



# Pathfinder: Applying graph theory for consistent tracking of daytime mixed layer height with backscatter lidar

Marco de Bruine<sup>1,2</sup>, Arnoud Apituley<sup>1</sup>, Dave Donovan<sup>1</sup>, Henk Klein Baltink<sup>1</sup>, and Marijn de Haij<sup>1</sup>

<sup>1</sup>Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, 3731 GA De Bilt, The Netherlands

<sup>2</sup>Institute for Marine and Atmospheric Research (IMAU), Utrecht University, Utrecht, the Netherlands

*Correspondence to:* Arnoud Apituley (apituley@knmi.nl)

**Abstract.** The height of the atmospheric boundary layer or mixing layer is an important parameter for understanding the dynamics of the atmosphere and the dispersion of trace gases and air pollution. The height of the mixing layer (MLH) can be retrieved, among other methods, from lidar or ceilometer backscatter data. These instruments use the vertical backscatter lidar signal to infer  $MLH_L$ , which is feasible because the main sources of aerosols are situated at the surface and vertical gradients are expected going from the aerosol loaded mixing layer close to the ground to the cleaner free atmosphere above. Various lidar/ceilometer algorithms are currently applied, but accounting for MLH temporal development is not always well taken care of. As a result,  $MLH_L$  retrievals may jump between different atmospheric layers, rather than reliably track true MLH development over time. This hampers the usefulness of  $MLH_L$  time series for e.g. process studies, model validation/verification and climatology. Here, we introduce a new method ‘Pathfinder’, which applies graph theory to simultaneously evaluate timeframes consistent with scales of MLH dynamics, leading to coherent tracking of MLH. Starting from a grid of gradients in the backscatter profiles, MLH development is followed using Dijkstra’s shortest path algorithm (Dijkstra, 1959). Locations of strong gradients are connected under the condition that subsequent points on the path are limited to a restricted vertical range. The search is further guided by rules based on presence of clouds and residual layers. Applied to backscatter lidar data from Cabauw, excellent agreement is found with windprofiler retrievals for a 12-day period in 2008 ( $R^2=0.90$ ) and visual judgment of lidar data during a full year in 2010 ( $R^2=0.96$ ). These values compare favourably against other  $MLH_L$  methods applied to the same lidar data set and corroborate more consistent MLH tracking by Pathfinder.

## 1 Introduction

The atmospheric boundary layer is the lowest part of the atmosphere where most of the interactions between surface and atmosphere take place. Knowledge of the processes and mechanisms in this layer is essential in meteorology and climate science. The height of the boundary layer, or mixing layer height (MLH) is an important parameter; it affects e.g. near-surface air quality, since it limits the volume of air into which pollutants are emitted, mixed and dispersed and therefore crucial in modelling pollution and smog and dispersion of greenhouse gases. Since the height of the mixed layer is driven by turbulent processes (Stull, 1988), it is one of the parameters that can be used to test model representation of the energy balance against



observations. MLH observations therefore, are of key importance to test models for realistic representation of the atmosphere ranging from short timescales (weather forecasting) to long timescales (climate change).

Several methods exist to measure MLH. Instruments used for this include backscatter lidar (Van Pul et al., 1994; Menut et al., 1999; Eresmaa et al., 2006; Haeffelin et al., 2012; Wiegner et al., 2006), Doppler lidar (Harvey et al., 2013b), radiosonde  
5 (Holtzworth, 1964; Vogelesang and Holtzlag, 1996; Zilitinkevich and Baklanov, 2002), sodar (Beyrich, 1995) and windprofiler (Angevine et al., 1994), to name just a few. From these instruments, the ceilometer has arguably the best combination of the number of installations, thanks to the deployment in many national meteorological networks and at aerodromes (Thomas, 2016) and excellent temporal/spatial resolution – typically 24/7 coverage of vertical profiles at time resolutions better than one minute and vertical resolutions in the order of 10 m.

10 Backscatter lidars use aerosol concentration as tracer for MLH, which is possible because the main sources of aerosols are situated at the surface. The turbulent motions in the mixing layer cause the aerosol concentrations to be relatively well-mixed within the layer, while exchange between mixing layer (ML) and free atmosphere (FA) is limited. As a result, higher aerosol concentrations are found within the mixing layer close to the ground and lower aerosol concentrations are found in the free atmosphere above. This difference in aerosol concentrations can be used to detect MLH.

15 Although the detection of the boundaries between atmospheric layers based on aerosol gradients has been used for many years, most of the backscatter lidar based techniques have difficulties to coherently track the MLH over time and may inadvertently jump between different atmospheric layers, such as a residual layer. Most algorithms search for the strongest gradient in a certain time window to which the MLH is then assigned, and the next time interval is treated independently. As a result MLH<sub>L</sub> retrievals may jump between different atmospheric layers. This limits the utility of the measurements for statistics  
20 and climatologies derived from these data records, as well as the use for process studies and model validation/verification.

To alleviate this problem, guidance can be sought from ancillary data (e.g. Pal et al. (2013)). This improves MLH retrieval at the cost of additional infrastructure and applicability will be limited to only those locations where the complete set of instruments is available. This also counteracts the attraction of the simplicity of the stand-alone backscatter lidar/ceilometer and to integrate these instruments into a network to monitor MLH over large, regional, areas. Such approaches are discussed  
25 in e.g. the COST action TOPROF (Illingworth, 2016) which aims to create an operational ground based profiling with ceilometers, lidars and microwave radiometers for improving weather forecasts and E-PROFILE (EUMETNET, 2016) in which a framework is developed to exchange lidar/ceilometer backscatter data.

Here, we describe the new algorithm ‘Pathfinder’ to track the development of the MLH during the day, based solely on single wavelength backscatter lidar data. The Pathfinder algorithm is based on graph theory and the algorithm published by  
30 Dijkstra (1959) for finding the shortest path in graphs, and therefore inherently takes temporal development into account. Section 2 introduces the Pathfinder method and gives a description of the different steps followed. Section 3 describes different instruments used for boundary layer observations as well as the methods to derive MLH that are used in this paper for validation and verification. Section 4 presents results from this new method applied to data from Cabauw. The section is concluded with a sensitivity test.



## 2 Methodology

### 2.1 Current lidar methods

The most commonly used techniques to detect MLH from lidar measurements are the gradient method (e.g. Van Pul et al. (1994); Flamant et al. (1997); Menut et al. (1999)), variance analysis (e.g. Hooper and Eloranta (1986); Menut et al. (1999))  
5 and continuous wavelet transforms (e.g. Haij et al. (2007)). Doppler lidar techniques also use aerosol backscatter, but have additional information about the velocity of particles (Harvey et al., 2013a).

Abovementioned measurement techniques all evaluate measurements per timestep individually. Different features, like the residual layer and advected aerosol layers can cause additional strong gradients in the lidar signal besides the MLH. These additional gradients can be of same order of magnitude as the backscatter gradient on the MLH or even stronger, possibly  
10 leading to a MLH estimate alternating between these different layers. However, MLH is a relatively slowly evolving quantity and large differences between subsequent MLH estimates can be rejected. This can be accomplished by processing information from both the spatial and temporal domain simultaneously.

### 2.2 Input lidar data

The lidar data used is the range corrected lidar signal (*RCS*). This is the signal that results after subtraction of the background  
15 sky light and correcting for the  $z^2$  geometrical term in the lidar equation:

$$\begin{aligned} RCS(z) &= z^2 P(z) \\ &= z^2 \cdot O(z) \cdot \beta(z) \cdot \exp\left(-2 \int_0^z \alpha(z') dz'\right), \end{aligned} \quad (1)$$

where  $\beta(z)$  is the backscatter profile in  $1/(m \cdot Sr)$ ,  $\alpha(z)$  is the extinction coefficient in  $1/m$  and  $O(z)$  is the overlap function between the receiving telescope field of view (FOV) and the area illuminated by the outgoing laser beam. For the MLH  
20 retrieval, we will only be looking at the uncalibrated properties of the lidar signal, and in particular, the vertical gradient in *RCS*. Calibration constants therefore cancel. The overlap function however is important, as it increases from zero to unity from the ground up to the altitude of full overlap:  $z_f$ . Above  $z_f$  we assume  $O(z > z_f) = 1$ . In the region of incomplete overlap below  $z_f$ , the gradient in *RCS* will be influenced by  $O$ , by adding a systematic positive contribution to the gradient. A further consequence of the lidar instrument geometry is that no lidar signal is obtained at very close ranges below  $z_b$ . In the lowest  
25 range between the instrument and  $z_b$ , the lidar is essentially blind.

### 2.3 The Pathfinder method

The Pathfinder method exists of three stages. First, parts of the data will be excluded as MLH based on four different guiding restrictions. In the second stage, the core of the algorithm, the lidar data is transformed into a mathematical graph where



all measurement points are represented by a node (vertex) and the connections (edges) are constructed based on the actual measurement values. Dijkstra's (1959) algorithm is then applied to find the combination of measurement points representing the temporal evolution of the MLH. In the last stage a quality label is assigned to the MLH estimates based on the ratio of the RCS above and below the MLH estimates. A flow diagram of the algorithm is given in Fig.1.

### 5 2.3.1 Guiding restrictions

Although the ML can exhibit a lot of different features, some structures are the same for all types of mixing layers. Using information of clouds, residual and advected layers, several restrictions are constructed to guide the algorithm to search for MLH. Because the ML is a phenomenon extending from the ground up, these the restrictions provide a maximum height to search for MLH.

10 The presence of clouds, in particular cumulus and stratocumulus clouds, can reveal the structure of the ML. Although there are numerous types of clouds and many mechanisms responsible for their formation, clouds only form above or on top of the ML. Whenever clouds are present, the maximum height of the MLH is restricted to the top of the first cloud layer. Cloud detection in the Pathfinder method is simple yet effectively based on the magnitude of the *RCS*, which is possible due to the high backscatter on cloud droplets compared to aerosols. For each timestep, the lowest altitude at which the *RCS* exceeds a  
15 prescribed threshold is marked as cloud base height (CBH). After that, the lowest altitude above the CBH at which the *RCS* drops below the same threshold is marked as cloud top height (CTH). The measurements above the CTH are then marked to be excluded from MLH search. To ensure inclusion of the cloud tops in the ML, a margin of 75 m above CTH is applied.

Note that CTH is an apparent cloud top height. For optically thick clouds, the lidar signal might not penetrate the complete cloud and the signal might drop below the threshold value before reaching the actual cloud top. Therefore, it is not completely  
20 certain CTH marks correct cloud top. Nevertheless, we do not limit the searching range to the cloud base, because, in case of shallow cumulus clouds, the air in the cloud is part of the boundary layer. The apparent cloud top marked as CTH is always closer to the correct cloud top and MLH than the cloud base.

