Convective Boundary Layer Depth Estimation from Wind Profilers: Statistical Comparison between an Automated Algorithm and Expert Estimations

LAURA BIANCO
Cooperative Institute for Research in Environmental Sciences, University of Colorado, and NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado

JAMES M. WILCZAK
NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado

ALLEN B. WHITE
Cooperative Institute for Research in Environmental Sciences, University of Colorado, and NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado

(Manuscript received 23 January 2007, in final form 8 April 2008)

ABSTRACT

A previous study showed success in determining the convective boundary layer depth with radar wind-profiling radars using fuzzy logic methods, and improvements to the earlier work are discussed. The improved method uses the Vaisala multiphase picking (MPP) procedure to identify the atmospheric signal in radar spectra in place of a fuzzy logic peak picking procedure that was previously used. The method then applies fuzzy logic techniques to calculate the depth of the convective boundary layer. The planetary boundary layer depth algorithm is improved with respect to the one used in the previous study in that it adds information obtained from the small-scale turbulence (vertical profiles of the spectral width of the vertical velocity), while also still using vertical profiles of the radar-derived refractive index structure parameter $C_n^2$ and the variance of vertical velocity. Modifications to the fuzzy logic rules (especially to those using vertical velocity data) that improve the algorithm’s accuracy in cloudy boundary layers are incorporated. In addition, a reliability threshold value to the fuzzy logic–derived score is applied to eliminate PBL depth data values with low score values. These low score values correspond to periods when the PBL structure does not match the conceptual model of the convective PBL built into the algorithm. Also, as a final step, an optional temporal continuity test on boundary layer depth has been developed that helps improve the algorithm’s skill. A comparison with independent boundary layer depth estimations made “by eye” by meteorologists at two radar wind-profiler sites, significantly different in their characteristics, shows that the new improved method gives significantly more accurate estimates of the boundary layer depth than does the previous method, and also much better estimates than the simpler “standard” method of selecting the peak of $C_n^2$. The new method produces an absolute error of the mixing-depth estimates comparable to the vertical range resolution of the profilers.

1. Introduction

The depth of the atmosphere’s turbulent planetary boundary layer (PBL) is well recognized as an important parameter for air quality monitoring and prediction studies, as well as for the evaluation of numerical weather prediction models. One potential method to routinely monitor the dynamically defined PBL depths uses high-resolution wind-profiling radars, as the maximum value of the radar-derived refractive index parameter $C_n^2$ (which usually emerges at the inversion due to small-scale buoyancy fluctuations associated with the entrainment process), which often provides a good estimate of the depth of the PBL (Angevine et al. 1994). More recently, an automatic fuzzy logic–based technique by Bianco and Wilczak (2002, hereafter BW02) has shown success in determining the convective PBL depth using wind-profiling radar measurements of $C_n^2$.
as well as the variance of vertical velocity, which is large within the PBL, but small above it. However, additional information on boundary layer structure is also contained in the Doppler width of the atmospheric spectral peak of the radar signal (Gossard et al. 1984, 1998), which is related to the turbulence intensity. Heo et al. (2003) have shown how inclusion of vertical profiles of Doppler spectral width can be used to help identify the PBL depth in wind-profiling radar data.

The structure parameter of refractive index \( C_n^2 \) is related to the signal-to-noise ratio (SNR) through the relation (Gossard et al. 1982; Stankov et al. 2003; White 1997):

\[
C_n^2 = \frac{1.54 \times 10^{-13} T_0}{\alpha^2 P n_c A_p} \lambda^{1/2} \left( \frac{R}{\Delta R} \right)^2 \text{SNR},
\]

where \( T_0 \) is the system noise temperature, \( \alpha^2 \) accounts for the losses in the transmission line, \( P \) is the transmitted power, \( n_c \) is the number of coherent integrations, \( A_p \) is the physical antenna area, \( \lambda \) is the radar wavelength, \( R \) is the range to the target, and \( \Delta R \) is the range resolution, and SNR is estimated from the spectral data.

An example of radar wind-profiling data that illustrates the behavior of range-corrected SNR, vertical velocity, and spectral width is shown in Fig. 1. This wind-profiling radar operates at 915-MHz, the interpulse period (IPP) is equal to 35 \( \mu \)s, the pulse width (PW) is equal to 400 ns, and the number of gates is equal to 50. The vertical resolution is 60 m, with height ranges from 157 to 3097 m, and the number of vertical acquisitions per hour is about 240. The unusual high number of vertically pointing acquisition used for this radar setting gives an excellent picture of the dynamic processes that take place in the planetary boundary layer. Important features are the development of a local maximum of SNR (top panel), which begins near ground level at sunrise [0533 local standard time (LST)] and the height of which increases with time. Below this peak in SNR, the vertical velocity (middle panel) alternates between intense updrafts and downdrafts, giving a large velocity variance. The Doppler spectral width (bottom panel) also shows larger values within the PBL and smaller values above.

An important aspect of radar-derived PBL depths is that they are a function of the turbulence itself, and therefore they can provide a measure of the depth of the turbulent PBL by some temporal or spatial average over the ensemble of turbulent structures around the time of measurement. In this sense the radar-derived PBL values represent the depth of the dynamically defined boundary layer, the layer of the atmosphere in contact with the earth’s surface through which rapid mixing is occurring due to the presence of turbulence. This is an important feature for tracking the evening transition from a convective to a stable boundary layer, as can be seen in Fig. 1. In the late afternoon hours (1600–1900 LST), the region of enhanced SNR, velocity variance, and spectral width first begins to decay aloft as the convective PBL begins to collapse. A residual layer of enhanced SNR remains at the height where the PBL top had been previously, which corresponds to the temperature inversion, but below this a turbulence-free zone expands downward with time, spanning the layer from 2.0 to 1.3 km at 1900 LST. In contrast, boundary layer depth detection methods that rely on profiles of mean temperature, winds, or aerosols, such as those obtained from balloon soundings or lidars, would continue to provide a depth corresponding to the temperature inversion at the earlier maximum depth of the PBL, and not to the present depth of the turbulence.

