and ceilometers are similar in aerosol detection, a Doppler lidar's additional wind measurement capability makes it more broadly applicable and at times more accurate than a ceilometer for PBLH measurement. The PBLH algorithm applied for this study combines several such aerosol and wind variables for PBLH measurement. Each PBLH retrieval involves measurement of turbulence intensity, horizontal wind profiles and backscatter intensity. The heights of steep gradients in these quantities are determined using empirical thresholds and wavelet transform techniques, and the three estimates are combined using fuzzy logic. This is described at length in Bonin et al. (2018). Additional lidar parameters and the application of the algorithm to PECAN data were presented in Carroll et al. (2019). The PBLH measurements were made from a repeating 25-minute lidar scan cycle. This Doppler lidar and PBLH algorithm combination are generally well-suited for accurate and precise measurement of the PBLH during the daytime boundary layer, nocturnal boundary layer, and morning transition period (Bonin et al. 2018, Carroll et al. 2019). The evening transition is the most challenging for this setup due to due to difficulties in defining a clear mixing layer during the decay of a turbulent daytime PBL (Lothon et al. 2014).

The question remaining is how to assimilate these observations into a numerical weather prediction (NWP) model. A number of studies have explored assimilating thermodynamic profile measurements from lidar (Hu et al. 2019, Coniglio et al. 2019, Degelia et al. 2019) and have shown that this increases the accuracy of model PBLH estimates. But we are interested assimilating the PBLH observations directly because the ceilometer network described above will focus on these retrievals, and satellite missions which measure PBLH are also planned. PBLH is a diagnostic variable in NWP parameterized physics models. This means any correction to PBLH will be lost during the model forecast unless the PBLH height observation is used to correct state variables such as tem-

perature and moisture. This could be done either by creating an adjoint of the PBL parameterization scheme, or through the use of an ensemble Kalman filter which would determine the error covariances between PBLH and state variables in the model. The structure of the covariance, and how the state variables are changed by assimilating PBLH, will depend on which PBL scheme is used. We will show how such a system could work by conducting a posteriori lidar PBLH observation impact experiments using forecast fields from a NASA Unified - Weather and Research Forecast (NU-WRF, Lidard-Peters, 2015) model runs for one day during the Plains Elevated Convection at Night (PECAN) campaign on July 11, 2015. The assimilation is done on 22 hourly WRF forecast fields throughout the day without cycling the analysis fields back into the model, using two different PBL parameterizations. In this paper, we demonstrate a new and promising method that uses the lidar-based aerosol backscatter and wind derived PBLH to correct model forecasted state variables. The purpose here is to show how ensemble computed error covariance can transfer observational information from PBLH to the state variable profiles.

## 2 Methodology

The assimilation methodology is based on the ensemble Kalman filter (EnKF)(Evensen, 2009), where the analysis state is the estimate with the minimum estimated errors, relative to the given error statistics. It differs from the EnKF in that the analysis is not used as an initial state for the next model forecast. Rather, two existing one day NU-WRF forecasts, with different PBL physics schemes, are used when lidar measurements are available at a single location. These forecasts were produced as a part of the PECAN campaign in 2015, and we resuse them here to demonstrate the assimilation algorithm that we have developed. These were not ensemble forecasts so we cannot build a standard ensemble Kalman filter from them. Instead we use Ensemble Optimal Interpolation (EnOI), in which profiles from neighboring model gridpoints are used to obtain an estimate of error statistics (Oke, *et al.*, 2010; Keppenne, *et al.*, 2014). This approach will allow for the construction of the vertical component of covariance, which is needed in order to understand how PBLH can be used to correct atmospheric profiles through the use of profile and PBLH statistics. We use profiles from nearby model grid points and have tested the system with varying numbers of grid points in the ensemble. An ensemble Kalman filter would likely give different covariance information, but the basic rela-

tionship between the state variable profiles and the PBLH are determined by the model in the same manner here.

The two NU-WRF simulations use the Mellor–Yamada–Janjic (MYJ)[Mellor and Yamada, 1974, 1982; Janjic, 2002] and Mellor-Yamada-Nakanishi-Niino level 2.5 (MYNN) [Nakanishi and Niino, 2009] which are local 1.5 and 2.5 order turbulence closure schemes respectively. The PBLH in each of these models is estimated using the total kinetic energy (TKE) method. The NU-WRF forecast state variables are temperature (T), moisture (Q) and velocity (U,V), and we define the forecast vector $\mathbf{x}^f = [T^f\ Q^f\ U^f\ V^f\ (PBLH)^f\ ]$, where we have combined PBLH with the state variables to enable the covariance calculation between them. The forecast runs are initiated from the NOAA global forecast system (GFS) reanalysis interpolated to the local domain of 30-48N and 84-110 W, with 220×220 lat/lon and 54 vertical levels, at 0 UTC. At this time, the initial state has assimilated all of the convential and satellite observations globally. The two WRF forecast experiments start at 0 UTC, and are run for 22 and 23 hours for the MYJ and MYNN experiments, respectively. We use an ensemble of the 20×20 nearest gridpoints, so that all of the ensemble members are within about 30 km of the lidar observations (since the grid spacing is about 3 km). Generally, larger ensembles using gridpoints farther away will result in larger forecast error covariance because the geographic variability. So this ensemble size was chosen as a balance between ensemble size and geographic localization. The forecast standard deviation for PBLH on the chosen ensemble was around 27 m at 22 UTC.