Even though the gradient in *RCS* at the top of the ML can be weak or even absent at times, a strong gradient is always an indication of a separation between air masses. In case of multiple strong gradients, the searching range can be restricted to  
25 the altitude of the gradient closest to the surface. For the detection of a strong negative gradient the lowest altitude at which the negative vertical gradient is stronger than a certain threshold is marked. Similar to the cloud restriction, measurements higher than 75 m above the restriction are excluded. Due to inhomogeneities within the ML associated with the structure of rising thermals, the restricting altitude is taken as the maximum altitude of a strong gradient found within a time window of 2 minutes around a timestep. Next to that, a more strict threshold is used before the possible convective period to better capture  
30 the layered BL structures at the beginning of the day.

The presence of strong positive gradients are also indicative of separation of air masses. Like the restrictions based on negative gradients, the measurements 75 m above the lowest altitude the gradient exceeds a certain threshold are excluded. With the enhanced *RCS* on cloud droplets, this restriction also marks cloud bases. To prevent exclusion of ML clouds, interaction



with previously described cloud detection is needed. When a cloud base is found within a range of 300 m above a positive gradient, the searching range is not restricted on this argument.

From literature, the scales associated with the ML in the mid-latitudes are well-known. For example, the ML in The Netherlands rarely extends above about 750 m during nighttime and about 3 km during daytime. This is implemented into the algorithm by restricting the searching range during the morning period to the night value. After that, a linear increase of 2.5 ms<sup>-1</sup> is allowed, until the maximum altitude for daytime is reached. For the remainder of the day, the search range is limited by this daytime maximum.

### 2.3.2 Applying graphs

Tracking of the evolution of MLH is actually selecting a series of points, with corresponding time and altitude from a dataset, based on certain criteria, much like planning an optimal route on a map. The mathematical representation of information in graphs are an excellent tool for this purpose. To apply graphs for MLH tracking from lidar measurements, we define the following four steps. First, a graph is set up with each point in the dataset as vertices. Secondly, connections between these vertices are made so that any collection of connected vertices (hereafter called ‘path’) in the graph represents a physically possible MLH evolution. Third, weights are assigned to the connections. Finally, the graph is searched by Dijkstra’s shortest path algorithm to select the optimal path following the MLH.

To set up a graph, every point in the dataset is translated into a vertex, regardless of the actual values in the dataset. For a dataset with N timesteps and M range gates, this will lead to a graph with NxM vertices. This is illustrated in Fig.2.a In this stage, the graph is unstructured and the vertices unconnected.

To restrict possible paths in the graph to a physically sensible MLH evolution, connections between vertices represent specific conditions. The first condition is that vertices only connect to vertices of the next timestep. Connections to vertices of the same timestep, previous timesteps or more than one timestep apart are not allowed. These excluded connections are represented by the red arrows in Fig.2.b. Next, connections are only allowed to vertices within a restricted vertical range – represented by the additional orange and blue arrows in Fig.2.b. The blue connections are allowed, but the orange connections are not because these exceed the vertical range limit. In our implementation, the maximum allowed MLH growth rate is 2.5 ms<sup>-1</sup>, so the allowed range is 75 m when the time between measurements is 30 s. An additional restriction is posed on a timescale of 15 minutes. Over this time window, connections are not allowed to vertices exceeding a growth rate of 1 ms<sup>-1</sup>. This is higher than found by e.g. Baars et al. (2008) who reports growth rates of 1 km/hr or 0.278 ms<sup>-1</sup> only in extreme cases. However, this value does not seem to represent timescales resolving individual rising thermals. Implementing lower values caused problems in the Pathfinder algorithm to track fast MLH development in the morning. These conditions ensure connections lead to the construction of paths following the evolution of the MLH.

To determine which path represents the MLH from the collection created in the previous steps, weights are assigned to the connections in the graph. For Pathfinder, weights  $W$  are based on the vertical gradient in RCS:

$$W = - \left( \frac{\partial}{\partial z} RCS \right)^{-1} \quad (2)$$



Connections pointing to a certain vertex get a weight corresponding to the point in the RCS gradient represented by that vertex. Using this conversion, strong negative gradients in RCS correspond to low weights and weak gradients correspond to high weights. Consequently, the path with the lowest total weight will include the vertices with strongest negative gradients. The assignment of weight is shown in Fig.2.c where the colors blue, green, yellow and red represent the gradient in RCS. Blue being the strongest negative gradient, with green and yellow to intermediate and red representing the weakest negative gradient. Note that the conversion minus sign in Eq.2 is needed for positive weights for negative gradients, since Dijkstra's algorithm needs positive weights. Whenever negative weights occur, these are given a prescribed high weight instead.

In the final step, Dijkstra's shortest path algorithm is applied to select the optimal MLH path. This method selects the path with lowest total weight originating from a specific vertex in the first timestep to a vertex satisfying the abovementioned conditions in the last timestep of the dataset (Fig.2.d). The starting vertex is defined as the vertex corresponding to the strongest negative gradient in the first timestep of the dataset.

Theoretically, any dataset of arbitrary length can be translated into a graph and analyzed simultaneously. However, even with the relative efficiency of Dijkstra's algorithm, this is unpractical and would lead to long processing times. Therefore, it was decided to split data into multiple time windows and apply the method to these windows separately. This improves the processing times substantially, making the method available near-realtime MLH tracking. Whereas computational cost determines the upper limit of the window size, the lowest limit is determined by the typical timescale within the ML. For the results shown in Sec.4, measurements are combined in time windows with a size of 15 minutes, sufficiently large to capture several rising thermals simultaneously. For a continuous series of MLH estimates, the last timestep in a time window overlaps with the first timestep of the subsequent time window. The MLH estimate in this timestep is used as the starting point of the next time window.

### 2.3.3 Quality flagging

In the final stage of the Pathfinder algorithm, a quality flag is added to each point of the MLH estimates. The quality criterion is based on the ratio  $R$  of the  $RCS$  above and below the estimated MLH, similar to the quality flag used in de Haij et al. (2007) and Morille et al. (2007). The ratio is calculated as:

$$R = \frac{\overline{RCS}(z = MLH, MLH + 150 \text{ m})}{\overline{RCS}(z = MLH - 150 \text{ m}, MLH)}, \quad (3)$$

where  $\overline{RCS}$  indicates the average  $RCS$  over the indicated range interval.

Since we expect  $RCS$  to be smaller above MLH than below it,  $R$  is expected to be smaller than 1. If the algorithm marked an erroneous, noise induced gradient within the ML or FA, the backscatter above and below the estimate are comparable and the ratio is close to 1. However, if the top of the residual layer is marked as MLH or aerosol concentrations in the ML as low, this ratio can also be small. Consequently, a high ratio is an indication of an incorrect MLH estimate, but a small ratio is no guarantee the MLH is correct. As a threshold, ratios above 0.9 are considered an indication of an incorrect MLH estimate.



### 3 Instrumentation

#### 3.1 Cabauw Site

The instruments for this study were located at the Cabauw Experimental Site for Atmospheric Research (CESAR) (Apituley et al., 2008) in The Netherlands, where a suite of operational and research mode remote sensing and in-situ measurements are performed to comprehensively characterise the atmospheric state. At CESAR, multiple instruments are operated that can be used to validate and verify the MLH retrievals described here. The instruments and methods used are briefly introduced below.

#### 3.2 Backscatter lidar

The main instrument used for the development and testing in this study was a single wavelength backscatter lidar, the Leosphere ALS450, operating at 355 nm. This particular instrument also measures depolarisation (Donovan and Apituley, 2013), but this information is not used here. The instrument is running 24/7 and records an averaged lidar signal every 30 seconds at a spatial resolution of 15 m. The ALS450 has a maximum operational vertical range for clouds and aerosols of 15 km during nighttime and a reduced range of 8 km during the day due to remaining solar background. Main characteristics are tabulated in Tab.1. The region of incomplete overlap between the probing laser beam and the receiving telescope extends up to about 500 m. All data from the ALS450 is available from the CESAR data base (Baltink, 2016).

The atmospheric clear air attenuation is not entirely negligible at 355 nm, and adds a negative contribution to the gradient. However, we will neglect contribution to the gradient, as it is very smooth compared to the gradients from the aerosol backscatter.

Also installed at Cabauw is an Impulsphysik LD40 ceilometer, which is part of the Dutch ceilometer network, from which MLH is routinely derived (Haij et al., 2007). The LD40 operates at 905 nm and has a maximum operational vertical range for clouds of about 10 km. However, for aerosols the range is often limited to 2 km or less due to the low signal to noise ratio (SNR). Pathfinder was applied to LD40 data, but due to the limited SNR no useful intercomparison could be made to the ALS450 or other instruments. Currently, the LD40 is being decommissioned and replaced by the more capable Lufft CHM15k for the entire network in The Netherlands. However, the data could not be included in this study, but is the subject of future work.

#### 3.3 Wind profiler

The windprofiler/RASS (Radio Acoustic Sounding System) is a clear-air radar measuring Doppler shift to collect information on the vertical wind profile. The profiler at Cabauw operates at 1290 MHz. To measure the wind profile it executes a measuring cycle consisting of measurements in four different oblique directions (15.5° off-zenith) and one in the vertical. Combining the measurements from the different directions gives horizontal and vertical wind speeds. MLH detection from the windprofiler is possible since the intensity of the returning signal primarily depends on inhomogeneities in the atmospheric moisture and temperature caused by turbulence. The SNR is directly proportional to these inhomogeneities, captured in the refractive index



structure function  $C_n^2$  (Ottersten, 1969). Due to the entrainment at the MLH,  $C_n^2$  exhibits a (local) maximum at this height. The width of the Doppler spectrum is also used in the MLH detection. The spectral width is related to the differences in wind speed in the probed volumes, and is smaller in the entrainment layer than in the mixing layer itself and also increases again into the free atmosphere. So, MLH is associated with a minimum in the spectral width. Results for the  $MLH_W$  used in this paper use  
5 the method described by Angevine et al. (1994).

### 3.4 Radiosonde

Radiosonde data from Vaisala RS92-GDP measuring profiles of relative humidity, temperature, pressure and wind is used from the IMPACT campaign (Roelofs et al., 2010) that were launched from the Cabauw site. Additional routine radio sonde data are available from De Bilt, which is located about 20 km to the East of Cabauw. Several methods exist to calculate MLH from the  
10 radiosonde measurements ( $MLH_S$ ), including the parcel method (Holtzworth, 1964), lapse rate method (Hayden et al., 1997) and Richardson bulk method (Vogelezang and Holtslag, 1996). Here, we will use the Richardson bulk method since it takes into account both instability and wind shear.

