

# WAVELET ALGORITHM FOR THE ESTIMATION OF MIXING LAYER HEIGHTS WITH CEILOMETERS

G. Teschke<sup>1</sup>, J. Reichardt<sup>2</sup>, D. Engelbart<sup>2</sup>

<sup>1</sup>*Hochschule Neubrandenburg - University of Applied Sciences, Brodaer Str. 2, 17033 Neubrandenburg, Germany, E-mail: teschke@hs-nb.de*

<sup>2</sup>*Meteorologisches Observatorium Lindenberg, Deutscher Wetterdienst, Am Observatorium 12, 15848 Lindenberg, Germany, E-mail: jens.reichardt@dwd.de, dirk.engelbart@dwd.de*

## ABSTRACT

An algorithm has been developed for the retrieval of mixing layer heights (MLH) from ceilometer, or lidar, measurements. The technique essentially relies on the computation of gradient information by use of the wavelet transform. The algorithm is described and a measurement example is presented.

## 1. INTRODUCTION

Currently, two major types of algorithms for the determination of the MLH from backscatter profiles measured by a lidar or ceilometer exist. The first one, which is sometimes called the peaks technique, is based on the analysis of the gradient and the variance of the backscatter profile. The second one, the wavelet method, uses the wavelet transform [1] of the backscatter profile. In principle, MLH detection is a pattern recognition problem. The basic assumption which is usually made is that the vertical distribution of aerosol can be used as a tracer for finding boundaries. The absolute value of the backscatter is typically not needed since the relevant information seems to be completely coded in the gradient (but possibly of different orders) of the backscatter profile. The peaks algorithm was invented for the detection of the top of multiple aerosol layers from backscatter profiles (see, e.g., [2]). The method is based on the analysis of collocated minima of the backscatter gradient and maxima of the backscatter variance. The wavelet algorithm has gained a great amount of popularity during the last years (see, e.g., [3]). Typically, the Haar wavelet transform is used because it is easy to implement and a powerful gradient locator and therefore a very promising mathematical tool; but recently published wavelet-based methods do not take full advantage of the wavelet theory. Here we present first results of our efforts to develop an advanced version of the wavelet algorithm and, thus, a reliable MLH detector.

## 2. PREPROCESSING OF CEILOMETER DATA

### 2.1. Cloud Detection

Because cloud-free backscatter profiles are the basic requirement for MLH detection, a cloud-detection algorithm is needed.

In a first step, the backscatter profile  $\beta$  is re-normalized by its variance  $\sigma$  (as taken from the ceilometer output

data file). By doing so, backscatter signals are increased in measurement cases in which clouds are likely present:

$$\beta_{\sigma}(h, t) = \beta(h, t)\sigma(t), \quad (1)$$

where  $h$  denotes the height and  $t$  denotes the time variable ( $N_h$  height bins,  $N_t$  profiles). Further, we define the global mean and variance:

$$\mu = \frac{1}{N_h N_t} \sum_h \sum_t \beta_{\sigma}(h, t), \quad (2)$$

$$\Sigma = \frac{1}{N_h N_t - 1} \sum_h \sum_t [\beta_{\sigma}(h, t) - \mu]^2. \quad (3)$$

Our studies show that a simple empirical estimate of a pixel-wise cloud threshold is given by:

$$e_C = \mu + 3\sqrt{\Sigma}. \quad (4)$$

Based on  $e_C$  a simple indicator function can be introduced:

$$I(h, t) = \int_0^h \chi_{\{\beta_{\sigma}(\xi, t) > e_C\}}(\xi, t) d\xi \quad (5)$$

where  $\chi$  is the characteristic function with  $\chi(\xi, t) = 1$  for  $\beta_{\sigma}(\xi, t) > e_C$  and  $\chi(\xi, t) = 0$  elsewhere. Therefore, at locations  $(h, t)$  where  $I(h, t) = 0$  we may assume that no clouds are present. The integration/summation with respect to  $\xi$  implies, that the first cloud layer marks the maximum height up to which MLHs will be detected. Hence, we may define cloud-free backscatter data  $f$  by

$$f(h, t) = \beta(h, t)\chi_{\{I(h, t)=0\}}(h, t), \quad (6)$$

with  $\chi(h, t) = 1$  for  $I(h, t) = 0$  and zero elsewhere. The filtered ceilometer data  $f(h, t)$  is the basis for all subsequent signal processing.