Although the example shown in Fig. 1 is of a clearly defined PBL at most hours, ambiguities may sometimes exist. These ambiguities can arise for example from weak levels of turbulence within the PBL, intermittent turbulence, and clouds. A particularly difficult period of the day is in the late afternoon with the collapse of the boundary layer. At this time the layer below the inversion is intermittently turbulent before it becomes fully decoupled from the inversion, and the depth of the PBL is nebulous. An automatic algorithm for detecting PBL depth should be able to differentiate between times when the PBL is a clearly defined phenomenon, and times when its existence is tenuous or ill-defined.

Based on these considerations, we have expanded on BW02 to improve the algorithm’s ability to determine the depth of the convective PBL using wind-profiler data. The key improvements in the new method are that it 1) includes the Doppler spectral width, 2) computes a reliability value indicating when the PBL is ill defined, and 3) provides the option for a time-continuity test on hourly PBL depth estimates.

The improved method is applied to wind-profiler data at two sites (i.e., Pittsburgh, Pennsylvania, and Plymouth, Massachusetts) that were taken during the 2004 New England Air Quality Study (NEAQS) field program. Statistical correlations are calculated between the PBL depths calculated from the PBL depth algorithm and those obtained by two meteorologists who visually inspected data similar to that shown in Fig. 1. All PBL depths are calculated as hourly averages. This represents a compromise between having a sufficiently long time series to provide a temporal average over an
ensemble of turbulent eddies, and a short enough interval to adequately represent the diurnal variation in PBL depth.

The outline of the paper is as follows. Section 2 illustrates the methods used in processing the profiler data. Section 3 describes the PBL depth detection algorithm. Section 4 introduces the two specific experimental datasets utilized in this work, the construction of the independent PBL depth dataset derived by the experts, and the methodology used for determining the value of different components of the algorithm. Section 5 provides a summary of the results, while section 6 presents the conclusions.

2. Data processing

Once the spectral data have been measured by the wind-profiling radar, the first step is to compute the spectral moments. More information on data processing algorithms used by wind profilers can be found in Barth et al. (1994), while evaluation of moment-estimation techniques can be found in Gaffard et al.
The purpose of this paper is not to provide another evaluation of these techniques. On the other hand, as discussed in BW02, the skill of the PBL depth algorithm depends not only on the rules incorporated into the fuzzy logic part of the algorithm, but also on the quality of the radar moment data that the algorithm uses. Hence, here we test three different processing methods to find the one most suitable to our purposes.

1) The standard peak picking (SPP) processing method used in the Vaisala wind-profiler software LAP-XM that detects and differentiates clutter and atmospheric spectral peaks using the technique described in Riddle and Angevine (1992), which attempts to remove ground clutter but not radio frequency interference or point target clutter.

2) A fuzzy logic–based procedure for moment estimation (FLPP), developed in our earlier work (BW02), which applies fuzzy logic methods for the signal peak identification and clutter recognition.

3) The multipeak picking (MPP) procedure (Griesser and Richner 1998), which is commercially available in Vaisala wind-profiler software and that has the advantage of being much faster than the FLPP method.

SPP, FLPP, and MPP are used to create moment files that are then used as input to the improved algorithm for the estimation of the PBL depth. All the moment-estimator methods produce one estimate of the moments at each range gate. In all the methods, the noise level is computed using Hildebrand and Sekhon (1974).

The PBL depth algorithm consists of the following steps:

1) Sunrise/sunset times are computed using the latitude and longitude for the profiler site and the day of the year. If the hour of the acquisition is located in the sunrise to sunset time range, then the PBL depth is computed, otherwise, it is not.

2) Mean profiles of the radar-derived SNR are computed. This includes range-correcting SNR on the vertical and each of the oblique beams, interpolating the oblique SNR values to the heights of each gate on the vertical beam, and calculating at each of these gates the arithmetic average of the oblique and vertical beam SNR values. The reason for averaging the SNR values for each beam is to reduce random noise and provide smoother profiles.

3) Hourly mean profiles of the radar-derived spectral width (second moment or spread of the spectrum) are computed for the vertical beam only.

4) Hourly mean values of vertical velocity variance are calculated for the vertical beam data that passed a consensus procedure (Strauch et al. 1984). The consensus window is 3 m s$^{-1}$ wide, and requires a minimum of 50% of the data values to calculate the variance. Consensus averaging is only applied to the vertical velocity data and not to SNR or the spectral width because when the profiler return becomes too small to detect an atmospheric signal, SNR and spectral width become small, which is information that can indicate that the range gate is above the PBL, while the vertical velocity measurement becomes random.

5) At the end of the hour we have three representative vertical profiles: one for the SNR (and therefore for $C_n^2$), one for the Doppler width of vertical velocity, and one for the variance of vertical velocity. Missing data of SNR and Doppler width are linearly interpolated. Then, these three profiles are smoothed using a ninth-order Savitzky–Golay (polynomial) smoothing filter (Savitzky and Golay 1964). The main advantage of this approach is that it tends to preserve features of the distribution such as relative maxima, minima, and width, which are usually “flattened” by other adjacent averaging techniques. These smooth vertical profiles are used as input to the fuzzy logic part of the code that estimates the PBL depth.

An example of smoothed vertical profiles used as input in the fuzzy logic PBL depth algorithm part is shown in Fig. 2. The vertical profiles correspond to the period from 1600 to 1700 LST 31 July 2004 at the Plymouth wind-profiler site. We note that the vertical profile of spectral width in Fig. 2 has a local maximum at a height slightly higher than the one where the vertical profile of SNR has its maximum. This characteristic can be explained by instrumental factors that contribute to the broadening of the spectrum.