## 4 Results

The performance of Pathfinder is demonstrated in a couple of case studies; under clear sky conditions in Sec.4.1, and in more  
15 complex atmospheres with clouds and precipitation in Sec.4.2. Subsequently, our new method is intercompared with MLH estimates from other instruments in Sec.4.3 and the same dataset is also processed by the STRAT2D method (Menut et al., 1999). Finally, to test the applicability of Pathfinder semi-operationally and to include a wide range of weather conditions, a comparison with manual MLH estimates for a full year is discussed.

The following data are all based on measurements of the Leosphere ALS450 at Cabauw, for which the Pathfinder tuning  
20 parameter values are listed in Tab.2.

### 4.1 Clear sky

The first case study is the evolution of  $MLH_L$  on 20 May 2010: a clear sky day with a well-developed afternoon ML. The measured backscatter and corresponding vertical gradients are shown in Fig.3. Also shown is the Pathfinder MLH estimate from sunrise to sunset. To point out different features of the algorithm, four regions will be highlighted: 1) the period between  
25 sunrise and the onset of convection, 2) the morning growth of the ML, 3) the plateau on the middle of the day and 4) the break-down of the ML and transition into the nighttime boundary layer. These regions can be seen in Fig.4.

From sunset it takes several hours before the surface temperature is high enough to produce convection visible in lidar observations. With the complete ML below the lidar detection range  $z_b$ , the solution tracks a gradient in stratification of the residual layer. Around 7 UTC, the ML rises above  $z_b$  and becomes visible to the lidar. The solution indicates the correct altitude  
30 as MLH. As the layer moves up and down, it sinks again below the detectable range, but when the layer reappears after some 30 minutes, the solution again indicates the correct height for MLH.



The greatest challenge in deriving MLH from lidar measurements is to separate the gradient associated with the ML from the gradients in the residual layer. The Pathfinder method will ignore additional layers when these are relatively far from the MLH. However, when an additional strong gradient exists close to the MLH, it might be included into the MLH estimate. The algorithms' decision depends on the proximity of the gradients and the ratio of their relative strength. For the solution to shift to a different layer it has to transition several points of weak gradients and receives a penalty for this in the form of a higher path weight. A relatively strong gradient on a residual layer can outweigh this penalty and still cause a shift in the path with the lowest total path weight. Fig.4b shows that for the major part of the morning transition the correct MLH is indicated. Only around 8:45 UTC the solution indeed triggers on a gradient in the residual layer, when the gradient on the real MLH becomes weaker.

Although individual rising thermals are also visible during morning growth, the differences between them are even more pronounced mid-day when the height of the ML is more or less stable. As shown in Fig.4c, the difference in altitude between subsequent thermals is in the order of 100-500 m, while the average MLH is nearly constant. The vertical limitation of 75 m between timesteps is enough to follow the differences between thermals, like the sudden decrease around 11:15 and 12:45 UTC.

## 4.2 Complex atmospheres

Even though a clear-sky day gives good insight in different ML features, completely cloud-free days are scarce in The Netherlands. The presence of clouds influences the evolution of the ML (e.g. by blocking incoming solar radiation) and causes additional gradients which can distract the algorithm from the correct solution. As an example for this, the next case study treats a day with abundant fair-weather cumulus clouds. As can be seen in Fig.5 two layers of clouds are present, mainly stratocumulus above cumulus clouds forming on top of the ML.

Around sunrise and sunset, the two cloud layers are well separated. Pathfinder correctly designates MLH at the top of the lowest cloud layer. This is mainly due to the guiding restriction that excludes the measurements above the first cloud layer. However, this is a broken cloud deck and the guiding restrictions cannot exclude the stratocumulus clouds for all timesteps. It is the combination of the guiding restriction together with the limited vertical searching range that ensures a correct tracking of the MLH even for broken cloud layers.

With the growth of the ML, the distance between the two cloud layers decreases up to a point where the two can no longer be distinguished when the signal extinction is too strong in the cumulus layer. During these periods, the solution tracks the top of the stratocumulus as MLH, leading to short peaks in MLH, e.g. around 10:00 UTC.

The limited vertical searching reach causes commitment to a cloud layer, beneficial for tracking broken cloud decks. But an example of the solution tracking the wrong cloud layer can be seen in Fig.6 between 15:45 and 16:45 UTC. The ML decreases in depth, but the cloud layer remains at the same altitude. The Pathfinder algorithm keeps tracking the strong gradients on the apparent cloud tops instead of the MLH. Because Pathfinder is designed to track gradients close to adjacent gradients, the algorithm will not divert from a cloud layer until either strong gradients or clouds trigger the guiding rules.

The last two cases consider the ML during precipitation events. An example of a day with (heavy) precipitation is 8 June 2010, seen in Fig.7. During and after precipitation, the lidar observations are affected by water on top of the instrument. Only



the strong backscatter on raindrops is strong enough to be distinguished from the background and noise. Therefore, a lidar based MLH estimate in these periods is unreliable if not impossible. Between 6:30 and 8:30 UTC the estimates fluctuate between 150 and 750 m, the complete range allowed by the guiding restrictions. During rain between 14:30 and 15:30 UTC, the signal is too low to calculate reliable MLH estimates.

- 5 Another day with frequent precipitation is 20 June 2010, seen in Fig.8. Characteristic of this day was that most of the precipitation evaporated before reaching the ground. Only 0.2 mm of rain was recorded at Cabauw that day. Without the observations affected by water on top of the instrument, this day gives insight in the effects of precipitation on the ML and the MLH solution of Pathfinder under these conditions. Evaporating precipitation can be grouped into two categories: evaporation in the lidar overlap region or well above it.
- 10 The period between 14:00 and 15:00 UTC is an example of precipitation reaching the overlap region. Because the raindrops evaporate on their way down, the highest droplet concentration and accompanying backscatter is strongest near the cloud base. Consequently, the *RCS* increases with altitude and no negative gradient is found until the apparent cloud top is reached. This altitude is then indicated as MLH.

15 Precipitation evaporating well above the overlap region can be seen in Fig.9b. Around 11:15 UTC, precipitation appears and with it the ML decreases in altitude. When the precipitation stops, air is no longer brought downward and the ML increases again in altitude. Under these conditions, the gradient on top of this layer is picked up well by the Pathfinder algorithm. Similar events can be observed between 11:45 and 12:00 UTC or 12:30 and 13:00 UTC.

### 4.3 Comparison to other methods

For comparison of the lidar methods Pathfinder and STRAT2D applied to the same data to radiosonde and windprofiler  
20 MLH retrievals, we used the observations from a 12 day-period in May 2008, obtained during the IMPACT campaign at Cabauw. Radiosonde observations were taken around 5, 10 and 16 UTC each day. The Richardson bulk method was used to estimate  $MLH_R$ . The windprofiler operated continuously during IMPACT, but  $MLH_W$  estimates could only be given for the above-mentioned 12 day-period. The ALS450 also operated continuously and both the Pathfinder and STRAT2D algorithms are applied to derive MLH estimates,  $MLH_{L,P}$  and  $MLH_{L,S}$  respectively. An overview of lidar measurements and the different  
25 MLH timeseries are shown in Fig.10.

#### 4.3.1 Windprofiler

The MLH estimates from windprofiler and Pathfinder algorithm are in excellent agreement, being within a vertical range of 100 m of each other for almost the complete period. Regression analysis shows a correlation of  $R^2 = 0.96$ . A few exceptions are found during mid-day. However, it is not apparent which MLH estimate is the better one. Next to that, Pathfinder could not  
30 track the MLH on some of the mornings because the ML was too shallow and stayed under the detection range for a large part of the morning.



### 4.3.2 Radiosonde

Agreement between radiosonde and Pathfinder strongly depends on the time of day. A good correlation  $R^2 = 0.90$  is found for the 16 UTC soundings, but is questionable at 5 and 10 UTC. At 5 UTC, the ML is often too low to be detected by lidar and cannot be tracked by Pathfinder, leading to a correlation of  $R^2 = 0.21$ . Comparing the 10 UTC soundings with the  
5 observations, it becomes clear that the Richardson bulk method often incorrectly designates the residual layer as MLH (e.g. 10 and 12 May). Investigation of these soundings shows that a theoretical adiabatically rising air parcel indeed has positive buoyancy almost up to the radiosonde estimate. The Richardson bulk method estimates MLH slightly higher because wind shear is taken into account as well. A well mixed-layer layer extending up to an altitude of 900 m can also be extinguished from the temperature profile corresponding to  $MLH_{L,P}$ . Based on this, it can be concluded that the Pathfinder estimate is correct and  
10 the Richardson bulk method is not. The reason may be that the Richardson bulk method does not take into account entrainment of environmental air into the rising thermal. By mixing in air with a lower (potential) temperature, the parcel loses its positive buoyancy at a lower altitude than the theoretical non-mixing rising parcel. Ignoring entrainment might be appropriate for deep convection, where the major part of the ascent takes place in non-turbulent air, but because of the turbulent nature of the ML, entrainment takes place at a much higher rate and cannot be neglected.

### 15 4.3.3 STRAT2D

For large parts of the 12-day period, the estimates by STRAT2D and Pathfinder agree well and also with the results of the windprofiler. But for STRAT2D large differences occur between subsequent timesteps when irregularities are found within the ML. This leads to unrealistic, erratic estimates of  $MLH_{L,S}$  evolution during the day under some conditions. Additionally, STRAT2D does not give an estimate due to the criterion that the ratio between backscatter above and below the MLH candidate  
20 cannot exceed a prescribed value. During the IMPACT campaign data used in this study, the ML air was relatively clean yielding too low backscatter values in the ML compared to the free atmosphere to fulfil the STRAT2D criterion.

## 4.4 Full year analysis

To quantify the performance of Pathfinder for a wide range of atmospheric conditions, it has to be applied to longer timeseries and validated against different methods, preferably based on multiple instruments. Continuous, collocated measurements  
25 needed for this comparison are scarce though, so the 12 UTC radiosonde observations from De Bilt were considered. However, a check of the manually derived MLH estimates for Cabauw against  $MLH_R$  showed that the radiosonde measurements appeared not representative for the MLH at Cabauw ( $R^2 = 0.64$  with an average error of 200 m). Also, the limitations of the Richardson bulk method as described in Sec.4.3.2 may be of influence here.

Therefore, manual estimates  $MLH_M$  at 12 UTC for all available days in 2010 were used. Lidar  $RCS$  and gradient fields from  
30 the ALS450 instrument were visually inspected to get an indication of weather conditions and evolution of ML during daytime. Plots zoomed in around 12 UTC were used to extract  $MLH_M$ . In 2010, ALS450 data was available on 336 days, and manual MLH estimates could be made for 252 days, limited by atmospheric conditions such as rain and fog or instrumental problems.