### 2.2. Temporal and Spatial Averaging

The ceilometer data  $f(h, t)$  are averaged in time and space. For temporal averaging, every  $k$  backscatter profiles are summed (thus reducing the number of profiles to  $N_t/k$ ). Spatial smoothing is accomplished with a sliding-average length of  $l$  height bins (for ceilometer data with a bin width of 15 m,  $l = 20$  is typical). The averaged data,  $F(h, t)$ , are then used for the wavelet analysis.

### 3. WAVELET ANALYSIS AND MLH DETECTION

The final step of the MLH detection procedure is the gradient analysis of  $F(h, t)$  of each individual spatial measurement, i.e. for each individual  $t$ . We define the wavelet transform as

$$W_\psi F(h, a; t) = \frac{1}{\sqrt{c_\psi}} \int F(\gamma, t) \frac{1}{\sqrt{a}} \psi \left( \frac{\gamma - h}{a} \right) d\gamma, \quad (7)$$

where the normalization constant  $c_\psi$ , given by

$$0 < c_\psi = 2\pi \int \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \quad (8)$$

reflects the constraint on the analyzing wavelets to produce an invertible and isometric transform. The isometry property is relevant since it ensures norm (or  $L_2$ -energy) equality between the signal  $F$  and its wavelet transform (allowing a physical interpretation of the wavelet-transformed signal).  $\hat{\psi}$  denotes the Fourier transform of the analyzing wavelet.

In recent publications, only the Haar wavelet transform has been considered. This was motivated by the fact that the Haar wavelet transform is nothing but the difference quotient at different scales  $a > 0$ . For  $a \rightarrow 0$ , the Haar wavelet transform converges to the derivative of the analyzed signal. This permits a direct interpretation of the wavelet transform, but is in principle just a different way of computing the difference quotient as done in classical approaches. However, the novelty (and the advantage) is to obtain the difference quotient of all desired scales by computing the wavelet transform once.

Remarkably, if  $\psi$  is chosen such that  $\psi = \phi^{(k)}$  (i.e.,  $\psi$  is defined via a derivative of some absolutely and quadratically integrable function  $\phi$ ,  $\phi \in L_1 \cap L_2$ ), it follows that

$$\lim_{a \rightarrow 0} W_\psi F(a, h; t) = C F^{(k)}(h), \quad (9)$$

which allows one to analyze also the behavior of derivatives of  $F$  of higher order, and this can give a hint at where the gradient growth (or decay) becomes maximum or minimum ( $C$  is a constant).

Our algorithm uses the Daubechies wavelet family (1st, 2nd, 3rd order) [4], thus yielding more structural information on the signals to be analyzed. In order to detect MLHs, we evaluate the wavelet spectrum at all scales simultaneously. This can be easily done by temporal spectrograms,

$$S(h, t) = \int |W_\psi F(a, h; t)|^p da, \quad (10)$$

with  $p > 0$ , which can alternatively be replaced by:

$$S(h, t) = \left( \int |W_\psi F(a, h; t)|^p \text{sign}[W_\psi F(a, h; t)] da \right)_+ \quad (11)$$

Equation (11) additionally takes into account the sign of the gradient information (which is, for our purpose, rather important). Finally, since we are interested in local maxima, the negative part can be skipped (which is denoted by  $(\cdot)_+$ ).

In general, the evaluation of  $S(h, t)$  will not yield a single, unique MLH but a set of MLH candidates. At this point, as will be discussed below, empirical MLH criteria (e.g., the range of gradients), or data from other instruments, are required to select the most probable MLH. So for maximum flexibility, we allow each profile to have  $N_{\text{MLH}}$  possible MLH events: For each  $t$  we consider  $S(h, t)$  as a function of  $h$  only. Then, we derive the  $N_{\text{MLH}}$  maximum values

$$V_{\text{MLH}}(\cdot, t) = \max_{1, \dots, N_{\text{MLH}}} S(\cdot, t) \quad (12)$$

and the corresponding  $N_{\text{MLH}}$  positions

$$P_{\text{MLH}}(\cdot, t) = \arg \max_{1, \dots, N_{\text{MLH}}} S(\cdot, t). \quad (13)$$