Three factors in particular can affect the clear-air spectrum measured by the profiler, and have been widely discussed by Gossard and Strauch (1983) and Hocking (1983, 1985, 1986). If there is a shear in the mean wind, the speed and direction of the mean wind may vary significantly within the pulse volume, causing broadening of the Doppler spectrum. This contribution to the variance by shear transverse to the beam is proportional to the transverse shear, $\beta_{\tau}$ (s$^{-1}$) (Sloss and Atlas 1968). Similarly, a shear in the radial wind across the pulse volume will contribute to the spectral width proportionally to the radial shear of the radial component of the wind, $\beta_{r}$ (s$^{-1}$) (Sirmans and Doviak 1973). Finally, even if there is no turbulence or shear in the
radial component, so that it is completely uniform in speed and direction, and even if the reflectivity is constant over the pulse volume, the Doppler spectrum will have a measurable width when measured with any real radar because of the finite width of the radar beam. This effect, often referred to as beam broadening, is dependent on the antenna properties, and is small if the beam width is small. Its contribution is proportional to the component of the mean wind that is transverse to the beam axis, $V_T (\text{m s}^{-1})$. Because the total variance of a group of independent Gaussian broadening processes is the sum of the individual processes, they contribute additively. A simple analysis of the various quantities involved in the broadening of the spectrum demonstrates that these contributions can be nontrivial. If we consider a range of values for $\beta_E$ and $\beta_T$ that goes from 0 to $3 \times 10^{-3} \text{ m s}^{-1}$ and a range of values for $V_T$ that goes from 0 to 10 m s$^{-1}$, the combined contribution of these three terms to the spectral width ranges from 0 to 1 m s$^{-1}$ for the parameter settings used in the case of the Dupont wind profiler and in the range of altitudes covered by the instrument. Therefore, even at times when the turbulence is not well developed, the measured spectral width of the vertical velocity can still be significant. In particular, in the entrainment zone the increasingly large contributions from the transverse and radial shear terms can produce a local maximum in spectral width, and the spectral width may not begin to decrease until several range gates above the height where SNR reaches its maximum value. This behavior is clearly evident in Fig. 2.

Corrections could be applied to the spectral width to account for the beam broadening and for the contribution by shear transverse to the beam using the profiler-measured horizontal wind and its vertical gradient. On the other hand, correcting for the contribution due to the shear in the radial wind across the pulse volume is not straightforward, as the radial shear is not measured. We have chosen to use the directly measured profile of spectral width in the PBL depth fuzzy logic algorithm, but to structure the algorithm so as to account for spectral width profiles such as that shown in Fig. 2.

**Fig. 2.** Hourly averaged vertical profiles for the period 1600–1700 LST 31 Jul 2004 at the PYM wind-profiler site. (left) Range-corrected SNR (dB), (middle) spectral width of the vertical velocity (m s$^{-1}$), and (right) variance of the vertical velocity (m s$^{-1}$) computed on the values of vertical velocity that have passed the consensus procedure. Cloud base was determined from a lidar ceilometer.
3. Boundary layer depth detection algorithm

The vertical profiles of radar Doppler moment data as computed above are combined together in the PBL fuzzy logic part of the algorithm similarly to BW02. However, that algorithm has been modified in five significant ways, some of which are not fuzzy logic based. These are 1) the inclusion of spectral width, 2) the use of a modified rule for vertical velocity variance, 3) the use of the highest local maxima in the score values instead of the absolute maximum score, 4) the incorporation of score values to denote the reliability of the estimated PBL depth value, and 5) an optional use of a temporal continuity test that compares each hourly PBL depth value with the overall evolution of PBL depth throughout the day. Each of these five new elements is discussed in detail below.

a. Spectral width

The rule for the spectral width in the PBL fuzzy logic algorithm is simply that if the spectral width is large at a given range gate, the probability of that level being the top of the PBL is large, and conversely, that if the spectral width is small, the probability of it being the PBL top is small.

b. Vertical velocity

In the previous approach, the PBL algorithm’s incorporation of vertical velocity was based on the well-known behavior of the vertical velocity variance in the clear-air convective PBL (i.e., that it increases with height to a maximum near the middle of the convective boundary layer, then decreases with height to the inversion). Consequently, the algorithm applied the rule that if the variance was large at a given range gate, the probability of that gate being the PBL top was small, and conversely, that if the variance was small, the probability was large. During cloud-free conditions, such as the data used in BW02, the profile of vertical velocity variance profile generally follows the expected pattern. However, many of the days at Plymouth and Pittsburgh were partly cloudy, especially with fair-weather boundary layer cumulus. When boundary layer clouds are present, the behavior of the profile of vertical velocity variance changes, and the interpretation of all three radar moments can be more difficult.

For stratus-topped marine boundary layers, White et al. (1991) have shown that the peak in radar reflectivity coincides with the top of the cloud layer, where a strong gradient in humidity is present. For continental boundary layers with intermittent fair-weather cumulus clouds, the situation is more complicated. Within the cloud itself, the greatest humidity gradients will exist at the cloud top. However, in the downdraft clear-air regions between the clouds, the maximum humidity gradient will occur near to (or somewhat below) the cloud base. Using a combination of lidar ceilometer and wind-profiler data, Grimsdell and Angevine (1998) find that for continental boundary layers, the mean profiler–derived PBL depth is equal to or slightly above the cloud base.

At the Plymouth site (and only this site) lidar ceilometer data were also available. Daily time–height plots that combined the three radar parameters, the expert’s and algorithm’s PBL depth estimates, and ceilometer cloud-base measurements were evaluated (not shown). These demonstrated that in fact most often the PBL height was associated with the cloud base, in agreement with Grimsdell and Angevine (1998).