5  $MLH_M$  is compared to a 10-minute average of  $MLH_{L,P}$  around noon. The results are shown in Fig.11a. The correlation with the manual estimates is as high as  $R^2 = 0.90$ , with an average error of 76 m. If the Pathfinder quality criterion is used that flags data where the ratio between backscatter below and above the candidate  $MLH_{L,P}$  (i.e.  $RCS(z > MLH_{L,P})/RCS(z < MLH_{L,P})$ ) is larger than 0.9 a number of days is excluded, and 196 days remain where confidence is highest. This data set yields a correlation between  $MLH_M$  and  $MLH_{L,P}$  of  $R^2 = 0.96$ , with an average error of 52 m.

The STRAT2D results  $MLH_{L,S}$  were also compared  $MLH_M$ . STRAT2D only gives an estimate if the same quality criterion is used, which leaves 235 days, with a correlation of  $R^2 = 0.74$  and an average error of 127 m. The larger scatter compared to Pathfinder is apparent in Fig.11b. Most of the mismatches are caused by an overestimation of the MLH by STRAT2D. In these cases  $MLH_{L,S}$  designates the top of the residual layer as MLH.

10 The color coding in Fig.11 indicates the month and reveals the seasonality. As expected, lowest MLH estimates are found in winter and highest estimates in spring and summer.

#### 4.5 Parameter sensitivity

Pathfinder limits the search for MLH to altitudes near previously found MLH estimates, which gives it its improved tracking behaviour. However, the method may be sensitive to initialisation parameters. Two parameters in particular that may have an effect on the retrieved  $MLH_{L,P}$  time series are the start time in the day of the first timestep and the time window size, i.e. the number of timesteps treated simultaneously.

##### 4.5.1 Variation of the initial start time

Change of the position of the first timestep will not change the size of the search range, but shifts it along the time axis. As a consequence, the first MLH candidate found may be different from the one found in the default run and this may influence the subsequent MLH estimates since they are bound to the previous MLH estimates. The default time window is 15 minutes, i.e. 30 timesteps for the ALS450. In the test, the time window was shifted forward and backward by 1 to 30 timesteps, leading to a total of 61 retrievals for each case considered. The analysis was applied to the observations of May 2010. As a measure of stability of the algorithm we take the correlation of the  $MLH_{L,P}$  estimates of a full day of the sensitivity runs with the default run, averaged over the complete month. This correlation did not drop below  $R^2 = 0.96$ , meaning that 93% of the test runs match the results of the default run.

##### 4.5.2 Extent of the search time window

Within a time window, the vertical search range increases with 75 m both upward and downward each timestep with our current settings. If the window size is increased, the search range consequently expands and more measurements are included in the search.

30 Next to the default timewindow of 30 timesteps, calculations are made for windowsizes between 10 and 70 timesteps with increments of 10 timesteps. Again, the ALS450 observations from May 2010 were used. Average correlations are above



$R^2 = 0.975$ , except for a timewindow size of 10 timesteps which is slightly below  $R^2 = 0.95$ . Different causes are found for the decrease in correlation deviating from the default timewindow. For small timewindows, the limiting factor of maximum change of the MLH of  $1 \text{ ms}^{-1}$  between starting and endpoint of the time window becomes more important. These time windows can no longer correctly track differences between individual rising thermals. As a result, many small differences occur between these runs and the default calculations.

All sensitivity runs with a larger time window showed the lowest correlations for a single day all on May 27. Besides some small differences, this was due to the period between 14 and 16 UTC where a residual layer with clouds existed in close proximity to the MLH. Pathfinder has to include several high cost timesteps to allow for a jump to another layer. For relatively small time windows, there are not enough timesteps left to compensate for this extra cost and the jump is not permitted in the solution. In case of larger time windows, this might not be the case and the solution is allowed to jump to another layer more easily. In this particular case, the solution did not jump to the residual layer for window sizes up to 30 timesteps. This divergence continues as long as the tracked feature (e.g. residual or cloud layer) exists or the guiding restrictions pick up gradients forcing the calculations back to the correct solution.

Therefore, as long as the time window is sufficiently large to capture the rising thermals, the Pathfinder produces similar MLH estimates during a day irrespective of the time window settings. Although solutions can diverge for short periods, the guiding restrictions force the calculations back to the correct solution.

#### 4.6 Discussion

The clear-sky cases show that the guiding restrictions successfully exclude large parts of the additional gradients of the residual layer. This allows the shortest path algorithm to track the evolution of the MLH in high resolution. In case the guiding restrictions do not exclude all additional gradients, a jump to another layer of high gradients is most often prevented by the shortest path algorithm itself. For a transition, several measurement points with low gradients have to be included in the path, increasing the total cost of the jump. A transition cannot always be prevented if the additional gradients are strong enough to compensate for the extra cost. Here, further study would be needed on how to tune the algorithm settings. Also, atmospheric conditions different from the Dutch conditions may require different settings.

Clouds can be a good indicator for MLH height, but the high backscatter on the cloud droplets cause a negative gradient typically an order of magnitude larger than gradients associated with differences in aerosol concentration. Because of these strong gradients, the solution found by the algorithm is drawn to cloud tops. In case of a cloud layer above the ML, guidance is needed to restrict the algorithm to the MLH, for instance, by a more rigorous cloud screening prior to the Pathfinder analysis.

During rain  $MLH_{L,P}$  derivation is difficult, because of the attenuation of the lidar signal. If precipitation is not reaching the surface, it was visible that the down draft of a rain shower lowered MLH. Pathfinder is able to correctly follow this decrease in MLH as long as the precipitation evaporates above the overlap region  $z_f$ . Otherwise, there is no negative gradient to assign  $MLH_{L,P}$  to. The top of the precipitating cloud will be marked as  $MLH_{L,P}$  instead, although it is not the actual MLH.



Because of the high temporal resolution of the lidar observations, changes in MLH by individual thermals are dominant. The typical spatial scale of the thermals is in the order of several hundreds of meters up to a kilometer. To compare results to other instruments with a similar time resolution, their collocation should be better than this typical scale.

When the analysis is started, an initial point is needed. For this, climatology is used. Since the MLH at night in The Netherlands is often below the minimum overlap region of the AL450 used in this study, no attempts were made to derive the nocturnal boundary layer. Nocturnal boundary layers could be tracked with if the lidar profiles would start at appropriately low heights.

Whereas Pathfinder was developed specifically for MLH retrieval from stand-alone single wavelength backscatter lidar data, the method may be generalised to be applied to data from other profiling instruments for MLH retrieval. Moreover, Pathfinder may be embedded in a chain of multiple algorithms, such as cloud screening.

## 5 Conclusions

A common feature in the existing methods to derive MLH from backscatter lidar measurements based on gradient detection in aerosol loading is that MLH is derived for each timestep individually, which allows for large jumps in the MLH between subsequent timesteps. These jumps are not consistent with the inherent gradual evolution of the layer which usually makes it easy for a somewhat trained individual to visually recognise the MLH development in lidar *RCS* plots. To accommodate for this, a new method was proposed called 'Pathfinder'. This method evaluates multiple timesteps within a configurable time window simultaneously. Graph theory is applied together with Dijkstra's shortest path algorithm (Dijkstra, 1959) to imitate the continuous character of the MLH.

Pathfinder divides a full day of lidar measurements into time windows of 15 minutes and translates these into a graph with vertices representing each individual data point. To estimate MLH one point is chosen for each timestep. The vertical gradient corresponding to a certain measurement point is used to assign a weight to the vertex inversely proportional to the gradient. This way, the path with the lowest total weight (cost) will contain the maximum sum of strong gradients and will be a good estimate for the MLH. To mimic the gradual evolution of the MLH, the distance between subsequent points is restricted. The threshold used for this is a maximum growth rate of  $2.5 \text{ ms}^{-1}$ . Between the first and last timestep of a time window this threshold is lowered to  $1 \text{ ms}^{-1}$ . To guide the shortest path algorithm, a set of rules is used. Features like clouds, strong negative and positive gradients exclude parts of the observations from the searching range for MLH.

Pathfinder was applied to data from a Leosphere ALS450 deployed at the Cabauw Experimental Site (CESAR) in The Netherlands. The results were checked against MLH estimates obtained from independent observations, such as windprofiler and radiosondes. Excellent agreement was found between MLH estimates of the Pathfinder method and windprofiler ( $R^2=0.90$ ), but comparison to collocated radiosonde data was more problematic ( $R^2=0.21$ ), we believe, due to limitations in the Richardson bulk method. Pathfinder results were also checked against manual MLH retrieval ( $R^2=0.96$ ), as well as the results from a different algorithm, STRAT2D, applied to the same data.



In in this study, Pathfinder gives less scatter than STRAT2D in the comparison of a full year analysis against manual MLH retrievals. This is due to the jumps between layers present in the STRAT2D estimates.

The Pathfinder method can be used operationally on stand-alone single wavelength backscatter lidar data, provided the signal to noise ratio is sufficient to detect aerosol layers up to a few kilometers above ground. The typical computation time is less than five minutes for an observations of a full day, based on the Leosphere ALS450 data set. Application to other lidar instruments, such as the Lufft CHM15k as now being deployed in the Dutch operational observation network, is currently under investigation.

*Acknowledgements.* The STRAT development team is gratefully acknowledged for making the STRAT2D code available, for their support and fruitful discussions.



