

Supplement of Atmos. Chem. Phys., 17, 10051–10070, 2017
<https://doi.org/10.5194/acp-17-10051-2017-supplement>
© Author(s) 2017. This work is distributed under
the Creative Commons Attribution 3.0 License.



Supplement of

PathfinderTURB: an automatic boundary layer algorithm. Development, validation and application to study the impact on in situ measurements at the Jungfraujoch

Yann Poltera et al.

Correspondence to: Giovanni Martucci (giovanni.martucci@meteoswiss.ch)

The copyright of individual parts of the supplement might differ from the CC BY 3.0 License.

1 Supplement

2 S1. PathfinderTURB algorithm and data

3 The S signal in eq.(1) in the accompanying article, is used by PathfinderTURB to retrieve the Convective
4 Boundary Layer (CBL) and the Top of the Continuous Aerosol Layer (TCAL) at Payerne (PAY) and at the
5 Kleine Scheidegg (KSE) with a time resolution of 1 minute (2 minutes at KSE) and a range resolution of 30 m
6 (45 m at KSE). The S -profile is smoothed in space and time (2-D smoothing) using a Gaussian filter with
7 FWHM equal to 5 bins in range corresponding to 150 m (225 m at KSE) and 5 steps in time corresponding to 5
8 minutes (10 minutes at KSE) and a standard deviation equal to 1.1 bins in range and 1.1 steps in time. Then
9 smoothed with the 2-D Perona-Malik anisotropic diffusion method (Perona and Malik, 1990, as suggested in
10 Haeffelin et al., 2012), based on 15 iterations and a conduction controlling coefficient of 2500. Values less than
11 1000 are set to 1000 and then the \log_{10} is taken. The corresponding gradient field is calculated with central
12 differences of this \log_{10} , smoothed S .

13 The central differences of the logarithm of S in time and space are proportional to the vertical ratio of the
14 logarithm of S at two range bins for the time t_j :

$$\nabla \left(\log_{10} \left(S(R_i, t_j) \right) \right) = \frac{\left(\log_{10} \left(S(R_{i+1}, t_j) \right) - \log_{10} \left(S(R_{i-1}, t_j) \right) \right)}{R_{i+1} - R_{i-1}} = \frac{1}{R_{i+1} - R_{i-1}} \log_{10} \left(\frac{S(R_{i+1})}{S(R_{i-1})} \right)$$

15 The criteria described in section 4.2 to restrict S and ∇S in range and time are applied to PathfinderTURB at
16 each time step and represent an important constraint for the algorithm. Next, we provide further details about the
17 restrictions based on strong negative and positive gradients, as well as about the calculation of the atmospheric
18 variance ($VAR(S)$) and the TCAL. The last section describes the guidelines and data used for the human expert
19 retrieval of the CBLH during the validation phase of PathfinderTURB in Payerne.

20 S1.1 Strong negative gradients

21 During the early morning period, the threshold is set to $\log_{10}(0.85)/(2dR)$, which is the value of the
22 derivative when S undergoes a drop larger than 15%. Else, the threshold is set to $\log_{10}(0.75)/(2dR)$, (signal
23 drop more than 25%). The signal is analysed from 250 m upwards. In case there is a cloud obstruction the
24 CBLH, the limit is set to the cloud base height (CBH).

25 S1.2 Strong positive gradients

26 During the early morning period, the threshold is set to $\log_{10}(0.95)/(2dR)$, which is the value of the
27 derivative when S undergoes an increase larger than 5%. Else, the threshold is set to $\log_{10}(0.85)/(2dR)$,
28 (signal increase of more than 15%). The signal is analysed from 250 m upwards. In case there is a cloud
29 obstructing the ceilometer's view to the CBLH, the limit is set at the CBH. If a strong negative gradient occurs at
30 less than 300 m above a strong positive gradient, the limit is set at the strong negative gradient's altitude.

1 The strong negative gradients, strong positive gradients and lower altitude restrictions are applied to a time
2 resolution of 5 minutes, i.e. at time t_j , the restriction corresponding to the highest altitude over 5-minutes centred
3 on t_j is retained.