The forecast error covariance, $\mathbf{P}^f$ is defined as

$$\mathbf{P}^f = \left\langle (\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^T \right\rangle \tag{1}$$

where the summation is over the grid points $i = 1, N_{lon}$, $j = 1, N_{lat}$ and $\mathbf{x}^t$ is the (unknown) true state, on the discrete model grid. We only assimilate the observation $y^o = PBLH = H(\mathbf{x}^f)$ where $H$ is the non-linear observation operator. The analysis equation is

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(y^o - H(\mathbf{x}^f)) \tag{2}$$

where the gain matrix, $\mathbf{K}$ is defined by:

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + (\sigma^o)^2)^{-1}, \tag{3}$$

166 $\sigma^o$ is the observation error standard deviation supplied with the lidar retrievals, and is
167 determined from the combined uncertainty of the vertical velocity variance, velocity gra-
168 dient and backscatter gradient. Generally, when these quantities change rapidly at the
169 top of the PBL, then the estimated error is small. The error estimates are larger when
170 (during the evening), the gradients are much more gradual. **H** is the linearized obser-
171 vation operator for PBLH. Because the PBLH is related to the state variables via the
172 two PBL physics schemes, determining **H** would require linearizing the PBL physics at
173 every analysis time. Rather, here we use the EnOI described above to get:

$$\mathbf{P}^f\mathbf{H}^T \approx \left\langle (\mathbf{x}^f - {\mu_\mathbf{x}}^f)\ (H(\mathbf{x}^f - {\mu_\mathbf{x}}^f))^T \right\rangle \tag{4}$$

174 and

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T \approx \left\langle H(\mathbf{x}^f - {\mu_\mathbf{x}}^f)\ (H(\mathbf{x}^f - {\mu_\mathbf{x}}^f))^T \right\rangle \tag{5}$$

175 where ${\mu_\mathbf{x}}^f$ is the mean forecast state of the ensemble of profiles. See Houtekamer and
176 Zhang (2016) for a review of ensemble Kalman filter techniques.

177 We expect the correlation between the airmass within the PBL and the free tro-
178 posphere to drop away rapidly, because of limited intereactions between them. We found
179 that this can cause errors in the analysis profiles if error covariance and PBLH is allowed
180 to continue into the troposphere. To reduce these errors we have added an exponential
181 decay starting at the model level closest to the PBLH ($k_{PBLH}$) to define a vertical lo-
182 calization factor:

$$C_{loc} = exp\left[-\alpha(\frac{k - k_{PBLH}}{k_{PBLH}})^2\right] \tag{6}$$

183 where $k$ is the model level and $\alpha = 8$ is an experimentally determined factor. This en-
184 sures that the covariance between the PBLH and the state variables becomes small within
185 a couple of model levels into the free troposphere.

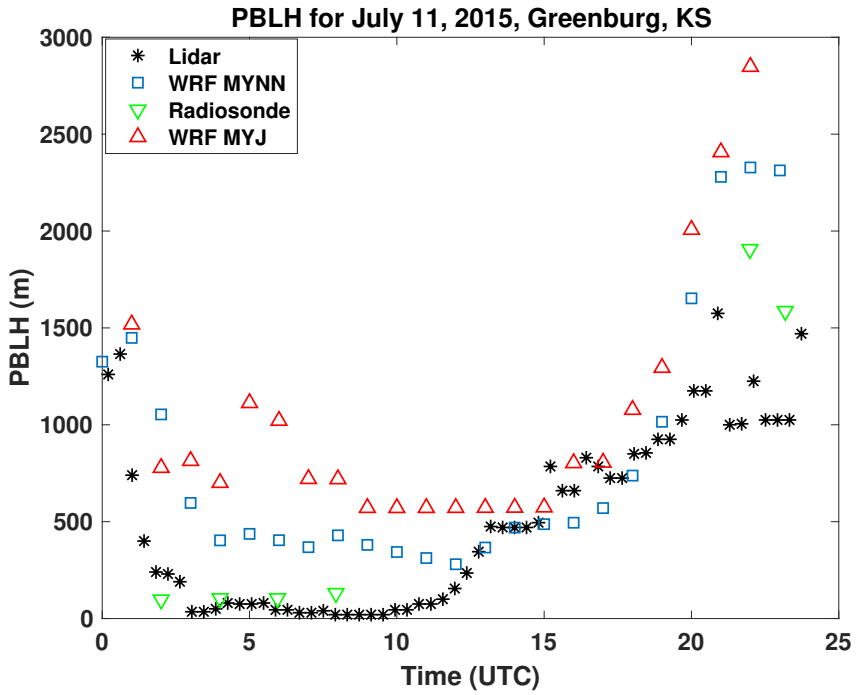
186 This system is solved at each hour using the nearest lidar profile observation in time,
187 and the resulting analysis fields are compared to radiosonde profiles when the latter are
188 also available. There are 22 or 23 analyses (for each forecast run), and 6 times where com-
189 parison with radiosonde profiles are made. We focus on the impact of the assimilation
190 on the state variables T, Q, U and V rather than the PBLH because only the state vari-
191 ables would be retained by a forecast.

## 3 Results

The NU-WRF simulations, taken from existing forecast runs used for the PECAN campaign (Santanello *et al.*, 2019) are initialized using a National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) reanalysis. The two forecast runs were conducted using MYJ PBL physics (2-22 UTC) and MYNN (2-23 UTC) on July 11, 2015. Lidar PBLH observations were made every 25 minutes on that day in Greensburg, KS (37.6 N, 99.3 W), while balloon soundings were launched from that location 6 times as part of the Plains Elevated Convection At Night (PECAN; Gerts et al. 2017). Figure 1 shows the PBLH during that day, derived from the two NU-WRF forecasts, lidar observations and soundings. We have determined the sounding PBLH using the parcel method (Holzworth, 1964), which defines the top as the height where the potential temperature first exceeds the ground temperature. The lidar PBLH (black *, derived using the method reported in Bonin, 2018) closely matches the radiosonde estimates (green triangles) in the late evening to early morning (2-7 UTC), while it is somewhat lower in the afternoon. The two NU-WRF forecasts differ from the observations depending on the time of day. In the early morning and early afternoon the MYJ forecasts (red triangles) both are higher than the observations, then rise less than the lidar observations in the late morning and early afternoon (12-17 UTC, there are no radiosonde measurements to compare to here) before rising much higher than the observations in the late afternoon (18-24 UTC).