## References

- Angevine, W., White, A., and Avery, S. K.: Boundary-layer depth and entrainment zone characterization with a boundary-layer profiler, *Boundary-Layer Meteorology*, 68, 375–385, doi:10.1007/BF00706797, <http://dx.doi.org/10.1007/BF00706797>, 1994.
- Apituley, A., Russchenberg, H., van der Marel, H., Boers, R., ten Brink, H., de Leeuw, G., Uijlenhoet, R., Arbresser-Rastburg, B., and Röckmann, T.: Overview Of Research And Networking With Ground Based Remote Sensing For Atmospheric Profiling At The Cabauw Experimental Site For Atmospheric Research (CESAR) – The Netherlands, in: *Proceedings IGARSS 2008*, Boston, Massachusetts, vol. III, pp. 903–906, 2008.
- Baars, H., Ansmann, A., Engelmann, R., and Althausen, D.: Continuous monitoring of the boundary-layer top with lidar, *Atmospheric Chemistry and Physics*, 8, 7281–7296, doi:10.5194/acp-8-7281-2008, <http://www.atmos-chem-phys.net/8/7281/2008/>, 2008.
- Baltink, H. K.: CESAR-database, <http://www.cesar-database.nl>, 2016.
- Beyrich, F.: Mixing-height estimation in the convective boundary layer using sodar data, *Boundary-Layer Meteorology*, 74, 1–18, doi:10.1007/BF00715708, <http://dx.doi.org/10.1007/BF00715708>, 1995.
- Dijkstra, E.: A note on two problems in connexion with graphs, *Numerische Mathematik*, 1, 269–271, doi:10.1007/BF01386390, 1959.
- Donovan, D. P. and Apituley, A.: Practical depolarization-ratio-based inversion procedure: lidar measurements of the Eyjafjallajökull ash cloud over the Netherlands, *Appl. Opt.*, 52, 2394–2415, doi:10.1364/AO.52.002394, <http://ao.osa.org/abstract.cfm?URI=ao-52-11-2394>, 2013.
- Eresmaa, N., Karppinen, A., Joffe, S. M., Räsänen, J., and Talvitie, H.: Mixing height determination by ceilometer, *Atmospheric Chemistry and Physics*, 6, 1485–1493, doi:10.5194/acp-6-1485-2006, <http://www.atmos-chem-phys.net/6/1485/2006/>, 2006.
- EUMETNET: E-PROFILE Website, <http://www.eumetnet.eu/e-profile>, 2016.
- Flamant, C., Pelon, J., Flamant, P., and Durand, P.: Lidar determination of the entrainment zone thickness at the top of the unstable marine atmospheric boundary layer, *Boundary-Layer Meteorology*, 83, 247–284, doi:10.1023/A:1000258318944, <http://dx.doi.org/10.1023/A%3A1000258318944>, 1997.
- Haefelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G. P., Lolli, S., O’Dowd, C. D., Sauvage, L., Xueref-Rémy, I., Wastine, B., and Feist, D. G.: Evaluation of Mixing-Height Retrievals from Automatic Profiling Lidars and Ceilometers in View of Future Integrated Networks in Europe, *Boundary-Layer Meteorology*, 143, 49–75, doi:10.1007/s10546-011-9643-z, <http://dx.doi.org/10.1007/s10546-011-9643-z>, 2012.
- Haij, M. d., Wauben, W., and Baltink, H. K.: Continuous mixing layer height determination using the LD-40 ceilometer: a feasibility study, *Scientific Report WR 2007-01*, KNMI, De Bilt, The Netherlands, 2007.
- Harvey, N. J., Hogan, R. J., and Dacre, H. F.: A method to diagnose boundary-layer type using Doppler lidar, *Quarterly Journal of the Royal Meteorological Society*, 139, 1681–1693, doi:10.1002/qj.2068, 2013a.
- Harvey, N. J., Hogan, R. J., and Dacre, H. F.: A method to diagnose boundary-layer type using Doppler lidar, *Quarterly Journal of the Royal Meteorological Society*, 139, 1681–1693, doi:10.1002/qj.2068, <http://dx.doi.org/10.1002/qj.2068>, 2013b.
- Hayden, K. L., Anlauf, K. G., Hoff, R. M., Strapp, J. W., Bottenheim, J. W., Wiebe, H. A., Froude, F. A., Martin, J. B., Steyn, D. G., and McKendry, I. G.: The vertical chemical and meteorological structure of the boundary layer in the Lower Fraser Valley during Pacific ’93, *Atmospheric Environment*, 31, 2089–2105, doi:[http://dx.doi.org/10.1016/S1352-2310\(96\)00300-7](http://dx.doi.org/10.1016/S1352-2310(96)00300-7), <http://www.sciencedirect.com/science/article/pii/S1352231096003007>, 1997.



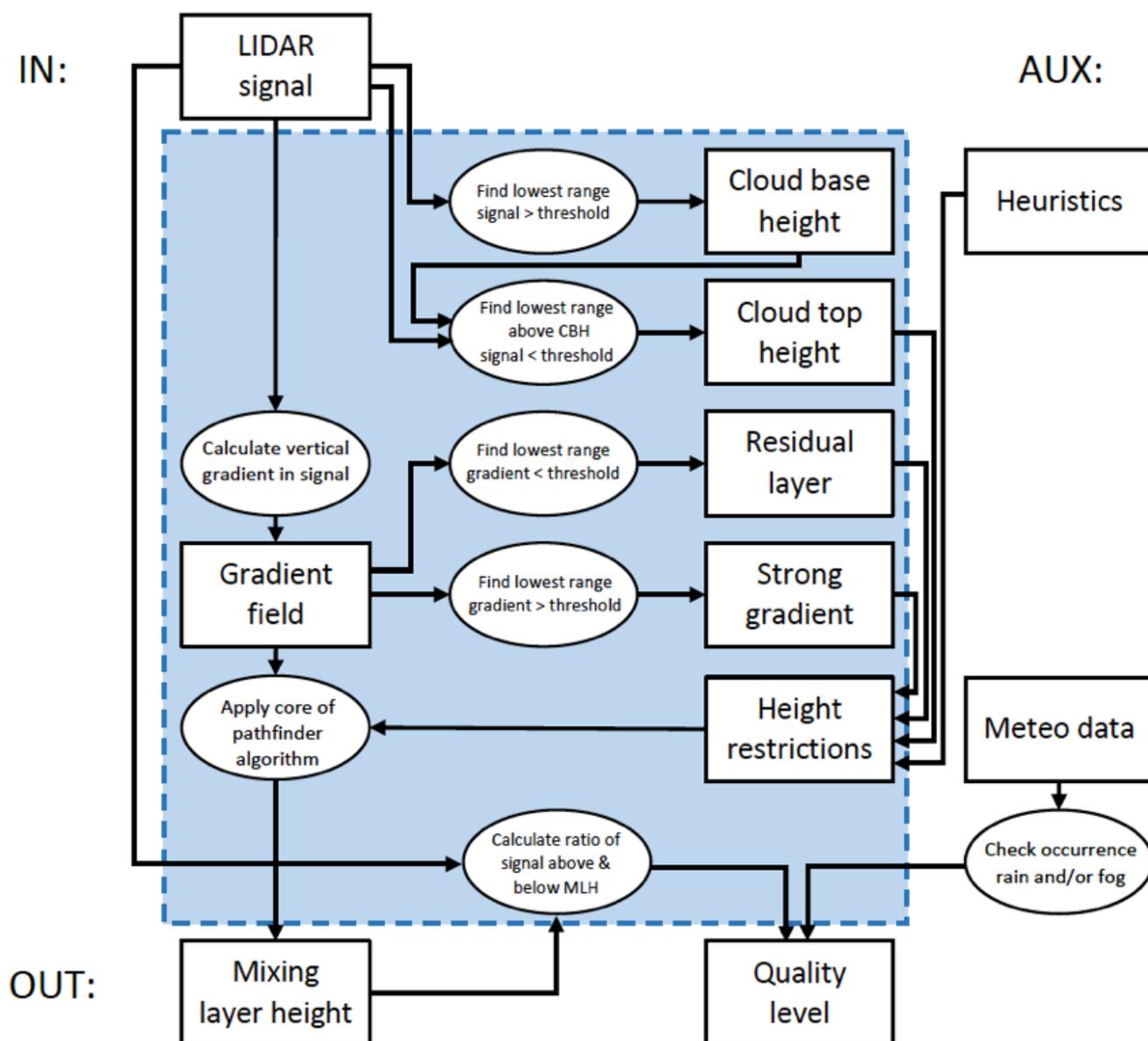
- Holtzworth, G. C.: Estimates of mean maximum mixing depths in the contiguous United States, *Monthly Weather Review*, 92, 235–242, doi:10.1175/1520-0493(1964)092<0235:EOMMMD>2.3.CO;2, [http://dx.doi.org/10.1175/1520-0493\(1964\)092<0235:EOMMMD>2.3.CO;2](http://dx.doi.org/10.1175/1520-0493(1964)092<0235:EOMMMD>2.3.CO;2), 1964.
- Hooper, W. P. and Eloranta, E. W.: Lidar Measurements of Wind in the Planetary Boundary Layer: The Method, Accuracy and  
5 Results from Joint Measurements with Radiosonde and Kyttoon, *Journal of Climate and Applied Meteorology*, 25, 990–1001, doi:10.1175/1520-0450(1986)025<0990:LMOWIT>2.0.CO;2, 1986.
- Illingworth, A.: TOPROF website, <http://www.toprof.imaacnr.it>, 2016.
- Menut, L., Flamant, C., Pelon, J., and Flamant, P. H.: Urban boundary-layer height determination from lidar measurements over the Paris area, *Appl. Opt.*, 38, 945–954, doi:10.1364/AO.38.000945, <http://ao.osa.org/abstract.cfm?URI=ao-38-6-945>, 1999.
- 10 Morille, Y., Haeffelin, M., Drobinski, P., and Pelon, J.: STRAT: An Automated Algorithm to Retrieve the Vertical Structure of the Atmosphere from Single-Channel Lidar Data, *Journal of Atmospheric and Oceanic Technology*, 24, 761–775, doi:10.1175/JTECH2008.1, <http://dx.doi.org/10.1175/JTECH2008.1>, 2007.
- Ottersten, H.: Atmospheric Structure and Radar Backscattering in Clear Air, *Radio Science*, 4, 1179–1193, doi:10.1029/RS004i012p01179, <http://dx.doi.org/10.1029/RS004i012p01179>, 1969.
- 15 Pal, S., Haeffelin, M., and Batchvarova, E.: Exploring a geophysical process-based attribution technique for the determination of the atmospheric boundary layer depth using aerosol lidar and near-surface meteorological measurements, *Journal of Geophysical Research: Atmospheres*, 118, 9277–9295, doi:10.1002/jgrd.50710, <http://dx.doi.org/10.1002/jgrd.50710>, 2013.
- Roelofs, G.-J., ten Brink, H., Kiendler-Scharr, A., de Leeuw, G., Mensah, A., Minikin, A., and Otjes, R.: Evaluation of simulated aerosol properties with the aerosol-climate model ECHAM5-HAM using observations from the IMPACT field campaign, *Atmospheric Chemistry and Physics*, 10, 7709–7722, doi:10.5194/acp-10-7709-2010, <http://www.atmos-chem-phys.net/10/7709/2010/>, 2010.
- 20 Stull, R.: *An Introduction to Boundary Layer Meteorology*, Springer, 1988.
- Thomas, W.: DWD Ceilometer Map, <http://dwd.de/ceilomap>, 2016.
- Van Pul, W. A. J., Holtzlag, A. A. M., and Swart, D. P. J.: A comparison of ABL heights inferred routinely from lidar and radiosondes at noontime, *Boundary-Layer Meteorology*, 68, 173–191, doi:10.1007/BF00712670, <http://dx.doi.org/10.1007/BF00712670>, 1994.
- 25 Voegelezang, D. H. P. and Holtzlag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, *Boundary-Layer Meteorology*, 81, 245–269, doi:10.1007/BF02430331, <http://dx.doi.org/10.1007/BF02430331>, 1996.
- Wiegner, M., Emeis, S., Freudenthaler, V., Heese, B., Junkermann, W., Münkler, C., Schäfer, K., Seefeldner, M., and Vogt, S.: Mixing layer height over Munich, Germany: Variability and comparisons of different methodologies, *Journal of Geophysical Research: Atmospheres*, 111, n/a–n/a, doi:10.1029/2005JD006593, <http://dx.doi.org/10.1029/2005JD006593>, 2006.
- 30 Zilitinkevich, S. and Baklanov, A.: Calculation Of The Height Of The Stable Boundary Layer In Practical Applications, *Boundary-Layer Meteorology*, 105, 389–409, doi:10.1023/A:1020376832738, <http://dx.doi.org/10.1023/A:1020376832738>, 2002.

