



An empirical method to correct for temperature dependent variations in the overlap function of CHM15k ceilometers

Maxime Hervo¹, Yann Poltera^{1,*} and Alexander Haeefe¹,

[1]{MeteoSwiss, Payerne, Switzerland}

[*]{now at: Institute for Atmospheric and Climate Science, ETH, Zurich, Switzerland}

Correspondence to: M. Hervo (maxime.hervo@meteoswiss.ch)

Abstract

Imperfections in a lidar's overlap function lead to artefacts in the background, range and overlap corrected lidar signals. These artefacts can erroneously be interpreted as an aerosol gradient or, in extreme cases, as a cloud base leading to false cloud detection. A correct specification of the overlap function is hence crucial to use automatic elastic lidars (ceilometers) for the detection of the planetary boundary layer or low clouds.

In this study an algorithm is presented to correct such artefacts. It is based on the assumption of a homogeneous boundary layer and a correct specification of the overlap function down to a minimum range, which must be situated within the boundary layer. The strength of the algorithm lies in a sophisticated quality check scheme which allows to reliably identify favorable atmospheric conditions. The algorithm has been applied to 2 years of data from a CHM15k ceilometer from the company Lufft. Backscatter signals corrected for background, range and overlap have been compared using the overlap function provided by the manufacturer and the one corrected with the presented algorithm. Differences between corrected and uncorrected signals reach up to 45% in the first 300m above ground.

The amplitude of the correction turned out to be temperature dependent being larger for higher temperatures. A linear model of the correction as a function of the instrument's internal temperature has been derived from the experimental data. Case studies and a statistical analysis of the strongest gradient derived from corrected signals reveal that the temperature model is capable to correct overlap artefacts with high quality, in particular such



1 due to diurnal variations. The presented correction method has the potential to significantly
2 improve the detection of the boundary layer with gradient based methods because it removes
3 false candidates and hence simplifies the attribution of the detected gradients to the planetary
4 boundary layer. A particularly high benefit can be expected for the detection of shallow stable
5 layers typical for nighttime situations.

6 The algorithm is completely automatic and does not require any intervention on site but
7 requires the definition of an adequate instrument specific configuration. It is therefore suited
8 for the use in large ceilometer networks.

9

10 **1 Introduction**

11 Due to technological advances in the recent decade state-of-the-art ceilometers can nowadays
12 be considered automatic elastic lidars. They are more and more used for profiling of aerosols
13 including the detection of volcanic particles (Flentje et al., 2010) and the determination of the
14 planetary boundary layer (Haeffelin et al., 2012). As for all lidars there is a zone close to the
15 ground where the telescope field of view (FOV) does not fully overlap with the laser beam
16 and hence geometric and instrumental effects distort the measured backscatter profile. This
17 effect is accounted for with the so called overlap function. This function describes the signal
18 loss due to the overlap effect as a function of altitude. A correct determination of the overlap
19 function is crucial for aerosol profiling in the zone of partial overlap, i.e. in the boundary
20 layer.

21 The overlap function can theoretically be modelled if the specifications and configuration of
22 the optical elements of the lidar are known (Kuze et al., 1998; Stelmazczyk et al., 2005). But
23 the precision of such models is not good enough in practice due to several instrumental effects
24 which are generally not known. For example the energy distribution of the laser beam can be
25 ambiguous (Sasano et al., 1979), the transmittance of interference filters may depend on the
26 incident angle (Sasano et al., 1979) or the laser beam might not be well focused on the
27 receiver and will thus alter the measured power (Roberts and Gimmestad, 2002). One of the
28 main issues is the impact of temperature on the optical components (Campbell et al., 2002;
29 Welton and Campbell, 2002).

30 To determine the overlap function experimentally there are several approaches like observing
31 a homogeneous atmosphere (Sasano et al., 1979; Welton et al., 2000), using a Raman signal



1 (Wandinger and Ansmann, 2002) or a hard target (Vande Hey et al., 2011), or using a
2 reference instrument with known overlap function (Guerrero-Rascado et al., 2010; Reichardt
3 et al., 2012). Most of these methods require rather costly installations or human intervention
4 and are thus not suited for larger networks of automatic lidars.

5 The only method that can potentially be applied to a large network without additional costs is,
6 in our opinion, the use of a vertically homogeneous atmosphere. Since this assumption is not
7 justified across the interface between the boundary layer and the free troposphere, this method
8 is only suited for instruments that reach full overlap within a few hundreds of meters, i.e.
9 within the boundary layer (Sasano et al., 1979) or for instruments with correctly specified
10 overlap down to a minimum range within the boundary layer (in this work).

11 Welton et al. (2000) proposed to perform horizontal measurements such that the assumption
12 of a homogeneous atmosphere also holds for instruments which reach full overlap only after a
13 few thousands of meters. Methods using horizontal or inclined measurements are the most
14 used in the scientific community and by manufacturers (Campbell et al., 2002; Biavati et al.,
15 2011). However, these methods assume that the overlap function does not change between
16 vertical and inclined alignment of the system. An assumption, which might not be justified for
17 certain instruments. Further, inclination of instruments requires important mechanical
18 developments or human intervention.