Since we are primarily interested in the impact of the assimilation on state variables within the boundary layer, in Figures 2 and 3 we plot the RMS difference between the model and the independent (unassimilated) radiosonde profiles from the surface to roughly the top of the boundary layer in the late afternoon. This corresponds to the first 8 layers, or about 800 mb. We use a fixed number of layers so as to make the comparisons of the RMS differences consistent during the day, rather than computing the RMS over a different number of layers as the PBL grows during the day. For the temperature forecast, the RMS difference would is

$$RMS(t_a) = \left[ \frac{1}{8} \sum_{i=1}^{8} (T_i^f - T_i^{sonde})^2 \right]^{1/2} \tag{7}$$




c

**Figure 1.** PBLH vs UTC time for July 11, 2015 for lidar backscatter (black *), WRF model - MYJ (red triangles), WRF model - MYNN (blue squares), and radiosonde observations using parcel method (green triangles).

where $t_a$ is the analysis time and $i = 8$ is the model level at the top of the PBL in the late afternoon. Figures 2 and 3 show the RMS differences with the radiosonde profiles throughout the day for the forecasts (blue x) and analyses (red squares) for potential temperature (upper left), water vapor mixing ratio WV (upper right) and the U (lower left) and V (lower right) components of velocity.

During the night (2-9 UTC), the assimilation has a relatively smaller impact on the potential temperature RMS differences (upper left) in the early morning (6 and 8 UTC), and the two forecasts have similar accuracy. By late afternoon (22 and 23 UTC, note that the MYJ forecast stops at 22 UTC) the radiosonde comparisons show that the assimilation reduces RMS differences in the potential temperatures by around $1.5K$ for MYJ and $2K$ for MYNN. The water vapor mixing ratio (upper right) also has little impact from the assimilation between 2 and 8 UTC, but at 22 UTC (the next radiosonde profile) the RMS difference for both MYJ and MYNN analysis increase by at least $1.5\times 10^{-3}kg/kg$ in the late afternoon. The U-velocity profiles (lower right) show small differences between the MYJ and MYNN through 8 UTC (3 a.m. local time) and the assimilation increases the RMS differences with radiosonde profiles by nearly $1m/2$ starting at 22 UTC for both models. The V-velocity profiles (d) begin to differ between MYJ and MYNN for the forecasts at 8 UTC ($0.5m/2$ decrease), and assimilation also decreases the RMS differences with radiosondes in late afternoon by $1.5 - 2m/s$.

We would like to understand why there is a smaller impact during night time and early morning, whereas there are decreases inthe RMS differences in temperature and V velocity and increases in moisture and U velocity in the late afternoon. To this end, we plot the forecast, analysis and radiosonde profiles (T, Q, U and V) at 4 UTC (11 p.m. local time) and 22 UTC (5 p.m. local time) in Figures 4-7. At 4 UTC, (Figures 4,5) these clearly indicate that there are small corrections made by the assimilation, as the red and blue profiles closely overlap. But it also shows that the profiles (particularly temperature and moisture) more accurately follow the radiosonde profiles (except for the U velocity above the PBL), meaning that there is less room for improvement to the forecast state. In contrast, Figure (1) shows that the forecast PBLH (particulary MYJ) is quite a bit higher than the lidar observation at 4 UTC. In the late afternoon (Figures 6, 7) indicate that there are large differences between the forecast and radiosonde profiles for

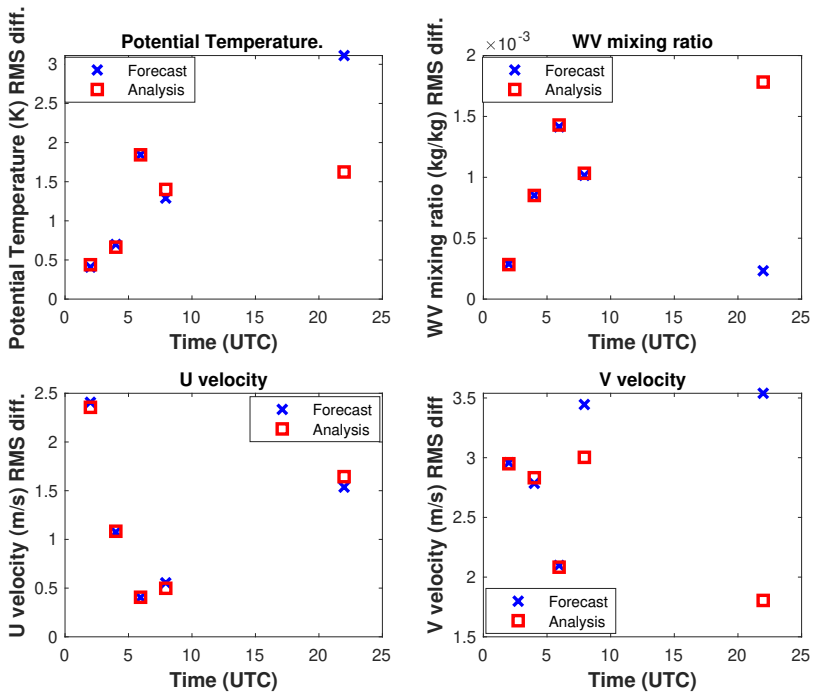



**Figure 2.** RMS difference for lowest 8 layers, vs. time of forecast (blue x) and analysis (red square) with radiosonde profiles for potential temperature (upper left), water vapor (upper right), U velocity (lower left) and V velocity (lower right).