In Fig. 2 vertical profiles of range-corrected SNR, spectral width, and vertical velocity variance are displayed for an hour in which the ceilometer shows a cloud-topped boundary layer situation. The variance of vertical velocity in this plot increases with height through the inversion layer/cloud base, demonstrating why the previously used fuzzy logic rule looking for large values of variance in the mid-PBL and small values near the inversion is of little value in cloudy boundary layer situations. We therefore had to modify the use of vertical velocity variance information in the new algorithm, so that the rule was changed to only restrain the PBL depth to be at or below the first range gate at which the consensus algorithm is unable to form a mean vertical velocity variance (i.e., less than 50% of the vertical velocity data pass the consensus test). As can be seen in Fig. 1, the vertical velocity data are usually continuously measured in time and height through the PBL until the later afternoon hours. Above the PBL velocity measurements rapidly become an almost pure noise signal. During the late-afternoon hours a nonturbulent layer develops below the inversion and the PBL begins to collapse from the top downward.

c. Vertical score profiles

In the previous approach presented in BW02 a score was calculated at each range gate, and then the depth of the convective PBL was taken as the range gate that had the maximum value of the score. The new algorithm uses a different approach. It calculates all local maxima of scores through the depth allowed by the vertical velocity variance criterion, and then selects the depth of the PBL as the highest gate from all of the local maxima of scores, rather than the height of the absolute maximum score. The reason for this change is that the modification made in the use of the vertical
velocity variance information to account for cloudy boundary layers profiles no longer reduces the scores in the lower portion of the PBL. Consequently, the profiles of scores frequently have a local maximum with a slightly smaller value at the true top of the PBL than at gates within the mid-PBL.

d. Reliability values

At this point the PBL depth estimation algorithm provides a time series of hourly data points of PBL depths and score values associated with that depth. The value of the score represents the degree to which the data fits our conceptual model of what the PBL looks like in wind-profiler data. Therefore, the higher the score the more confidence we have that the algorithm has accurately selected the PBL depth, and the more reliable is the algorithms value. In the improved algorithm we use the score value to set a reliability threshold for the acceptance of a PBL estimate. If the threshold value is set low, many data points are accepted as “good,” although the percentage of wrong values is large. If the threshold is high, few wrong estimates are accepted, but at the expense of eliminating some good values as well. Through an analysis of score histograms (section 5b) we find a suitable value of the threshold value.

e. Temporal continuity

Finally, as an optional step, an additional temporal continuity test over a period of 1 day can be performed on the boundary layer depth estimations to improve the results of the analysis. If the user requires a real-time PBL depth estimate at the end of every hour, then this option would not be applied. However, if the user requires the PBL depths after a complete day’s worth of data have been collected, this part of the code can be applied, eliminating further outliers. The reason why we decided to add this step is related to the fact that ahead in the analysis we found some outliers between the expert estimations and the results of the PBL depth method. Upon reexamination of the three panel time-height cross sections for the days with these outliers, the two experts concluded that if they had examined the $C_n^2$, vertical velocity, and spectral width only for the individual hour of each of the outliers, they would have chosen a PBL depth in very close agreement with the algorithm. The reason that they chose a different value was because by viewing an entire day’s worth of data at one time, they were mentally comparing data from one hour to the next, incorporating time continuity into their analysis to eliminate large transient, temporal changes in PBL depth.

The temporal continuity test requires a minimum of 4 h of PBL depth estimates for a given day. We remove one estimate at a time and perform an interpolation with a third-order polynomial over the remaining PBL estimations the algorithm generated for that day. The process is repeated for all the estimates, removing them one at the time. Each time, we measure the distance of the removed estimate from the best third-order polynomial fit calculated from the remaining PBL depth estimations, at its same temporal position. If this distance is greater than a threshold value, the removed point is considered to be an outlier and therefore eliminated from the set of daily estimations. If this distance is smaller than the threshold, the removed point is considered a good estimate for the PBL depth at that hour and therefore included in the set of daily estimates. The determination of the threshold value is explained in section 5d. We note that a potential disadvantage to applying any temporal continuity test is that in rare circumstances when the boundary layer does in fact change very rapidly a continuity test may eliminate these events.

4. Data analysis

To test how the new algorithm is performing compared to the previous one and to other procedures, we decided to focus on two datasets, collected at Pittsburgh (PIT) and Plymouth (PYM). PIT was chosen because it is a representative midlatitude continental site, while PYM, which is located 6 km from the Atlantic coast, is frequently influenced by sea breezes and maritime air masses. One important difference between the limited data used in BW02 and the data from PIT and PYM is that solar radiation measurements and satellite imagery indicate that the data used in the earlier analysis was for mostly clear-sky conditions, while the latter two sites often had intermittent clouds, especially fair-weather boundary layer cumulus.

Data were collected from 27 July to 16 September 2004, giving a total of 101 days for analysis. For both sites, local standard time is equal to UTC – 5 h. Both wind-profiling radars operated at 915 MHz, and sampled interleaved high-resolution and low-resolution modes. High-resolution data have been used almost exclusively in the present analysis as the PBL depths were nearly always within the high-resolution height range. For the PIT radar the IPP was equal to 25 $\mu$s, the PW was equal to 416 ns, and the number of gates was equal to 45. For the PYM radar the IPP was equal to 25 $\mu$s, the PW was equal to 400 ns, and the number of gates was equal to 38. The two different settings correspond to vertical resolutions of 62 (PIT) and 60 m (PYM);
vertical ranges of 125–2853 m (PIT) and 142–2362 m (PYM); and the number of vertical acquisitions per hour were 32 (PIT) and 9 (PYM). The reason for the larger number of samples at the PIT profiler is that it was using a developmental digital receiver card that allowed for increased sampling time by reducing the length of the dead time between pulses. Details for wind-profiler data processing are given in Strauch et al. (1984).