For each  $t$  we thus obtain two corresponding vectors containing at most  $N_{\text{MLH}}$  candidates for MLHs ( $p$ -averaged gradient values and spatial location; for different choices of  $p$ , the gradient is amplified or attenuated; currently, we use  $p = 1$ ). To attribute a quality index to each of these MLH candidates, we use the ‘‘noise’’ floor information  $f_{\text{noise}}$  (as determined in the spatial-averaging process). In the presence of clouds, the noise contains structured patterns (which typically manifests itself in singularly appearing larger signal amplitudes). Therefore, the local signal variance of  $f_{\text{noise}}$  around the MLH locations stored in  $P_{\text{MLH}}$  must differ compared to the total variance of  $f_{\text{noise}}$ . This motivates the definition of the following variance spot indicator at location  $P_{\text{MLH}}(l, t)$ :

$$\Sigma_m[P_{\text{MLH}}(l, t)] = \int_{P_{\text{MLH}}(l, t) - m}^{P_{\text{MLH}}(l, t) + m} \left[ f_{\text{noise}}(h, t) - \int_{P_{\text{MLH}}(l, t) - m}^{P_{\text{MLH}}(l, t) + m} f_{\text{noise}}(\xi, t) d\xi \right]^2 dh, \quad (14)$$

where  $t$  is the particular time step and  $l$  the specific MLH candidate. This quantity is then compared to the global variance of  $f_{\text{noise}}(\cdot, t)$ :

$$\Sigma(t) = \int \left[ f_{\text{noise}}(h, t) - \int f_{\text{noise}}(\xi, t) d\xi \right]^2 dh, \quad (15)$$

which leads to a quality index for the MLH candidates:

$$P(l, t) = |\Sigma_m[P_{\text{MLH}}(l, t)] - \Sigma(t)|^q \quad (16)$$

for some  $q \geq 1$ . If  $P(l, t)$  is small, the local and global noise characteristics are nearly the same [no rapid gradient (flows) are within the wavelet transform]. In summary, values of  $P(l, t)$  close to zero indicate a large likelihood of a MLH events whereas larger values indicate

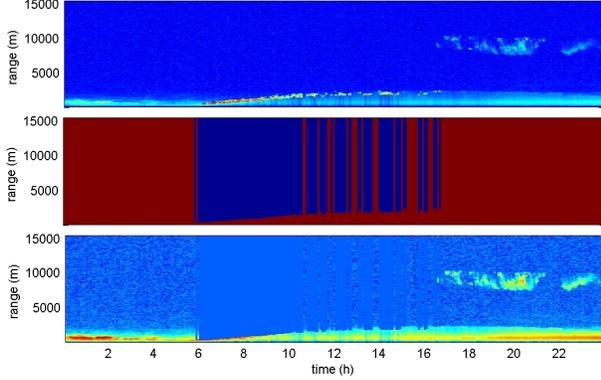


Figure 1: MLH retrieval for 8 July 2007, filtering of the raw data. From top to bottom, re-normalized backscatter data  $\beta_\sigma$  [Equation (1)], characteristic function  $\chi$  of Equation (6), and filtered ceilometer data  $f$  [Equation (6)].

unlikely MLH events, or something completely different (e.g., clouds). To obtain an intuitive from-zero-to-one scale of decision numbers, we define

$$\hat{P}(l, t) = \frac{1}{1 + P^2(l, t)}. \quad (17)$$

The function  $\hat{P}(l, t)$  gives an ordering of the MHL events by its quality index. The most likely MLH at time  $t$  is then the height with the maximum value of  $\hat{P}(\cdot, t)$ :

$$\text{MHL}(t) = \arg \max_{l \in \{1, \dots, N_{\text{MLH}}\}} \hat{P}(l, t). \quad (18)$$

#### 4. EXAMPLE

We apply our MLH detection algorithm to ceilometer data obtained with the CHM-15K ceilometer [5] manufactured by Jenoptik, Germany, which is the new standard ceilometer of the German Meteorological Service operational network. With its powerful diode-pumped Nd:YAG laser and photon-counting receiver electronics it is capable of detecting optically thin aerosol layers and cirrus clouds. The data used to demonstrate MLH detection is taken from the CHM-15K instrument which is operating at the Meteorological Observatory in Lindenberg and which is co-located with the water vapor Raman lidar RAMSES.

Figure 1 visualizes the first step in the retrieval process, which is cloud detection. The first six hours of 8 July 2007 showed cloud-free conditions with particle-backscattering confined to the boundary layer. Boundary-layer clouds started to develop at 06:00 UT, and were intermittently present until  $\sim 17:00$  UT. Base height increased over the course of time from near the ground to about 2000 m. After 17:00 UT an optically thin cirrus cloud was observed. Indicator function  $I$  [Equation (5)] and, hence, characteristic function  $\chi$  reliably filter out ranges of the backscatter profiles with and above clouds the ceilometer cannot penetrate.