**Table 1.** Leosphere ALS450 main characteristics.

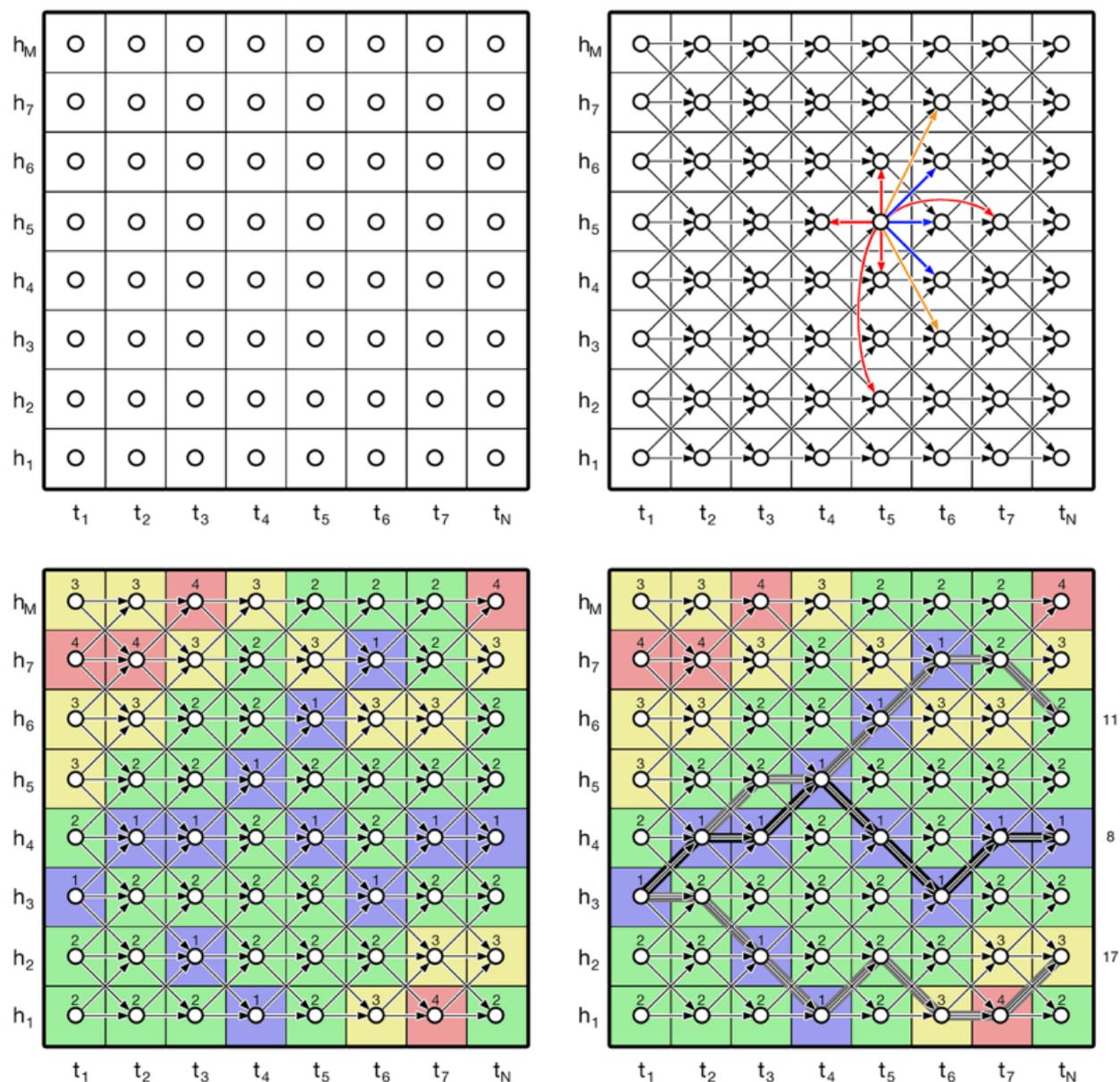
Emission	Wavelength	355 nm
	Power	16 mJ/pulse
	PRF	20 Hz
	Divergence	<0.5 mrad
Detection	Wavelength	355 nm

**Table 2.** Overview of Pathfinder parameters applied to Leosphere ALS450 lidar data.

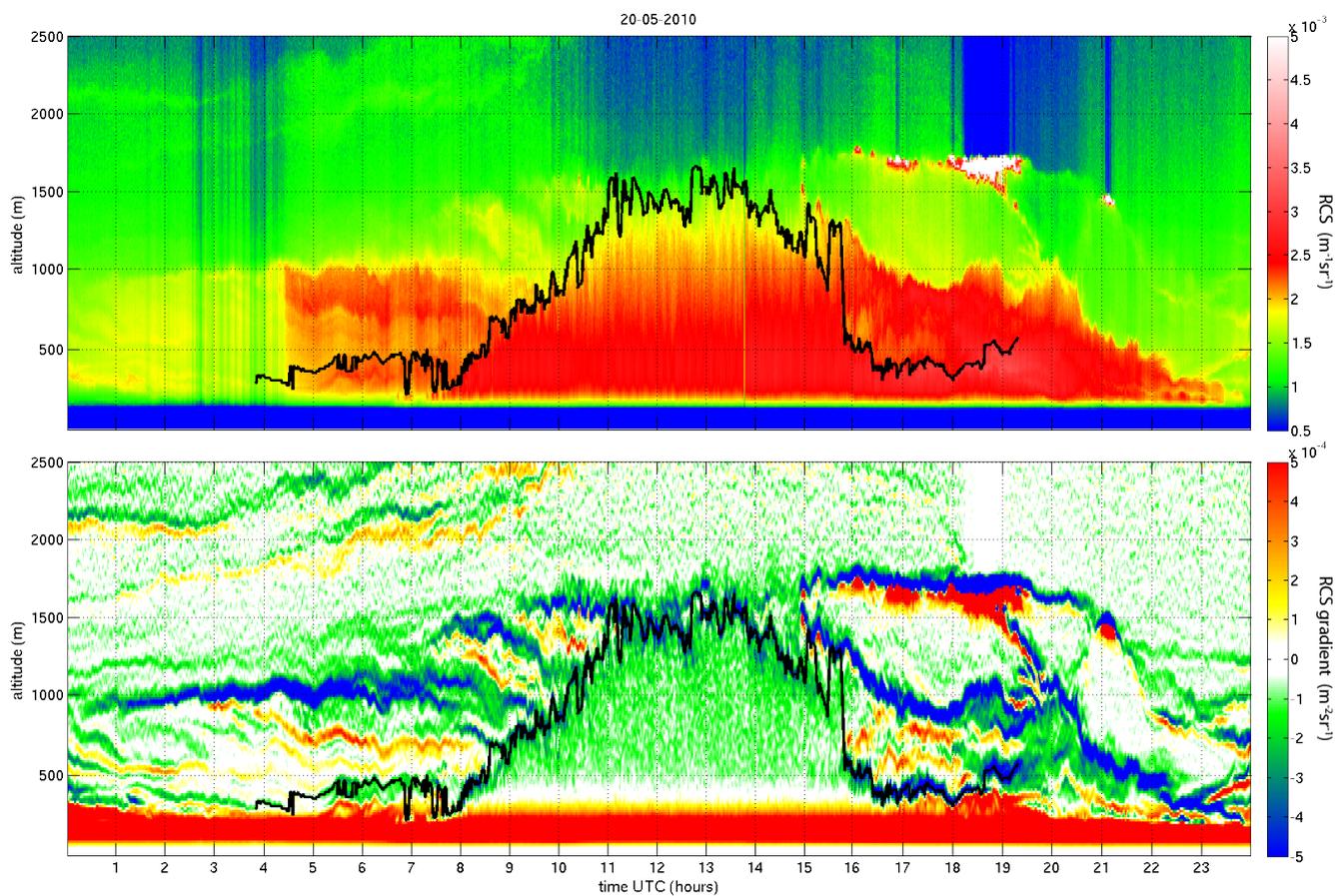
Parameter	value
Gaussian smoothing	$\sigma = 1.1$ range gates
Convective delay	3 hours
<b>Guiding restrictions:</b>	
clouds	$> 5 \times 10^{-3} \text{m}^{-1} \text{sr}^{-1}$
negative gradient	$> 5 \times 10^{-4} \text{m}^{-2} \text{sr}^{-1}$
positive gradient morning	$< -1 \times 10^{-5} \text{m}^{-2} \text{sr}^{-1}$
positive gradient day	$< -2 \times 10^{-4} \text{m}^{-2} \text{sr}^{-1}$
maximum altitude night	750m
maximum altitude day	3000m
minimum altitude	175m
<b>Graph:</b>	
window size	15 minutes
vertical range 1-to-1	75m
vertical range total	900m