**a. Experts’ data analysis**

It is recognized by users of radar wind-profiler data that human experts provide the best estimates of PBL depths derived from this data, because of the ability of a human to recognize patterns in the data and cross correlate multiple variables simultaneously over both space and time. For this reason human expert selection or editing of PBL depths is the commonly used basis for published analyses of wind-profiler-derived PBL depths (W. Angevine 2007, personal communication) except for those that use the BW02 fuzzy logic algorithm. The BW02 algorithm can be viewed as an imperfect attempt at replicating the human expert’s thought processes in pattern recognition for selecting the PBL depth in wind-profiler data. We therefore use human expert depth estimations as our “true” PBL depth values, and compare these to the automated algorithm in its various stages of development.

Although human experts are recognized as capable of providing accurate PBL depths from wind-profiler data, no two experts will always be in close agreement. Differences usually occur when the convective boundary layer is not well defined in the data, and both experts have some uncertainty as to what the PBL depth is, but venture to make an estimation anyway. To eliminate these uncertain values, we require the human experts to be in reasonable agreement with one another, using the analysis procedure described below. The intent of this process is to provide a dataset of accurate PBL depths against which the automated algorithm can be evaluated.

Two “experts” (authors White and Wilczak) separately examined all 100 days of the combined PIT and PYM dataset, visually examining three-panel (i.e., range-corrected SNR, vertical velocity, and spectral width) time–height cross sections similar to those shown in Fig. 1. Using these data they selected the PBL depth for each daytime hour, estimating a PBL depth only when they were reasonably confident they could identify it, and not giving a value if it was ill defined. This visual analysis of the data was completed before work on the new automated algorithm was begun, and these values remained fixed throughout the study. A statistical comparison of the experts’ estimations is discussed in the following paragraphs.

In the case of Pittsburgh there were 285 h in which both the experts gave a value for the PBL depth, shown in Fig. 3a. Statistics presented in the upper-right corner of the figure shows that the correlation coefficient between the experts is very good and the percentage error (PE) is equal to 11%, where percentage error is defined as the mean value of the absolute difference between the experts estimations divided by the corresponding experts’ mean estimation, times 100.

As can be seen in Fig. 3a, there are several cases in which the experts disagreed considerably in their PBL depth estimates. In addition, there is a slight bias between the two experts. Therefore we created an aver-
aged set of these estimations, which used the mean value of the estimations given by the experts, and does not include any value of the PBL height for the hours in which either of the experts failed to provide a PBL depth. Moreover, this set excluded the hours when the difference between the two experts was larger than one standard deviation relative to the best-fit (solid) line shown in Fig. 3a, which are those data points that fall outside of the two dashed lines. Therefore, we define the percentage of valid estimations (PercValid) as the total number of estimations that fall within the one standard deviation limits, divided by the total number of estimations, times 100. This procedure eliminated 25 data values at PIT, reducing the combined experts’ dataset there to 260.

The same kind of statistical comparison for the experts’ estimation was performed for the PYM site. Here there were 313 h in which both experts gave an estimated value for the PBL depth (Fig. 3b). The correlation coefficient is again very high, with a percentage error equal to 9%, and the root-mean-square error (RMSE) of the experts’ estimations is 105 m. Following the same procedure as for PIT, an averaged experts’ dataset was created that eliminated 34 values, leaving 279 PBL depth estimates.

b. PBL depth algorithm analysis

For our analysis of the improved PBL depth algorithm, we compare its results against those from the BW02 algorithm, and against those from a simpler algorithm that uses only $C_n^2$ profile information. In particular, we want to look at the performances provided by the inclusion of the spectral width of the vertical velocity into the algorithm, while also still using vertical profiles of both the radar-derived refractive index structure parameter $C_n^2$, and the variance of vertical velocity. The three different procedures for the PBL depth estimation tested are the following:

1) BLFL1: The standard boundary layer depth algorithm that is similar to that of Angevine et al. (1994) in that it relies solely on the maximum of the $C_n^2$ profile, using only one input variable.

2) BLFL2: This is identical to the BW02 algorithm and uses vertical profiles of both the radar-derived refractive index structure parameter $C_n^2$ and the variance of vertical velocity, or two input variables.

3) BLFL3: The improved method that adds information obtained from vertical profiles of the spectral width of the vertical velocity, while also still using vertical profiles of both $C_n^2$ and the variance of vertical velocity, for a total of three input variables. It also incorporates modifications to the use of the vertical velocity variance presented previously in section 3.

Results obtained by the three moment calculation techniques discussed in section 2 (SPP, FLPP, and MPP) were combined with the PBL depth algorithm that uses three inputs in the fuzzy logic part of the procedure (BLFL3). We will refer to these combinations with the acronyms: SPP-BLFL3, when the moments computed with the standard peak picking algorithm are combined with BLFL3; FLPP-BLFL3, when the moments computed with the fuzzy logic peak picking procedure are combined with BLFL3; MPP-BLFL3, when the moments for each range gate are computed with the MPP software, and the height of the PBL is chosen with the BLFL3 procedure. As will be shown from these first combinations, the MPP moments provide better results than either the SPP or FLPP moments. Therefore, later only MPP moments were tested in combination with the three different boundary layer depth algorithms. We will refer to these combinations with the following two acronyms: MPP-BLFL1, when the moments are computed with the MPP software and the height of the PBL is chosen with BLFL1; and MPP-BLFL2, when the moments are computed with the MPP software and the height of the PBL is chosen with the BLFL2 procedure. Finally we want to test the consequence of including the information on the spectral width of the vertical velocity in the improved PBL depth method. Therefore, we remove this input from the fuzzy logic part of the process and analyze the result. We refer to this combination as MPP-BLFL3 no SW.