4 **S2. Calculation of the first transition to turbulence, atmospheric variance and total variance**

5 The backscatter signal shows usually high temporal fluctuations at the top of the CBLH. However, when
6 calculating $VAR(S)$ from the timeseries of S , mesoscale to microscale fluctuations contribute to the $VAR(S)$
7 value. We can use spectral analysis (FFT analysis) to try to recover the fluctuations solely due to aerosol number
8 density fluctuations, but also to estimate whether we are in a fully developed turbulent regime, by looking at the
9 spectral power decay. This gives, at each range and time step, an estimation of the atmospheric variability as an
10 expression of turbulence.

11 Let us be given a 1h-long S timeseries, centered at time step t_j , and at range R_i . The method, based on (Pal et al.,
12 2013), goes as follows:

- 13 • De-spike the timeseries recursively with a sliding histogram as in (Lenschow et al., 2000). This step is used
14 to remove extreme values due to, for example, clouds.
- 15 • De-trend with a polynomial regression of order 2 (i.e. subtract the fit from the timeseries).
- 16 • Calculate the FFT.
- 17 • Reduce the influence of mesoscale variability (periods of 30 min and more) with a high-pass filter applied
18 on the FFT coefficients. We use a raised cosine filter (high-pass cutoff $f_{cutoff} = 1/1800$ Hz and roll-off factor
19 0.5).
- 20 • Integrate the power of the high-pass filtered FFT between $-0.75 f_{Nyquist}$ and $0.75 f_{Nyquist}$. This gives us the
21 atmospheric variance $VAR^{atm}(S)$.
- 22 • Integrate the power of the high-pass filtered FFT everywhere. This gives us the total variance $VAR^{tot}(S)$.
- 23 • Perform a linear fit of the power spectrum, in logarithmic scale, between the 10-minutes frequency and the
24 frequency corresponding to $0.75 f_{Nyquist}$, and calculate the slope of the fit. Calculate then the relative error
25 err_{rel} of the slope compared to the theoretical value of $-5/3$ (Kolmogorov's power law for the inertial
26 subrange) and apply the power 4 to this relative error.

27 $VAR^{atm}(S)$ is an hourly approximation of the atmospheric variance. We divide it by $VAR^{tot}(S)$ and apply the
28 power 4 to have a larger span of the values between 0 and 1, because this ratio of $VAR^{atm}(S)$ vs. $VAR^{tot}(S)$ is
29 about 0.75 for a signal with fluctuations dominated by noise and is close to 1 for a signal with fluctuations
30 dominated by aerosol number density fluctuations. The resulting value,
31 $VAR(S) = (VAR^{atm}(S)/VAR^{tot}(S))^4$, is a measure of the atmospheric variability.

32 In mixing and entrainment zones and periods, the $-5/3$ scaling law of turbulence for the inertial subrange can be
33 observed. The relative error is then a good proxy to see if the region observed is in a turbulent regime or not.
34 Again, we apply the power 4 to have a larger span of the values between 0 and 1 and so to better discriminate
35 between low and high turbulence zones. The resulting value, $turbfit = (1 - err_{rel})^4$, is a measure of turbulence.

1 **S2.1 Implementation**

2 From 00:00 to 24:00, in steps of 10 minutes (in order to reduce the computational time), the atmospheric
3 variability $VAR(S)$ and the turbulence $turbfit$ are calculated over periods of one hour centered on each time
4 steps, for each altitude between the ground and the maximum climatological value set before.

5 The signal used is the standard signal S ($beta_raw$) generated automatically by the CHM15K ceilometer, whose
6 time and range resolutions are reduced to 1 minute and 30 m, respectively. The Nyquist frequency $f_{Nyquist}$ is then
7 2 minutes and the first noise frequency $0.75 f_{Nyquist}$ is 2 minutes and 40 seconds.

8 The atmospheric variance $VAR(S)$ and the turbulence $turbfit$ are first smoothed in range with a running mean of
9 11 points (i.e. over 300 m), then both in range and time (2-D) using a Gaussian filter with FWHM equal to 5 bins
10 in range corresponding to 150 m and 5 steps in time corresponding to 50 minutes (11 bins in range and 11 steps
11 in time for $turbfit$) and a standard deviation equal to 1.1 bins in range and 1.1 steps in time. Finally, they are
12 interpolated on the time-and-range grid of the gradient field calculated before (i.e., 1 min and 30 m in PAY; 2
13 min and 45 m in KSE).