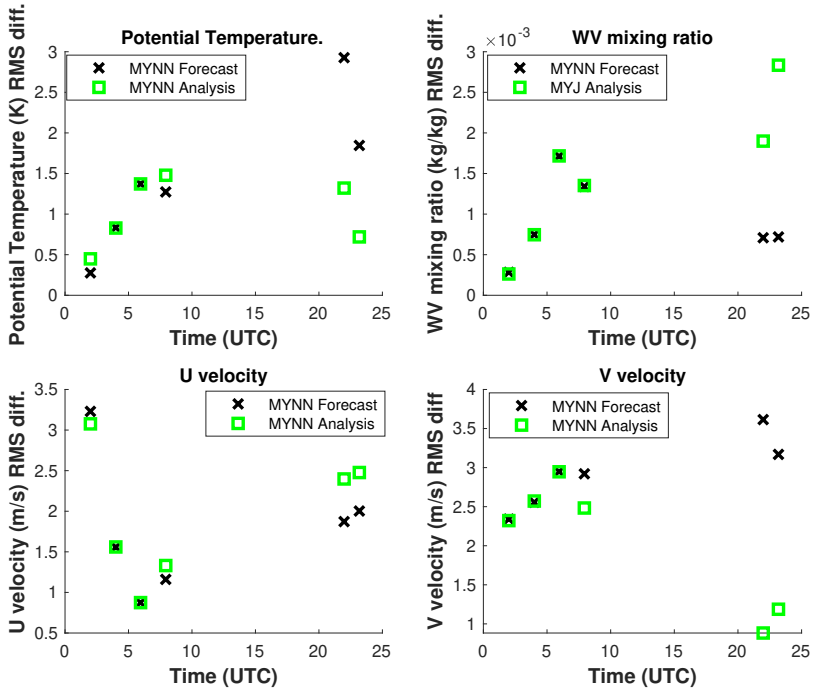



**Figure 3.** Same as Figure 2, but for MYNN PBL model, with forecast (black x) and analysis (blue square).

all of the state variables, and the forecast PBLH values differ substantially from the lidar measurements as well. The correction to to the forecast profiles generally pushes the analyses towards the independent radiosonde profiles, particularly for temperature and V velocity.

So the forecasts that that predicted both PBLH and state variables with relatively greater accuracy in the early morning were not corrected, while the less accurate afternoon forecast was drawn towards the independent radiosonde measurements. The assimilation also made changes to the vertical velocity (W) in the afternoon, but there is no independent data to compare with so we have not included it.

The WV profile is shown to be increased by the assimilation (since WV and PBLH are negatively correlated and higher PBLH corresponds to lower WV levels in the PBL models), but the analysis overshoots the radiosonde WV profile, hence causing the increase in the water vapor RMS difference in Figures 2 and 3. Compared to temperature, WV is highly variable in time and space and it has been shown in the past that slanted balloon trajectories under estimate the WV present (Demoz et al 2006; Crook, 1996). The U velocity difference with the radiosonde is larger for the analysis, but this correction is more difficult because the differences (at least for MYJ) are both positive and negative and the PBLH observation only contains a single piece of information. The V velocity is, on the other hand, greatly improved by the assimilation. These analysis profiles in show that, for this one analysis time, the assimilation is pushing the state variables in the proper direction for temperature, V velocity and moisture, though the moisture correction overshoots the readiosonde profile. The reason for these corrections to the state variable profiles is that the error covariance between PBLH and each state variable, $\mathbf{P}^f\mathbf{H}^T$, can be computed from the ensemble of profiles that was collected from the model grid. The forecast PBLH for each profile was computed using the full PBL physics, and therefore contains the essential correlation information between these variables.

The increasing differences between PBLH and profile forecasts from early morning to late afternoon only partly explain the much larger impact of the assimilation at 22 UTC. We can also analyze this by plotting the error covariance between PBLH and each of the state variables, seen in Figure 8 at different times during the day. The covariance with temperature is always positive, and grows by a factor of 4 by late after-

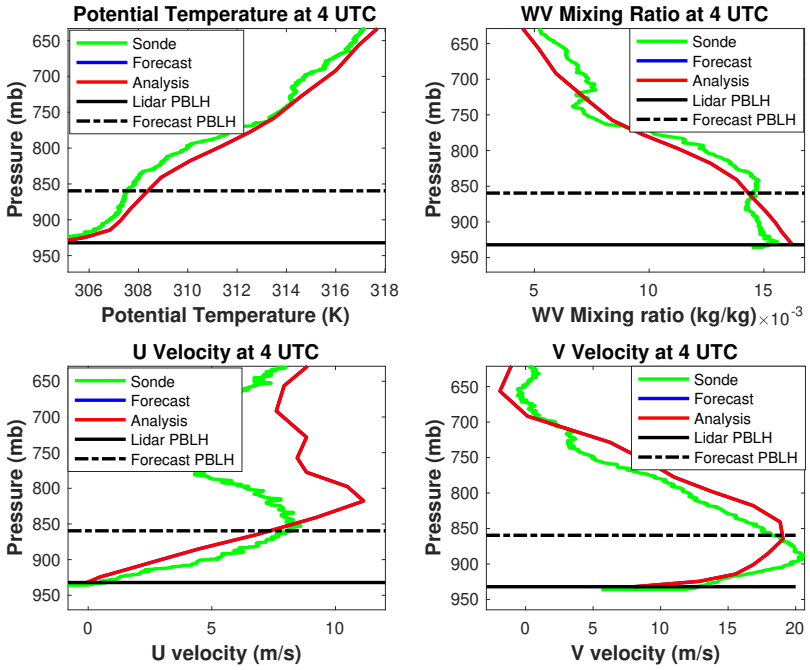



**Figure 4.** Profiles from radiosonde (green), forecast (blue) and analysis (red) for potential temperature (upper left), water vapor mixing ratio (upper right), u-velocity (lower left) and v-velocity (lower right) at 4 UTC, July 11, 2015 in Greensburg, KS. The model uses the MYJ physics parameterization.