5. Results

All the statistical results obtained in this work are summarized in Table 1 for the PIT site, in Table 2 for the PYM site, and in Table 3 for the PIT and PYM sites combined together.

a. Evaluation of the SPP, FLPP, and MPP moments

To test the results obtained with SPP-BLFL3, FLPP-BLFL3, and MPP-BLFL3, we compare the PBL depth estimations with those obtained by the experts’ estimations, for both PIT and PYM sites. The first step of the analysis doesn’t involve the use of a reliability value threshold to select good values of PBL depths, so that all of the estimations are included in the comparison, regardless of their reliability. Moreover, no temporal continuity test has been used at this stage of the analysis.
As an example of the automated PBL detection results, in Fig. 4 we show scatterplots for PIT and PYM of PBL depths from the MPP-BLFL3 algorithm versus those obtained by the experts. This first step of the analysis does not involve the use of a confidence value threshold to select good values of PBL depths, so that all of the estimations are included in the comparison, regardless of the confidence level. Moreover, no temporal continuity test has been used at this stage of the analysis.

The statistical results for all three algorithms are presented Tables 1 and 2. In particular, the table lists the number of points involved in the analysis (Npt), the squared correlation coefficient ($R^2$), the RMSE, the bias measured between the PBL depth method and the experts’ estimations (Bias), and the percent error (PE). We note that number of points involved in the statistical comparison differs among the methods, ranging from 238 to 256 for PIT and from 254 to 258 for PYM. This is due to the fact that the BLFL3 algorithm requires a local maximum in the profile of scores over the height interval for which a variance of vertical velocity can be computed after passing the vertical velocity data through the consensus procedure. The amount of vertical velocity data that passes through the consensus procedure varies among the three moment schemes, with fewer velocities passing consensus with SPP and FLPP than with MPP. Consequently, SPP and FLPP produce more cases with a smaller vertical height interval to search for the local maximum in the score, and there are more cases in which a local maximum is not found because the score increases monotonically over the allowed height range.

Comparing the statistics obtained by the different methods at this stage of the analysis (first three rows of Tables 1 and 2), the FLPP-BLFL3 method is performing better than the other algorithms (SPP-BLFL3 and MPP-BLFL3) at the PIT site, while overall the three methods perform similarly at PYM. At this point of the analysis no definitive conclusions can be determined, thus, it was decided to apply other considerations in order to improve the results.

### b. Reliability values

Figures 5a,b present histograms of the values of the scores for all of the PBL depths at both PIT and PYM. They illustrate that most of the PBL depth estimations for both sites have high values of scores. In fact, many of the estimations have a score equal to the maximum value of 100, and most of them have a high value (i.e., greater than 95) for both sites. In this step of the analysis, we examine the correlation between the value of these scores and the agreement of the estimations with the experts’ values.

The relationship between score value and the accuracy of the algorithm’s PBL depth estimate is presented in Figs. 6a,b for PIT and PYM, respectively. Here we plot the fuzzy logic score for each estimate versus the difference ($D$) between the MPP-BLFL3 and expert depth estimations in Figs. 4a,b. It is clear that most of the significantly erroneous PBL depth values (large values of $D$) have a lower value of the score, while most of the high score values have accurate depth estimates. This supports our contention that a higher score in the fuzzy logic algorithm for the PBL depth estimation is related to a higher level of confidence in the algorithm’s
estimate of the PBL depth. Using a relatively high-reliability threshold for the score equal to 95 (black vertical line) allows us to reject most of the PBL depth estimates that differ considerably from the experts’ values. However, we note that even with a threshold value of 95, five outliers at PIT and four at PYM remain (circled stars at the right-hand side of the 95 vertical lines). We will take these outliers under further consideration in section 5d. The algorithm was hence modified to include a reliability threshold value equal to 95 on the score associated with each PBL depth estimate.

We note that the choice of the threshold for the score is ultimately a subjective choice depending on each us-

![FIG. 4. (a) Scatterplot of PBL depth estimations from the MPP-BLFL3 method vs PBL depth estimations made by eye by the experts for the PIT site. (b) Same as (a), but for the PYM site.](image)

![FIG. 5. (a) Histogram of the distribution of the scores that the PBL depth estimations obtain at the end of the fuzzy logic (BLFL3) algorithm for the PIT site. (b) Same as (a), but for the PYM site.](image)
er’s requirements to exclude outliers at the expense of also potentially excluding numerous good data points. Also, although the score threshold value of 95 does an equally good job of excluding outliers for both the PIT and PYM sites, it is possible that different values may apply at other profiler sites.

The last three rows of Table 1 present the statistical results of the comparison of SPP-BLFL3, FLPP-BLFL3, and MPP-BLFL3 with the expert estimations for PIT, but now with a score threshold equal to 95. The last three rows of Table 2 are for PYM. From Tables 1 and 2 we note that the number of points included in the analysis is higher for MPP-BLFL3 compared to SPP-BLFL3 and FLPP-BLFL3, both for PIT and PYM. The use of a threshold equal to 95 rejects fewer PBL depth estimations in the MPP-BLFL3 algorithm than the other two algorithms. Moreover, values of $R^2$, RMSE, and PE are better when using the MPP-BLFL3 algorithm compared to the other two methods (i.e., SPP-BLFL3 and FLPP-BLFL3). Figure 7a shows the scatterplot between MPP-BLFL3 and expert estimations for PIT and Fig. 7b shows the same for PYM. Outliers are circled.

c. Testing the improvement of the PBL depth algorithm

Knowing that MPP gives the best moments estimation for use in the PBL depth algorithm when the reliability threshold is used, we now want to test how much better the new algorithm is compared to the previous, simpler algorithms that only used one or two input variables, and to determine the improvement that is attrib-

![FIG. 6. (a) Representation of the fuzzy logic scores values vs $D$ (difference between the estimates obtained by MPP-BLFL3 and the experts in Fig. 4a) for the PIT site. (b) Same as (a), but for the PYM site.](image)