14 **S2.2 Inferior limit for altitude from transition to turbulence**

15 After the early morning period and before sunset, the altitude of the first transition corresponding to the onset of
16 turbulence is calculated by applying a Haar-wavelet covariance transform to the profiles of $turbfit$ and averaging
17 it over multiple scales (60 m to 240 m at PAY, 90 m to 360 m at KSE), then we look at the first local minimum
18 of the averaged transform.

19 If necessary, the retrieval of the CBLH could be made in an even more robust manner by varying the starting
20 times and window widths, as described in the next subsection. However, doing so did not change much the
21 results presented in the related article.

22 **S3 Robust choice of CBLH**

23 The pathfinder has an inherent uncertainty, and its sources are due to the a-priori range restrictions, the gradient
24 and variance calculations, the window's width and the starting time. We consider here only the latter two
25 uncertainty sources, and run the algorithm from sunrise to sunset with offsets of -30 to 30 minutes in steps of 10
26 minutes on the sunrise time, and window's widths of 15 minutes to 45 minutes in steps of 5 minutes. This 49
27 runs of PathfinderTURB, which increases the computational time to about 300 s. Parallelization of this step is
28 straightforward such that, if needed, the computational time could be reduced easily on nowadays standard
29 multicore machines.

30 To select the final CBLH time series, we create a count matrix, which counts the number of CBLH candidates at
31 a given altitude and at a given time step and divides this number by the number of CBLH candidates at this time
32 step. This gives a value between 0 and 1. It is sometimes not enough to select the final CBLH time series based
33 on the count matrix, and so we multiply the count matrix by a ratio matrix, which at each altitude and time step
34 calculates the ratio of the mean of S over $z + 150$ m to the mean of S over $z - 150$ m for the considered altitude z .
35 Negative ratios or ratios over 1 are set to 1. We subtract then the ratio from 1 and multiply the result by 2. This
36 gives a value between 0 and 2, values close to 2 belonging to clouds, else the values are mostly below 1.

1 For each CBLH time series, we add the values of the product of the count and ratio matrices along the time
2 series path. The time series having the maximal value is then selected as final CBLH time series.

3 **S4. Calculation of the TCAL**

4 The same time resolution and the 2-D Gaussian smoothing applied to the S signal (Sect. S1) are applied also to
5 the detected noise level by using error propagation and ignoring non-diagonal covariance terms. The signal-to-
6 noise ratio (SNR) is then calculated simply dividing S by the noise level.

7 The signal is considered too noisy when the SNR drops below 0.6745 or when S becomes artificially negative
8 (which happens due to the manufacturer automatic background light subtraction procedure). This creates a
9 binary mask the same size of S with the limitation that isolated points are not seen as noisy even if they have
10 neighbouring noisy points and vice versa. Following (Biavati, 2014), image processing morphological operators,
11 three erosions followed by 20 dilations, are used to make such isolated points disappear. Then, at each time step,
12 the SNR mask is set to 0 at and above the first range bin above 600m where the mask is 0.

13 The discrimination of pure molecular signal from the aerosol signal is performed here by setting a threshold
14 value on the profile of $\log_{10}(S)$. Once we have determined the instrumental calibration constant using the
15 procedure suggested by (Wiegner and Geiß, 2012) and after we have calculated the molecular profiles over one
16 year with the formulas given in (Bucholtz, 1995) (Pressure and temperature from radiosounding), a threshold Th
17 to discriminate aerosols from molecules can be derived as $Th=4.625$. We then define a binary mask for aerosols
18 based on the threshold Th . Each profile of $\log_{10}(S)$ is smoothed in range by applying a 11-point running mean at
19 all ranges above the first range bin where the overlap function is larger than 0.05, and the aerosol mask is set to 0
20 at and above the first range bin where the running mean is lower than Th . The aerosol mask is also set to 0 at all
21 ranges above the detected cloud base. Then, three erosions followed by 10 dilations are applied to the mask. For
22 profiles where S is already smaller than Th at the ground, the mask is set to 0 at all ranges (purely molecular
23 atmosphere). At each time step, the TCAL is given by the first range bin where the product of the SNR mask and
24 of the aerosol mask is 0. The actual temporal resolution of the calculated TCAL is then decreased to 10 minutes
25 (5 time steps), i.e. for each time step, the highest TCAL calculated during 5 time steps centred on the current
26 time step is selected. In case of low cloud, the TCAL is not considered for the analysis.