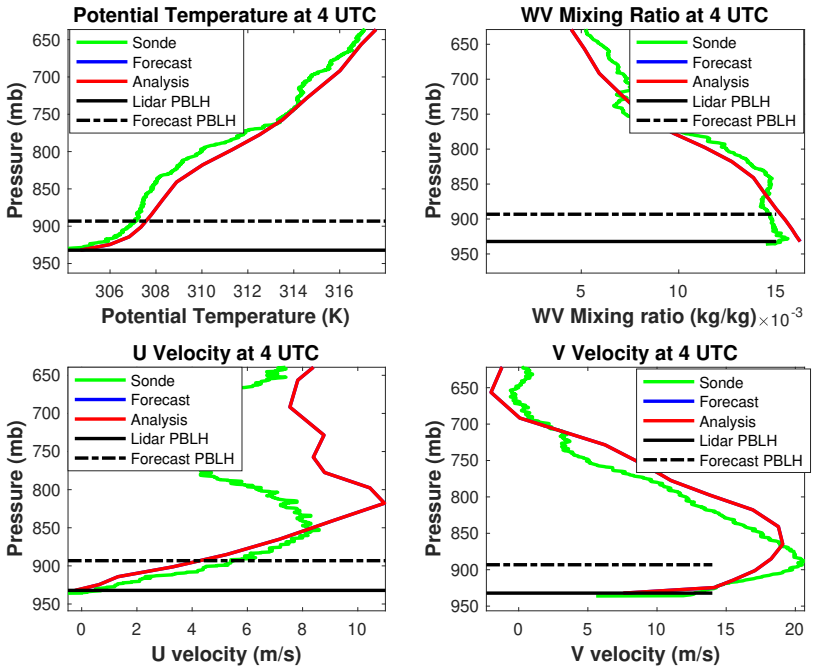



**Figure 5.** Same as figure 4 except using MYNN model.

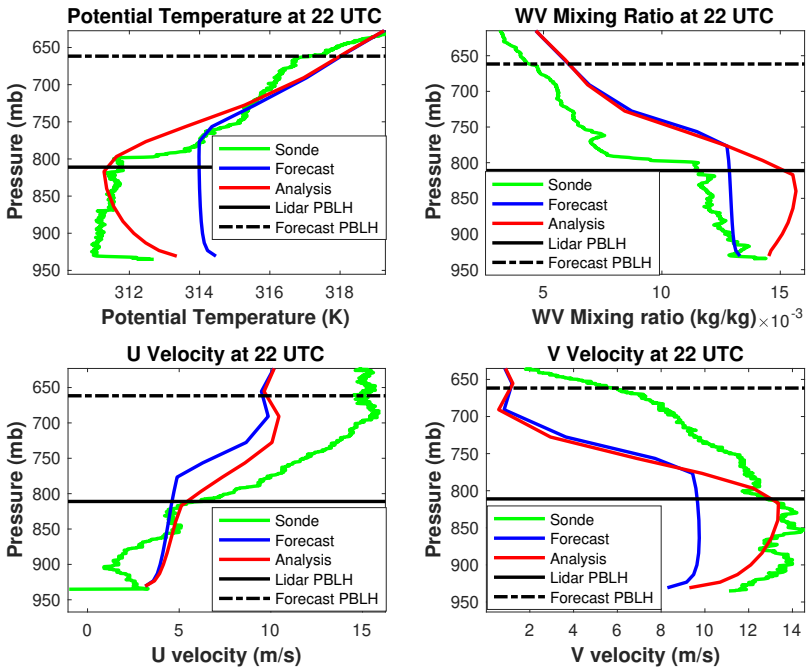



**Figure 6.** Same as figure 4 except using except at time 22 UTC.

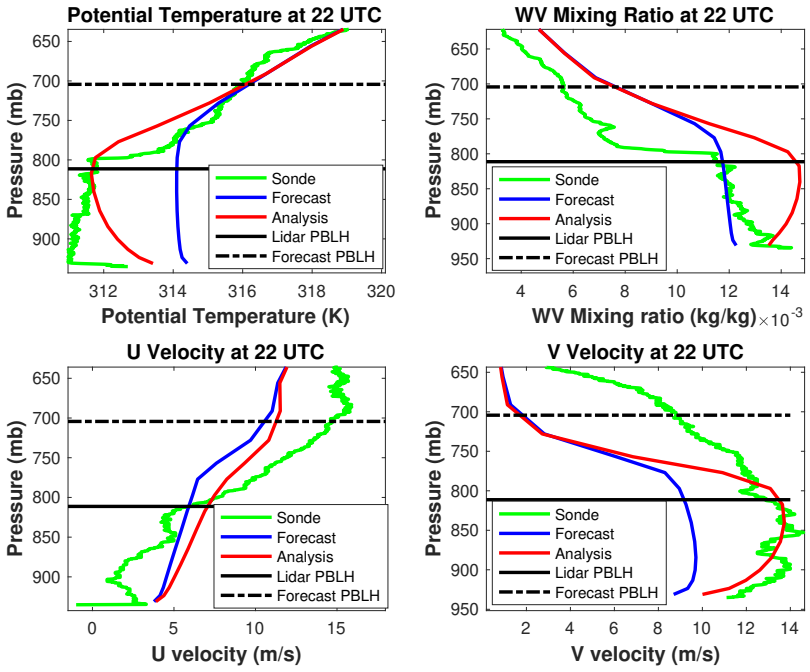



**Figure 7.** Same as figure 6 except using MYNN model.

noon near the surface. The covariance with WV is mostly negative and grows by roughly a factor of 5, while the covariance with the two components of velocity oscillate between positive and negative and shows less