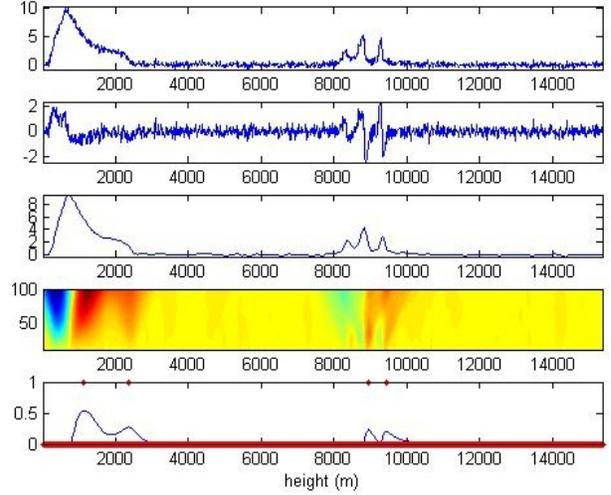


Figure 2: MLH retrieval for 8 July 2007, determination of MLH candidates at 18:10 UT. From top to bottom, backscatter profile  $f$ , noise floor  $f_{\text{noise}}$ , averaged backscatter profile  $F$ , wavelet transform  $W_\psi F$  [Equation (7)], and temporal spectrogram  $S$  [Equation (11)] with detected MLH candidates (red dots at  $y = 1$ ).

The filtered ceilometer data are then used for the wavelet analysis presented in Figure 2. The top panel shows a backscatter profile in raw-data resolution (15 m and 30 s). A profile at around 18:10 UT was selected to demonstrate the effect of aerosols and clouds on the noise level (as determined in the spatial-averaging process), which is shown in the panel below. Note the pronounced features of  $f_{\text{noise}}$  that are related to the cirrus cloud. These features are later in the retrieval process used to assign a low quality index to MLH candidates at cloud altitudes.

The averaged backscatter profile and its wavelet transform are shown in the third and fourth panel of Figure 2, respectively. The Daubechies wavelets of order one were used here. Results for scales  $a$  between 10 and 100 are shown, small values are coded in blue colors, large values in red colors. Finally, the bottom panel highlights the temporal spectrogram of the wavelet transform. Because Equation (11) is used, only heights with dominantly positive wavelet transform values produce significant spectrograms. The number and positions of MLH candidates are derived from the spectrograms by Equations (12) and (13), respectively. In this example, four MLH candidates are found.

Figure 3 highlights the temporal evolution of the MLH candidates over the course of the full day. In the top panel, MLH candidates are superimposed over the ceilometer raw data, in the panel below MLH candidate traces are color-coded according to MLH heights, with the blue color for the MLH candidates closest to the ground, the green color for the second-lowest, and so on. In the boundary layer, there are only one or two MLH candidates for most of the time (third and fourth MLH

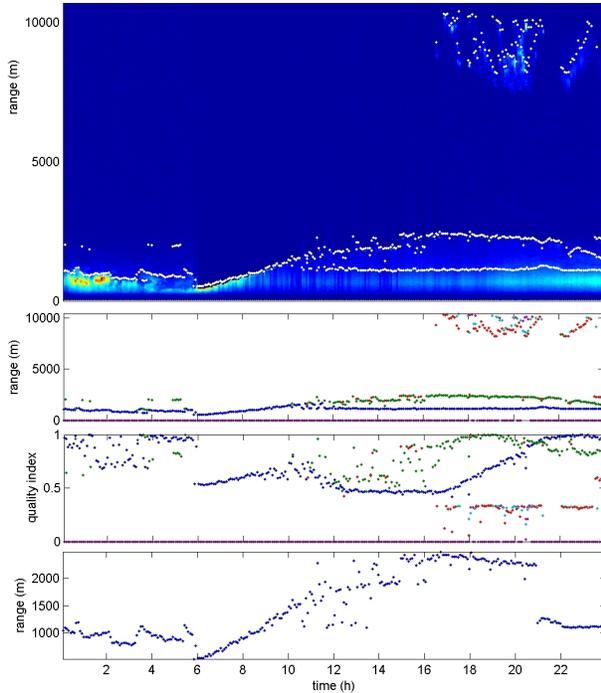


Figure 3: MLH retrieval for 8 July 2007, retrieval of MLH optimum estimates. From top to bottom, ceilometer raw data with superimposed MLH-candidate positions  $P_{MLH}$  [Equation (13), white diamonds], MLH-candidate positions  $P_{MLH}$ , quality index  $\hat{P}$  [Equation (17)], and most likely MLH. The algorithm detects up to four MHL candidates for each time step (color-coded symbols: blue, green, red, cyan).