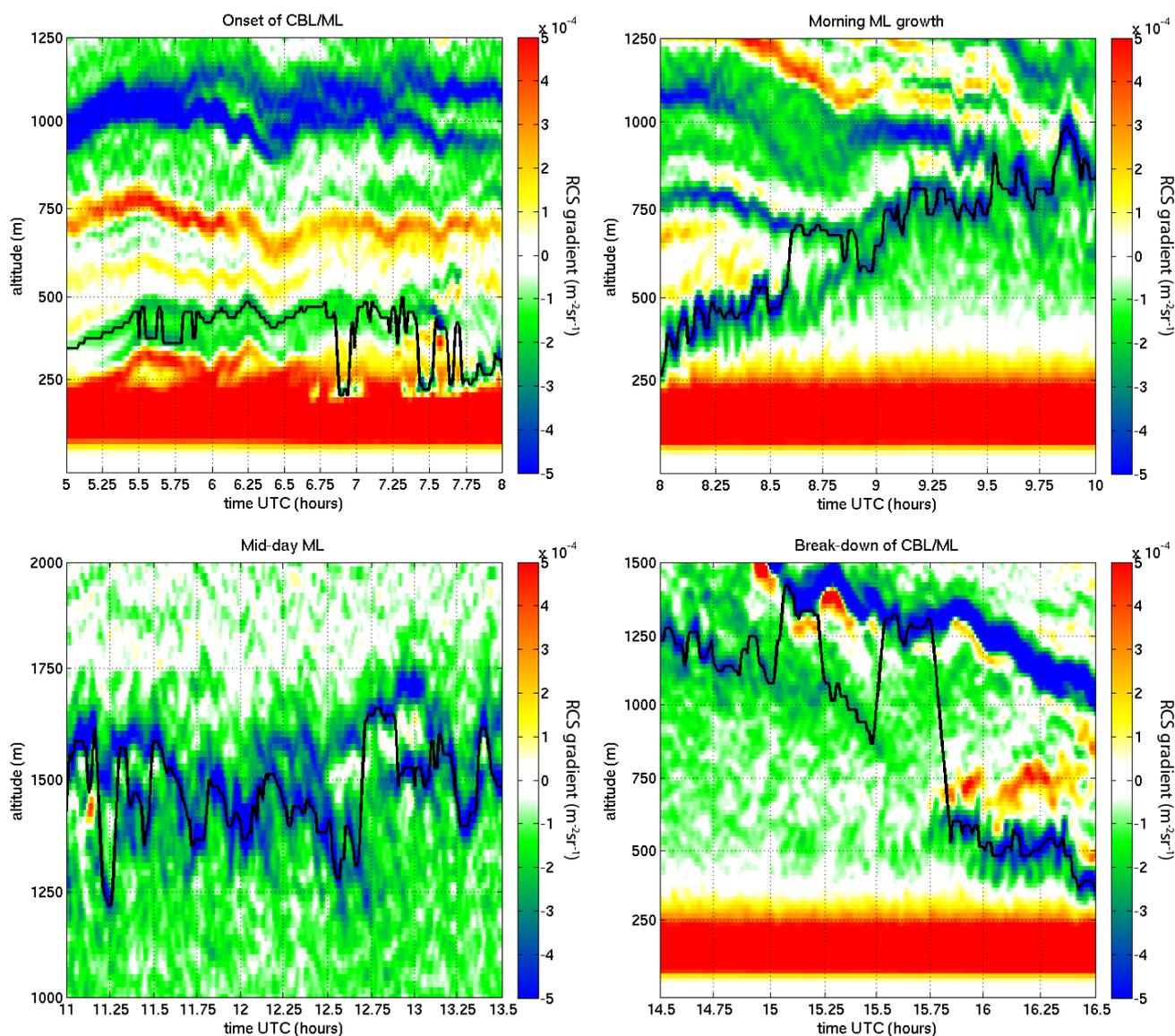
**Figure 1.** Flow diagram of the Pathfinder algorithm. Internal operations and variables are grouped within the blue area. Input data is shown at the top and output variables at the bottom. Auxiliary data is displayed at right.



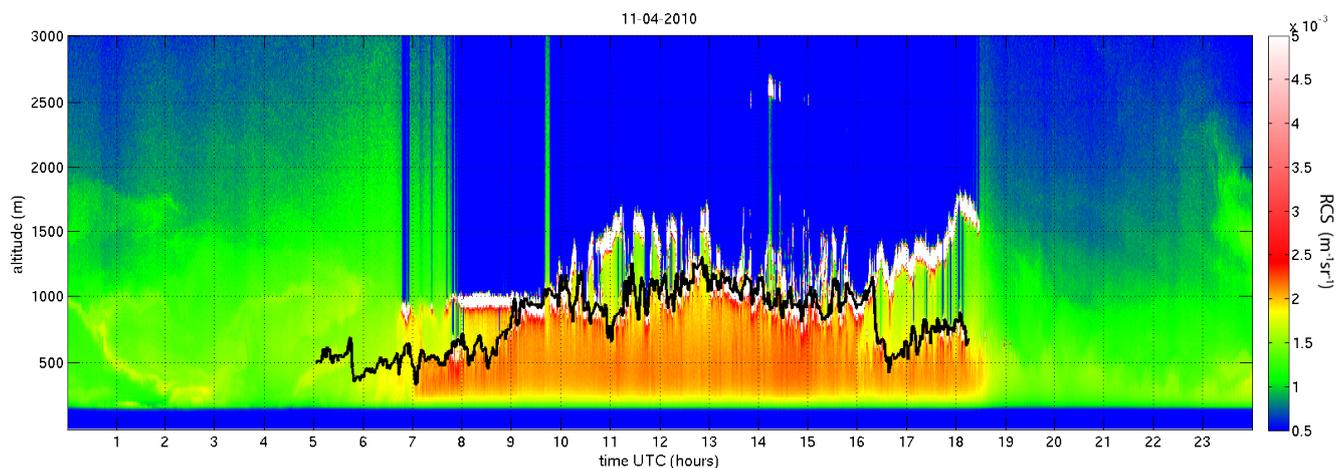
**Figure 2.** Overview of the translation of data into a mathematical graph and subsequent steps to determine MLH. a) one-to-one translation of datapoints to graph vertices. b) Grid of connections between vertices, for a single vertex allowed connections are shown in blue, examples forbidden connections shown in orange and red. c) Assignment of weights to the graph connections based on RCS vertical gradients. d) Selection of correct path with Dijkstra's algorithm, with 3 possible paths originating at the same point. Their total cost is displayed at the right. The middle path will be selected as MLH development estimate as this is the path with the lowest possible cost.



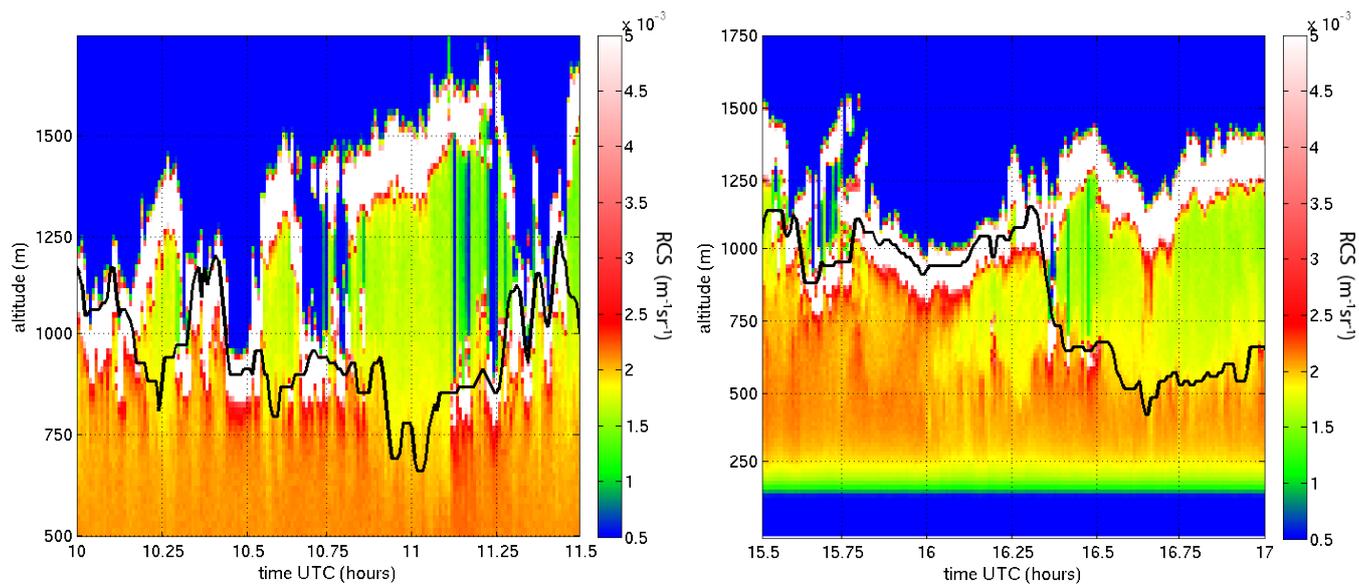
**Figure 3.** (a) A full diurnal cycle of backscatter lidar profiles from the ALS450 system at Cabauw on 20 May 2010. The data presented is the range corrected signal *RCS*. (b) The gradient field calculated from the backscatter lidar data.



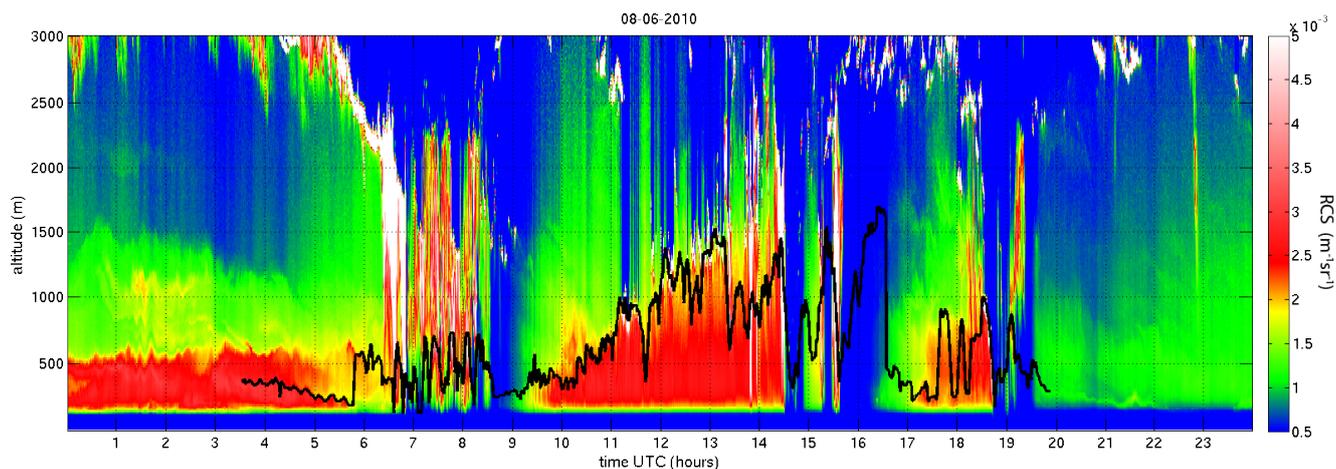
**Figure 4.** Highlights of the MLH evolution at Cabauw on 20 May 2010. See text in Sec.4.1. Gradients in the *RCS* and Pathfinder MLH estimate during a clear sky day. These include (a) the period between sunrise and onset of the ML (b) the morning growth of the ML (c) the mid-day structure of the MLH and (d) the break-down of the ML. Time is indicated in decimal hours.