consistent growth. Thus, the largest impact of assimilation on temperature and moisture occurs in late afternoon while more limited velocity corrections are largely constrained by the correlations determined by the ensemble of model forecast states.

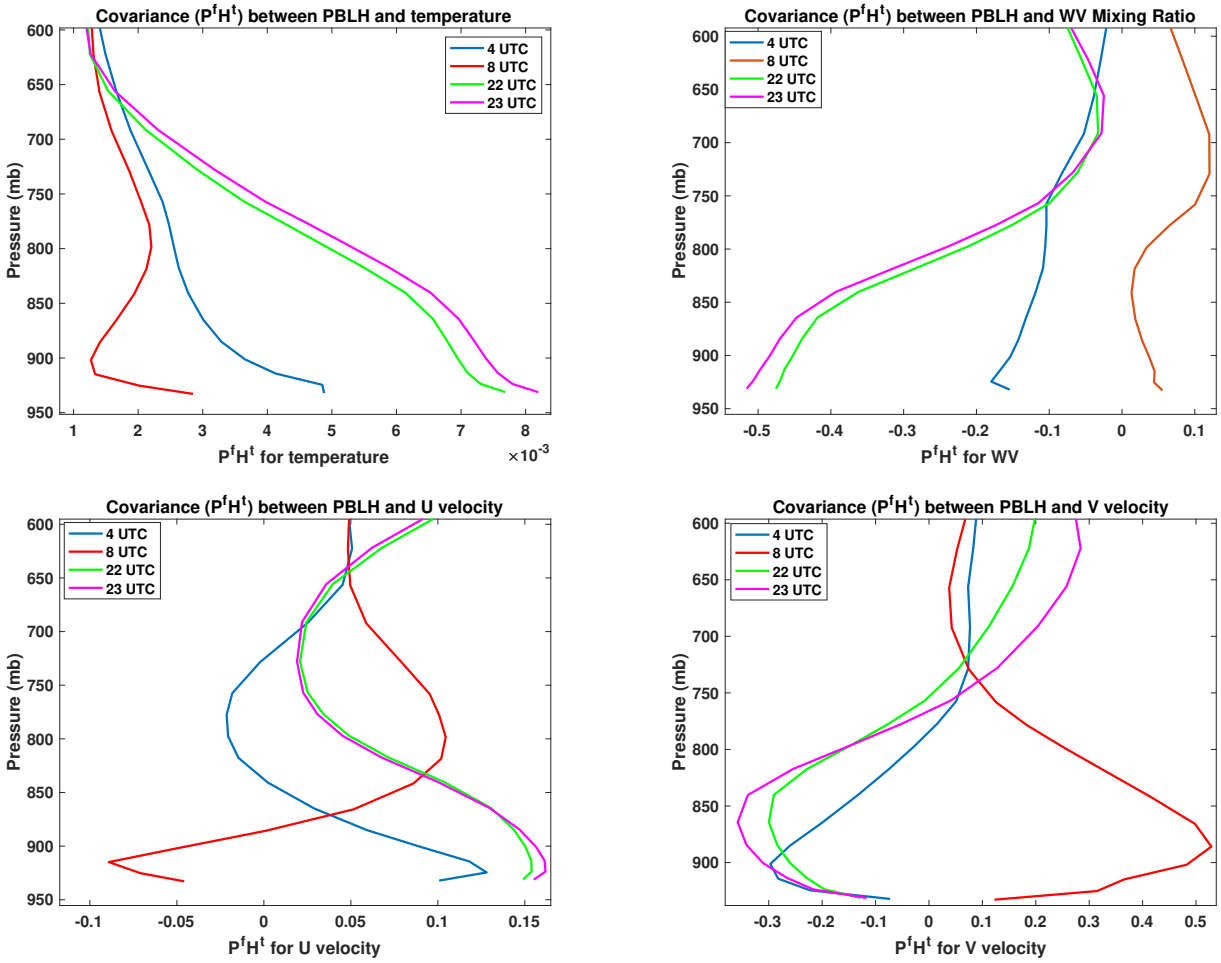



**Figure 8.** Covariance $\mathbf{P}^f\mathbf{H}^T$ between PBLH and temperature (upper left), water vapor (upper right), U velocity (lower left) and V velocity (lower right), at times 4, 8, 22 and 23 UTC, for PBL physics model MYHH.