candidates are mostly restricted to cirrus altitudes). With the help of the quality index one can now select the MLH candidate that most likely is the correct MLH. As the two bottom panels of Figure 3 show, the most likely MLH is either the MLH candidate closest, or second-closest, to the ground, MLH candidates at cirrus levels have low quality index values and are rejected without exception. Before 11:00 UT, there is virtually only one MLH candidate. Around that time, however, a second one appears which is causing scatter in the most likely MLH in the bottom panel. While the lower MLH candidate is stationary at heights around 1200 m, the higher boundary-layer MLH candidate is trending upwards, reaching altitudes of up to 2500 m. Until 18:00 UT, the quality index of the latter is significantly larger than that of the former, but afterwards the spread diminishes. Around 20:45 UT, the quality index of the lowest MLH candidate even surpasses the one of the next-higher candidate, causing the most likely MLH to drop abruptly and significantly. However, the quality indices of the two candidates remain both at high values and close to one another until midnight, so the retrieval results can be considered ambiguous here.

In Figure 4 the results of the MLH retrieval are shown together with the water-vapor and aerosol fields as measured with RAMSES. This lidar is currently only operat-

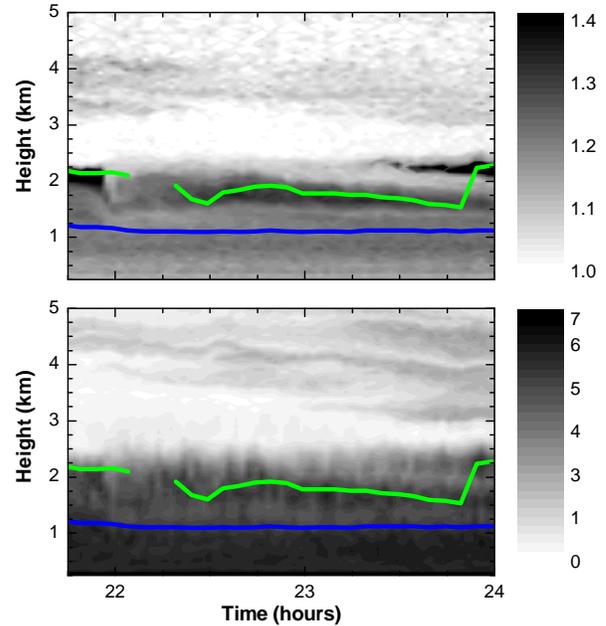


Figure 4: Height versus time display of backscatter ratio at 355 nm (top) and water vapor mixing ratio (in g/kg; bottom) measured with the water vapor Raman lidar RAMSES on 8 July 2007. Height resolution is 60 m, integration time is 120 s. The temporal evolution of the two MLH candidates with highest quality index are shown for comparison (blue and green curves, cf. Figure 3).

ing at nighttime, so RAMSES data are not available before 21:45 UT. The MLH candidate second-lowest to the ground follows nicely layered features of the aerosol and water-vapor fields, while this is not obvious in the case of the lowest MLH candidate. In summary, the first results obtained with our MLH detection algorithm are promising, but the quality of the retrieval has to be assessed in future studies. Questions that have to be addressed in this context evolve around the definition of MLH in general, and the threshold and parameter values used in the retrieval algorithm.

## REFERENCES

1. Louis, A. K., Maaß, P., and Rieder, A. *Wavelets*. Teubner, Stuttgart, Germany, 1998.
2. Menut, L., Flamant, C., Pelon, J., and Flamant, P. H. 1999: Urban boundary layer height determination from lidar measurements over the Paris area. *Appl. Opt.*, **38**, pp. 945-954.
3. Brooks, I.M. 2003: Finding boundary layer top: application of a wavelet covariance transform to lidar backscatter profiles. *J. Atmos. Ocean. Technol.*, **20**, pp. 1092-1105.
4. Daubechies, I. *Ten Lectures on Wavelets*. SIAM, Philadelphia, U.S.A., 1992.
5. [http://www.jenoptik-los.com/data/downloads/521/CHM\\_15k\\_en.pdf](http://www.jenoptik-los.com/data/downloads/521/CHM_15k_en.pdf)