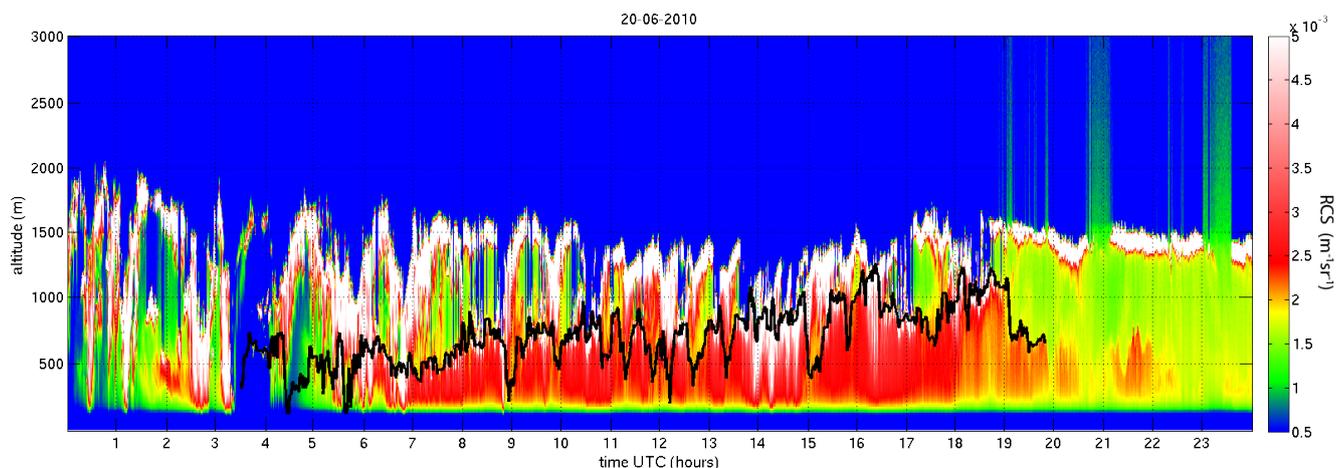
**Figure 5.** Lidar *RCS* together with Pathfinder MLH estimate on 11 April 2010. A day with cumulus cloud forming on top of the ML with an additional stratocumulus cloud layer above.



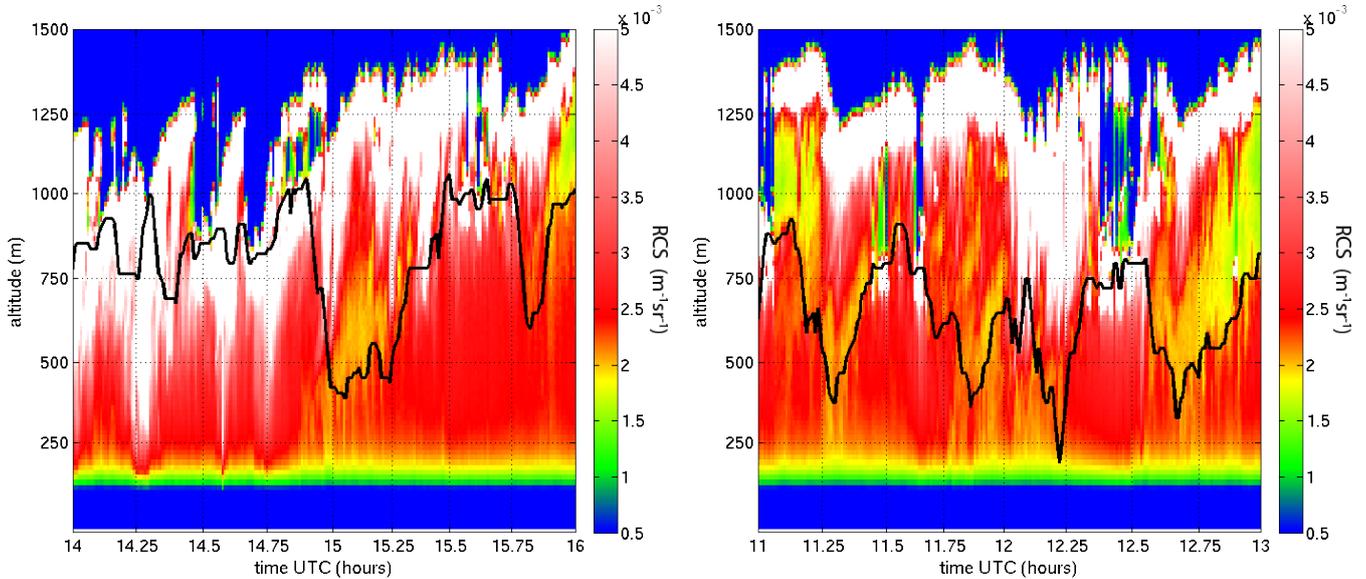
**Figure 6.** Highlights of the MLH estimate found by the Pathfinder algorithm. See text in Sec.4.2. Shown are (a) difficulties in distinguishing ML and non-ML clouds and (b) commitment to a cloud layer that detaches from the ML.



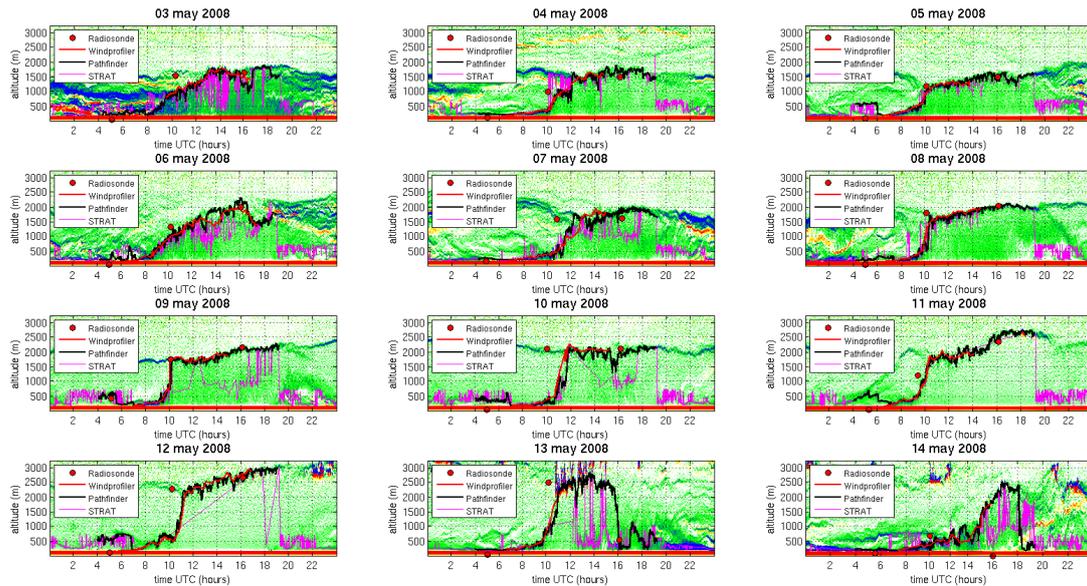
**Figure 7.** Lidar *RCS* together with Pathfinder MLH estimate on 20 June 2010. A day with a completely overcast sky and precipitation falling from the cloud but (almost) completely evaporating before reaching the surface.



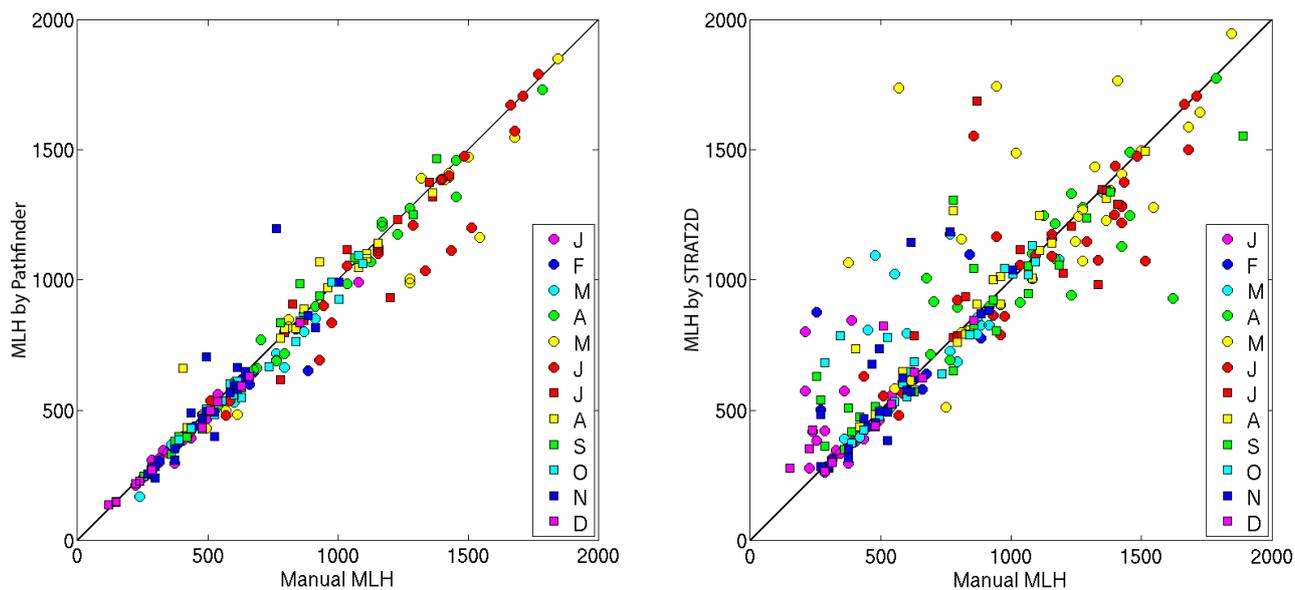
**Figure 8.** Lidar *RCS* together with Pathfinder MLH estimate on 20 June 2010. A day with a completely overcast sky and precipitation falling from the cloud but (almost) completely evaporating before reaching the surface.



**Figure 9.** Highlights of the MLH estimate found by the Pathfinder algorithm on 20 June 2010. See text in Sec. 4.2. Shown are (a) precipitation events reaching the lidar overlap region and (b) precipitation evaporating well above this region.



**Figure 10.** Overview of MLH retrievals from ‘Pathfinder’  $MLH_{L,P}$ , STRAT2D  $MLH_{L,S}$ , Windprofiler  $MLH_W$  and Radiosonde  $MLH_R$  for the 12 days of IMPACT data from 2008, shown against the background of the ALS450  $RCS$  gradient field.



**Figure 11.** Scatter plots of (a) Pathfinder  $MLH_{L,P}$  against  $MLH_M$  ( $R^2 = 0.96$ ) and (b) STRAT2D  $MLH_{L,S}$  against  $MLH_M$  ( $R^2 = 0.74$ ) for the full dataset of 2010. Color coding indicates the month of the year.