## 4 Conclusions

These offline data assimilation experiments indicate that assimilating ground based lidar backscatter and wind measurements of PBLH into a regional NWP model will likely

lead to corrections to profiles within the PBL, particulary when this approach is applied to an EnKF assimilation system with cycling. Using two NU-WRF forecasts over a period of one day with different PBL physics models, we show how the state variables, T, WV, U and V can be corrected using an assimilation system with ensemble based error covariances. During the night and early morning the assimilation has relatively little impact on the state variables, but by late afternoon the temperature field is drawn closer to independent radiosonde measurements. We have shown that the lack of data impact early in the day is the due to the relatively higher accuracy of the model and lack of correlation between the forecast PBLH and temperature profiles at that time. Later in the day, when the model is less accurate in predicting the growth of the boundary layer, the data begins to draw the analyses mostly toward the independent radiosonde profiles. The assimilation over corrected the water vapor mixing ratio in the direction of radiosonde data, and this could likely be tuned in an assimilation system. And it corrected the the V velocity component by a smaller amount, and reduced differences with the radiosonde profiles for the V velocity. These corrections are the result of ensemble computed error covariances between the PBLH and the state variable profiles within the PBL. The results here indicate that this approach has some potential to be used in a forecast system in a way that that the PBLH observational information could be carried forward in time so as to impact the forecast accuracy within the PBL. An additional value of assimilating PBLH is its close connection with the PBL scheme used in the model. The covariances between PBLH and the different state variables are defined through the PBL physics scheme. This has an impact on the corrections made to the profiles within the PBL, which can be used as another way to evaluate the physics parameterizations. For example, the MYJ and MYNN result in analysis profiles that differ, particularly in WV in the late afternoon. And the differences in reponse to assimilation are an indication of how the two different PBL schemes affect the covariance between PBLH and the state variables. However, a full evaluation would require that the assimilation be implemented into a cycling data assimilation system.

This work is intended only to demonstrate a necessary first step in terms of how ensemble statistics can help to constrain profiles within the PBL by assimilating PBLH observations. A more complete demonstration of this approach will require the construction of an EnKF, and run over many days with a variety of weather patterns, including significantly warmer(cooler) and wetter(drier) days. This is needed to show how the as-

similated PBLH observations will impact future forecasts within the PBL. In addition, an EnKf will involve spatial covariances in both horizontal and vertical directions, and will allow for both inflation and horizontal localization. This will enable further tuning of the system to optimize the analysis state relative to the independent radiosonde observations. The PBLH assimilation withn the EnKF framework could be done in any of numerous existing enKF assimilation systems that connect with WRF, including NU-WRf (Lidard-Peters *et al.*, 2015) and WRF-DART (Anderson *et al.*, 2009).

## 5 Acknowledgments

B. Demoz was funded by National Science Foundation award (AGS-1503563) to the University of Maryland, Baltimore County and through NOAA Cooperative Science Center in Atmospheric Sciences and Meteorology, funded by the Educational Partnership Program at NOAA in collaboration with Howard University.

## 6 Data Sets

PECAN (`https://data.eol.ucar.edu/master_list/?project=PECAN\verb`) data are archived by NCAR/EOL, which is funded by NSF. The forecast and analysis fields produced for this work are stored at https://alg.umbc.edu/pecan/.

## 7 Competing Interests

The authors declare that they have no conflict of interest.

## 8 Author Contributions

Andrew Tangborn built the assimilation system, with input from Jeffrey Anderson on the algorithm. Belay Demoz and Brian Carroll provided the lidar observations. Joseph Santanello provided background information on PBL physics. All of the authors contributed to writing and revising the paper.

## 9 References

Anderson, J.L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn and A. Arellano (2009), The Data Assimilation Research Testbed: A Community Facility, *Bull. Amer. Met. Soc.*, 90, 1283-1296 doi:10.1175/2009BAMS2618.1.

Bonin, T.A., B.J. Carroll, R.M. Hardesty, W.A. Brewer, K. Hajney, O.E. Salmon and P.B. Shepson (2018), Doppler Lidar Observations of the Mixing Height in Indianapolis Using an Automated Composite Fuzzy Logic Approach, *J. Atmos. Ocean Tech.*, 35, 473-490.

Brooks, I.M. (2003), Finding Boundary Layer Top: Application of a Wavelet Covariance Transform to Lidar Backscatter Profiles, *J. Atmos. Ocean Tech.*, 20, 1092-1105.

Browning, K. A., and Coauthors (2007), The Convective Storm Initiation Project. , *Bull. Amer. Meteor. Soc.*, 88, 1939–1955, https://doi.org/10.1175/BAMS-88-12-1939.

Carroll, B. J., Demoz, B. B., and Delgado, R. (2019). An overview of low-level jet winds and corresponding mixed layer depths during PECAN. Journal of Geophysical Research: Atmospheres, 124(16), 9141-9160. https://doi.org/10.1029/2019JD030658.

Coniglio, M. C., G. S. Romine, D. D. Turner, and R. D. Torn, 2019: Impacts of Targeted AERI and Doppler Lidar Wind Retrievals on Short-Term Forecasts of the Initiation and Early Evolution of Thunderstorms. *Mon. Wea. Rev.*, 147, 1149–1170.

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

Degelia, S. K., X. Wang, and D. J. Stensrud, 2019: An Evaluation of the Impact of Assimilating AERI Retrievals, Kinematic Profilers, Rawinsondes, and Surface Observations on a Forecast of a Nocturnal Convection Initiation Event during the PECAN Field Campaign. *Mon. Wea. Rev.*, 147, 2739–2764.

Delgado, R., Carroll, B. and Demoz, B. (2016). FP2 UMBC Doppler Lidar Line of Sight Wind Data. Version 1.1 [Data set]. UCAR/NCAR - Earth Observing Laboratory. Accessed 29 May 2017. https://doi.org/10.5065/d6q81b4h.

Demoz, B., C. Flamant, T. Weckwerth, D. Whiteman, K. Evans, F. Fabry, P. Di Girolamo, D. Miller, B. Geerts, W. Brown, G. Schwemmer, B. Gentry, W. Feltz, and Z. Wang, 2006: The dryline on 22 May 2002 during IHOP-2002: Convective scale measurements at the profiling site. Mon. Wea. Rev., 134(1), 294-310.