27 **S5. Human expert CBLH retrieval in Payerne**

28 The following guidelines were adopted to perform the retrieval of the CBLH by human experts:

- 29 • A CBLH detection must lie on or be very close (within ± 2 range bins) to an aerosol gradient.
- 30 • A CBLH detection must always exist during clear-sky periods. It is however subject to restrictions in
31 presence of cloud, fog and precipitation:
 - 32 ○ Cloud: the detection must match the cloud base.
 - 33 ○ Fog or low stratus cloud (<200 m a.g.l.): detection allowed when it starts to dissipate.
 - 34 ○ Precipitation: detection possible only during time intervals without precipitation of at least 2
35 hours.
- 36 • Meteorological and astronomical information should be taken into account in order to respect the
37 following criteria:

- 1 ○ The detection starts at sunrise and ends at sunset.
- 2 ○ The CBL starts developing from the ground after sunrise, i.e. elevated layers shall not be
- 3 selected during early morning stage.
- 4 ○ The CBLH shows a decay/breakdown when the vertical heat flux at the ground drops to about
- 5 0 and ΔT has a negative local minimum value normally at mid-afternoon till the end of the day,
- 6 with ΔT calculated hourly by taking the difference between two consecutive hourly means of
- 7 the surface temperature measurement. If no decaying layer can be followed, no detection is to
- 8 be provided.
- 9 • The expert selects as many points as it is necessary in order for a linear interpolation between two
- 10 adjacent points to represent trustworthily the short-term evolution of the observed layer.

11 All information are taken into account (station measurements, CBLH estimations from remote-sensing
12 instruments and radio-sounding launches, synoptic conditions, etc.) and compared to the expert detection. In case
13 of clear mistakes by the expert, the detection points should be edited accordingly.

14 **S5.1 Determination of the expert consensus**

15 The CBLH detections by each expert are linearly interpolated over 30 s and truncated before sunrise and after
16 sunset. Before comparing the CBLH values retrieved by PathfinderTURB with those detected by the experts, a
17 number of additional restrictions were defined. Manually-detected CBLH values for a given day must fulfil the
18 following conditions in order to be “valid”:

- 19 • No precipitation (station measurement) and no fog (ceilometer measurement) for more than 2
- 20 consecutive hours.
- 21 • Only time periods of at least 30-minute duration, containing interpolated data from at least two experts
- 22 and with spread (i.e. difference between the maximum and minimum values) of less than 10% of the mean
- 23 CBLH plus 100 m are taken into account. The allowed spread increases with altitude because the probability to
- 24 lie on different layers decreases with increasing altitude. The offset of 100 m added to the maximum allowed
- 25 spread is an empirical value that translates into a permitted 300-m spread for a CBLH at 2000 m, which is a
- 26 conservative estimate of the depth of the entrainment zone at that altitude (EZ depth can be 40% of the CBLH).

27 The CBLH detections respecting the above criteria are retained for comparison with PathfinderTURB. The mean
28 value of all the valid detections (reference and test group) for each time step was calculated along with the lower
29 and upper error bounds determined as the minimum and the maximum detected CBLH. In total, an expert
30 consensus was reached for 135 days, out of the initial 174, covering a total of 43914 minutes. On 6 days, no
31 ceilometer data were available. On 13 days, poor weather conditions prevented all detections. On 20 days, either
32 the spread was too large or the duration of the matched temporal interval was too short.

33 Because the number of CBLH detections decreases as the days become shorter in winter, midday is the period of
34 the day with the highest availability of detections (and the best agreement). Both morning and afternoon periods
35 present difficulties, when detecting the CBLH from the timeseries of S, nevertheless the morning provides better
36 availability of detection than the afternoon. The limitations of morning detections are related to the fact that
37 during the first 2-3 hours after sunrise the CBL top is still low above the ground (< 200 m a.g.l.) and the overlap-

1 corrected, normalized signal, S , is affected by the incomplete overlap in that zone, which makes the detection in
2 the first few hundred meters difficult and more uncertain. Another source of complexity in the morning
3 detections is that the residual nocturnal layer may remain aloft and very close to the developing CBLH, which
4 increases the uncertainty related to the layer attribution. During the late afternoon, the CBL transforms into the
5 RL as soon as buoyancy weakens and it becomes neutrally stratified. Under these conditions, the CBL top drops
6 rapidly, but usually without displaying a clear aerosol gradient, which leads to a large (and somewhat
7 unphysical) spread among the experts' detections.