![FIG. 7. (a) Scatterplot of PBL depth estimations from the MPP-BLFL3 method (with a value of the threshold equal to 95) vs PBL depth estimations made by the experts for the PIT site. Outliers are circled. (b) Same as (a), but for the PYM site.](image)
utable directly to the inclusion of the spectral width in the algorithm. To answer these questions we compare the results obtained by combining MPP with BLFL1 (i.e., choosing the height of the PBL where the hourly profile of $C_n^2$ shows its maximum value), with BLFL2 (i.e., using vertical profiles of both the radar-derived refractive index structure parameter $C_n^2$, and the variance of vertical velocity, identical to BW02), with BLFL3 minus the use of the spectral width of the vertical velocity as an input, and with BLFL3. Since the purpose of this comparison is to determine the relative skill of the new algorithm compared to previous algorithms, only MPP-BLFL3 uses the reliability threshold. Also, since we have shown that data from the PIT and PYM sites behave similarly, especially in terms of their sensitivity to the reliability threshold value, we combine the data from these two sites in the remaining part of the analysis.

Figures 8–11 show the statistical comparisons between MPP-BLFL1 (Fig. 8), MPP-BLFL2 (Fig. 9), MPP-BLFL3 no SW (Fig. 10), and MPP-BLFL3 (Fig. 11) and the experts' PBL depth estimates using the PIT and PYM data combined together. In this part of the analysis we use a value of the threshold equal to 95 for MPP-BLFL3 no SW and MPP-BLFL3, rejecting PBL depth estimates with values of the score lower than this. Because of the threshold, the number of points involved in the analysis for MPP-BLFL3 no SW and for
MPP-BLFL3 is lower than the other two algorithms, which always provide a PBL depth estimate. The statistical results (first four rows of Table 3) demonstrate that the MPP-BLFL3 algorithm does better than the other three methods. The new BLFL3 algorithm is therefore a clear improvement over that used in BW02, or the simpler BLFL1 algorithm, or the BLFL3 algorithm minus the spectral width input. In the next step we will focus on the outliers found for both sites and on the additional and optional step of using the temporal continuity test over a period of 1 day to further remove outliers.

d. Temporal continuity test

For this final step of the analysis, we will refer to Fig. 11 where MPP-BLFL3 results are compared to the experts’ PBL depth estimates at PIT and PYM. In particular, this figure still presents nine outliers (five of which are for PIT and four for PYM), outlined with circles. We perform the analysis introduced in section 3e and in Fig. 12 we plot the dependence of the difference (DIFF), calculated between the value of each PBL depth estimation from the third-order polynomial fit to the remaining PBL depth estimations of the day under consideration, versus the distance D between the estimations presented in Fig. 11 for PIT and PYM combined together.

For both sites we find that a value of DIFF equal to 600 m (vertical line) would eliminate four outliers (circled in solid), while not rejecting too many good points (13 points in total among PIT and PYM). This choice still leaves five outliers (circled with a dashed line). Of the five outliers, one was fooled by the presence of ground clutter; the remaining four occurred in cloudy conditions, and two of these occurred during the first 2 h after sunrise. After this procedure we have the statistical results presented in Fig. 13 and in the last row of Table 3. Thus, with the time-continuity threshold the algorithm produces an absolute error of the mixing-depth estimates comparable to the vertical range resolution of the profilers, although at the possible expense of eliminating some data that were in good agreement with the expert estimations as well.

e. Analysis of the “missing values”

To this point the analysis has focused on a statistical comparison of boundary layer depths for cases when both the experts and the automated algorithm were able to provide BLD estimates. However, there are also many cases when neither the experts nor the algorithm provided BLD estimates, cases when the experts gave estimates but the algorithm did not, and finally cases when the algorithm gave estimates but the experts did not. Figure 14 shows the frequency at which each of these possibilities happen, as a function of the hour of the day. An analysis of the cases where either the experts or the algorithm failed to provide an estimate is undertaken to provide a more complete understanding of the skill of the algorithm.

There are three categories of situations for which the
expert BLD estimates will be missing: 1) neither expert was confident in the identification of the BLD, 2) only one of the two experts was confident in giving a BLD estimate and therefore the mean experts’ value was not computed, and 3) both the experts provided an estimate, but the difference between the two experts was larger than one standard deviation and therefore the mean experts’ value was not computed. The experts did not give a BLD estimate when the convective boundary layer was not well defined in the data. The following are some of the reasons that this happens: the turbulence is too weak to be detected by the profiler; the growth of the boundary layer is too rapid over the course of 1 h making it difficult to assign a single value to the depth; there is intermittent rain; the boundary layer is too shallow to be visible in the range of wind-profiler measurements; the turbulence intensity diminishes at the end of the day and the boundary layer rapidly collapses; intermittent clouds give BL depths that vary greatly from one hour to the next; and in some cloudy situations when the turbulence is well developed in and below the cloud but the reflectivity does not indicate a clear inversion layer.

Figure 14a shows the number of hourly data values when both the experts and the algorithm provide BLD estimates. Here we see that during the midday hours (1400–2100 UTC; 0900–1600 LST), close to 50% of the time both the experts and the algorithm identify the BLD, with fewer simultaneous estimates occurring in the early morning and late afternoon hours. The maximum number of values occurs during the late morning hours (1400–1500 UTC; 0900–1000 LST), when the boundary layer has grown to a significant depth. Later in the afternoon the number of values decreases slightly when convective BL clouds usually form, sometimes making it difficult to assign a PBL depth. Figure 14b shows those cases when neither the experts nor the algorithm provide a BLD estimation. Considering only daytime hours, there are a total of 558 such cases. These situations occur at nearly constant frequency through most of the day, indicating that the reasons for a missing BLD value (listed above) are diverse enough that they generate a roughly uniform distribution of missing values. There is however a slight minimum in the late morning followed by a small increase in missing values in the afternoon, again due to the increasing presence of clouds. Figure 14c presents the occurrence of cases when the experts are confident and the algorithm is not. There are a total of 159 such cases, and they occur at nearly the same frequency for any daytime hour. Most of these episodes occur when for a particular hour of the day the BLD is only marginally well defined, but the experts can visually rely on the previous and following hours when the BL may be better defined, and visually interpolate between them.