Evensen, G. (2009), *Data assimilation: the ensemble Kalman filter*, Springer.

Geerts, B., and Coauthors, (2017), The 2015 Plains Elevated Convection At Night field project. *Bull. Amer. Meteor. Soc.*, 98, 767–786, https://doi.org/10.1175/BAMS-D-15-00257.1.

Hegarty, J.D., J. Lewis, E.L. McGrath-Spangler, J. Henderson, A.J. Scarino, P. DeCola, R. Ferrare, M. Hicks, R.D. Adams-Selin and E.J. Welton (2018) Analysis of the Planetary Boundary Layer Height during DISCOVER-AQ Baltimore–Washington, D.C., with Lidar and High-Resolution WRF Modeling, *J. Appl. Meteo. Climat.*, 57, 2679-2696.

Hicks, M., D. Atkinson, B. Demoz, K. Vermeesch and R. Delgado (2016), The National Weather Service Ceilometer Planetary Boundary Layer Project, *The 27th International Laser Radar Conference (ILRC 27)*, https://doi.org/10.1051/epjconf/201611915004.

Hicks, M., B. Demoz, K. Vermeesch and D. Atkinson (2019), Intercomparison of Mixing Layer Heights from the National Weather Service Ceilometer Test Sites and Collocated Radiosondes, *J. Atmos. Ocean Tech.*, 36, 129-137.

Holzworth, G.C. (1964), Estimates of mean maximum mixing depths in the contiguous United States, *Mon. Wea. Rev.*, 92, 235-242.

Hong, S.-Y. and H.-L. Pan (1996), Nonlocal boundary layer vertical diffusion in a medium-range forecast model, *Mon. Wea. Rev.*, 124, 2332-2339.

Hong, S.-Y. and H.-L. Pan (1998), Convective Trigger Function for a Mass-Flux Cumulus Parameterization Scheme, *Mon. Wea. Rev*, 126, 2599-2620.

Houtekamer, P.L. and F. Zhang (2016), Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, *Mon. Wea. Rev.*, 144, 4489-4532.

Hu, J., N. Yussouf, D. D. Turner, T. A. Jones, and X. Wang, 2019: Impact of Ground-Based Remote Sensing Boundary Layer Observations on Short-Term Probabilistic Forecasts of a Tornadic Supercell Event, *Wea. Forecasting*, 34, 1453–1476.

Janjic, Z.I. (1994), The Step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure, *Mon. Wea. Rev.*, 122, 927-945.

Janjic, Z.I. (2002), Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model (NCEP Office Note No. 437).

T. N. Knepp, J.J. Szykman, R. Long, R. M. Duvall, J. Krug, M. Beaver, K. Cavender, K. Kronmiller, M. Wheeler, R. Delgado, R. Hoff, T. Berkoff, E. Olson, R. Clark, D. Wolfe, D. Van Gilst, D. Neil (2017), Assessment of mixed-layer height estimation from single-wavelength ceilometer profiles, *Atmos. Meas. Tech.*, 10, 3963-3983.

Lothon, M., Lohou, F., Pino, D., Couvreux, F., Pardyjak, E. R., Reuder, J., et al. (2014). The BLLAST field experiment: Boundary-Layer late afternoon and sunset turbulence. Atmospheric Chemistry and Physics, 14(20), 10931–10960. https://doi.org/10.5194/acp-14-10931-2014.

Mellor, G.L. and T. Yamada (1974), A Hierarchy of Turbulence Closure Models for Planetary Boundary Layers, *J. Atmos. Sci.*, 31, 1791-1806.

Mellor, G.L. and T. Yamada (1982), Development of a turbulence closure model for geophysical fluid problems, *Rev. Geophys.*, 20, 851-875.

Nakashini, M. and H. Niino (2009), Development of an improved turbulence closure model for the atmospheric boundary layer, *J. Met. Soc. Japan*, 87, 895-912.

National Research Council (2009), Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks, in: Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks, 1–234, Natl. Academies Press, 2101 Constitution Ave, Washington, DC 20418 USA.

425 NCAR Technical Note (2012), Thermodynamic Profiling Technologies Workshop Report
426 to the National Science Foundation and the National Weather Service, National Cen-
427 ter for Atmospheric Research.

428 Oke, P.R., G.B. Brassington, D.A. Griffin, and A. Schiller (2010), Ocean data assimi-
429 lation: a case for ensemble optimal interpolation, *Austr. Meteor.Ocean. J.*, 59, 67-76.

430 Peters-Lidard, C.A. and Co-authors (2015), Integrated modeling of aerosol, cloud, pre-
431 cipitation and land processes at satellite-resolved scales, *Environ. Mod. Soft.*, 67, 149-
432 159.

433 Santanello, J.A. and Co-authors (2018), Land–Atmosphere Interactions: The LoCo Per-
434 spective, *Bull. Amer. Meteor. Soc*, https://doi.org/10.1175/BAMS-D-17-0001.1.

435 Santanello, J.A.,, S.Q. Zhang, D.D. Turner, P. Lawston, and W.G. Blumberg, PBL Ther-
436 modynamic Profile Assimilation and Impacts on Land-Atmosphere Coupling, AGU Fall
437 Meeting, San Francisco, CA, Dec. 9-13, 2019.

438 Wulfmeyer, V., R.M. Hardesty, D.D. Turner, A. Behrendt, M.P. Cadeddu, P. Di Giro-
439 lamo, P. Schlussel, J.Van Baelen and F. Zus (2015), A review of the remote sensing of
440 lower tropospheric thermodynamic profiles and its indispensable role for the understand-
441 ing and the simulation of water and energy cycles, *Rev. Geophys.*, https://doi.org/10.1002/2014RG000476.