Finally, the most interesting situation is when the algorithm provides an estimate but the experts do not, presented in Fig. 14d. The total number of these cases is 229. This situation happens at a near-constant frequency through the morning and early afternoon, with a peak in the late afternoon hours (2100–2200 UTC; 1600–1700 LST). This is consistent with the fact that 1) in the late afternoon transition BL, large hour-to-hour changes in the BLD can occur, and the experts, being aware of the time of the day, are more conservative, while the algorithm behaves the same at all hours of the day; and 2) in the presence of late-afternoon intermittent clouds the experts are less likely to provide an estimate even for a well-defined clear-sky hour if in the preceding and following hours the data are confused by clouds.

We next examined time–height cross sections of SNR, vertical velocity, and spectral width for each of the 229 cases shown in Fig. 14d to see if we could make some assessment of the skill of the algorithm, despite the fact that initially the experts provided no BLD estimates. This is of course a “posterior-subjective evaluation,” and the results are merely qualitative. Of the 229 points, we identified 7.9% of the algorithm’s estimates to be clearly wrong, 14.8% of them to be uncertain (when the boundary layer depth could possibly have been at that height), and 77.3% of them to be, in retrospect, reasonably good estimates. The 7.9% of clearly wrong estimations is equivalent to 18 BLD data
points. Of these 18 values, 7 occurred in the first 2 h of the day after sunrise, when the BLD cannot be possibly as high as that selected by the algorithm, and 5 of them are estimates fooled by the presence of strong ground clutter. In the future we will investigate how to modify the algorithm to choose an "unreasonably" large value of BLD during the first few hours after sunrise. This additional constraint requires more extensive study of these situations for other sites and other period of the year, if we wish to avoid adding simple threshold values that may not be valid in other geographical locations or for wintertime periods, for instance. Ground clutter is a subject of serious investigation for postprocessing data analysis and improvements in its elimination would be beneficial for the BLD algorithm. Finally, we note that if we combine both categories when the algorithm provided a BLD estimate (Figs. 14a,d), the total number of estimates by the algorithm is 610, of which 3.8% would be outliers. If we further remove the nine points that occur in the first few hours after sunrise (seven of which come from the data in Fig. 14d, and two were among the five outliers found in the statistical comparison of section 5d), the percentage of outliers decreases to 2.3%.

6. Conclusions

An improved algorithm for the automatic determination of the convective boundary layer depth with wind-profiling radars has been developed and tested. The improved method is an extension of the fuzzy logic method presented in BW02, with several new components that have been evaluated separately. The evaluation used two sets of 915-MHz wind-profiler data, collected during the summer of 2004 at Pittsburgh, Pennsylvania, and Plymouth, Massachusetts. Approximately 50 days’ worth of profiler data were analyzed from each site. Improvements to the algorithm were evaluated by comparing the automated PBL depth estimates to independent boundary layer depth estimations made by human experts’ visual inspection of time–height cross sections of SNR, vertical velocity, and spectral width of the vertical velocity.

First, the improved algorithm uses the Väisälä multiplex procedure (MPP) for the computation of the moments of radar spectra, rather than the separate fuzzy logic (FLPP) method used previously. Boundary layer depths computed from the MPP moments are shown to be more accurate than those using FLPP or the standard peak picking (SPP) method used in the Väisälä wind-profiler software.

Second, the algorithm for the PBL depth estimation has been modified to include the Doppler radar spectral width parameter in the fuzzy logic part of the code. The algorithm was also modified regarding the way in which it incorporates the vertical velocity variance so that it can also be applied to cloudy boundary layers.

Third, we used the score assigned at the end of the fuzzy logic process as a reliability value for the estimation of the PBL depth. The value of the score reflects how well each point in the profile matches the characteristics of being at the top of the PBL, with a higher value meaning a greater level of confidence in the PBL depth estimate. Through a statistical analysis of two profiler datasets, we identified a value of the reliability threshold that successfully eliminated many outliers without also rejecting too large a number of correct estimates.

Fourth, as an additional and optional step, a temporal continuity test has been added to the algorithm that is applied during a period of 1 day. If the user does not require PBL depths in real-time each hour, then this part of the code can make the estimations more precise.

A statistical comparison of the new algorithm with the previous BW02 algorithm, which only uses \( C_n^2 \) and the vertical velocity variance, and with a simpler test using only the maximum value of \( C_n^2 \) for the boundary layer depth, demonstrated the superiority of the new algorithm. A further test in which the spectral width data were removed from the new algorithm demonstrated that inclusion of the spectral width data adds skill to algorithm. The new algorithm provides very accurate estimates of the boundary layer depth, with an absolute error comparable to the vertical range resolution of the profilers. To achieve this level of accuracy, the algorithm is by necessity quite conservative, producing valid estimates only for approximately 75% of the hours during which the experts also agreed on a PBL depth value. In part this is due to the fact that the algorithm has been developed to work in both clear and cloudy boundary layers, which have different radar characteristics. Inclusion of independent measurements of the presence of clouds, for example, with a ceilometer or a pyranometer, could allow for separate algorithms to be used in clear-sky and cloudy conditions, which may allow for a higher percentage of the PBL depths to be accurately estimated. This will be investigated in a future study.

Acknowledgments. The authors thank three anonymous reviewers and the editor for their constructive comments. This research was partially funded by NOAA’s Health of the Atmosphere program.

REFERENCES