19 Since instrumental parameters are not perfectly constant in time the overlap function needs to
20 be re-evaluated in regular intervals. Hence for dense networks of lidars an automatic approach
21 which requires minimal system modifications is needed. We propose in this study an
22 extension of the method by Sasano et al. (1979) combined with the assumption that a first
23 guess of the overlap function is available. We will show that this method can be implemented
24 for existing instruments without intervention on site and that it is suited for use in large
25 networks of automatic lidars. The algorithm as presented here is optimized for the CHM15k
26 built by the company Lufft Mess- und Regeltechnik GmbH (previously Jenoptik) but can in
27 principle be adapted to other instruments. The paper is organized as follows: the instrument
28 for which the method has been implemented and tested is described in Section 2 and in
29 Section 3 the detailed description of the method is given. Results are presented in Section 4
30 and in Section 5 we discuss temperature effects on the overlap function and propose a model
31 to correct for such effects. Examples of the performance of the correction for the



1 determination of the boundary layer height are presented in Section 6 followed by a summary
 2 and conclusions.

3 **2 The CHM15k-Nimbus ceilometer**

4 The CHM15k-Nimbus ceilometer is a bi-axial photon-counting lidar (1064nm, 6.5KHz, 8μJ)
 5 manufactured by the company Lufft Mess- und Regeltechnik GmbH (www.lufft.com),
 6 hereafter referred to as Lufft. The emitter and the receiver are placed next to each other in the
 7 optical module, with a center-to-center distance of 12 cm. More information about a similar
 8 instrument can be found in Wiegner and Geiß (2012). For the instrument considered in this
 9 study, the lowest level of non-zero (full) overlap is at approximately 180 (800) m. Its relevant
 10 parameters are given in Table 1.

11 The manufacturer Lufft provides an individual overlap function for each optical module which
 12 has been determined in the factory using a reference instrument. However, this overlap
 13 function cannot account for changes over time due to mechanical and thermal stress and can
 14 thus show significant deficiencies as shown in Section 4.2. It has been noted that artefacts due
 15 to differences between the assumed and the true overlap function are visible in the first few
 16 hundreds of meters. Such artefacts are detrimental for various applications like the
 17 determination of the planetary boundary layer height or the retrieval of aerosol optical
 18 properties.

19 **3 Method**

20 **3.1 Physical basis**

21 The lidar equation relates received power, P , as a function of range, r , and time, t , to
 22 instrumental and atmospheric parameters as follows:

$$P(r, t) = \frac{1}{r^2} C_L(t) C_{CHM}(t) O(r, t) \beta(r, t) e^{-2 \int_0^r \alpha(r', t) dr'} + B(t) \quad (1)$$

23 C_L is the time dependent calibration factor, and C_{CHM} is a factor accounting for variations in
 24 the sensitivity of the receiver. C_{CHM} is the product of the variables p_{calc} and $scaling$
 25 provided by the manufacturer. α and β are the extinction and backscatter coefficient,
 26 respectively, and B is the background. $O(r, t)$ is the range and time dependent overlap
 27 function which can be expressed with a temporally constant overlap function provided by the
 28 manufacturer, $O_{CHM}(r)$, and a correction function, $o(r, t)$, as follows:



$$O(r, t) = O_{CHM}(r)/o(r, t). \quad (2)$$

- 1 The standard instrument output, β_{raw} (variable *beta_raw* provided by the manufacturer), is the
 2 normalized and background, range and overlap corrected signal defined as :

$$\beta_{raw}(r, t) = \frac{(P(r, t) - B(t))r^2}{C_{CHM}(t)O_{CHM}(r)}. \quad (3)$$

- 3 We define the corrected instrument output as :

$$\beta_{corrected}(r, t) = \beta_{raw}(r, t) o(r, t), \quad (4)$$

- 4 which is proportional to the attenuated backscatter coefficient defined as

$$\beta_{att}(r, t) = \beta(r, t) e^{-2 \int_0^r \alpha(r', t) dr'}. \quad (5)$$

- 5 The factor of proportionality is the calibration factor as can be shown using Equation (1) and
 6 (4). The algorithm to calculate the correction function $o(r, t)$ is based on two main
 7 assumptions:

- 8 1. The aerosol extinction and backscatter coefficients are constant in a range interval
 9 $[0, R]$ and during the time period of observation (assumption of homogeneous
 10 atmosphere).
- 11 2. The overlap function is known with low uncertainty in the range interval $[R_{OK}, \infty]$,
 12 with $R_{OK} \leq R$.
 13

14 Under these assumptions the aerosol lidar ratio is constant in the range $[0, R]$, hence the
 15 aerosol backscatter coefficient is proportional to the aerosol extinction coefficient in the
 16 considered range. The molecular backscatter coefficient depends on atmospheric density and
 17 varies with range.

18 In the range $[0, R]$ Equation (1) to (3) can be written as (time dependence neglected for
 19 clarity):

$$\log(\beta_{raw}(r)) + \log(o(r)) = \log(C_L) + \log(\beta_p) - 2\alpha_p r + \log\left(1 + \frac{\beta_m(r)}{\beta_p}\right) - 2 \int_0^r \alpha_m(r') dr'. \quad (6)$$

- 20 Using the aerosol lidar ratio L and a molecular lidar ratio equal to $8 \frac{\pi}{3}$, Equation (6) can be
 21 rewritten as follows:

$$\log(\beta_{raw}(r)) + \log(o(r)) = \log(C_L) + \log\left(\frac{\alpha_p}{L}\right) - 2\alpha_p r + \underbrace{\log\left(1 + \frac{3L\alpha_m(r)}{8\pi\alpha_p}\right)}_{A_1(r)} - 2 \underbrace{\int_0^r \alpha_m(r') dr'}_{A_2(r)}. \quad (7)$$



1 For a standard atmosphere and at a wavelength of 1064 nm, a lidar ratio between 20 and
 2 120 Sr and a particle extinction coefficient between 0 and 100 Mm⁻¹, the 5th term (A_2) is in the
 3 order of 0.01% of the total signal. It is being neglected for the rest of the calculations. Noting
 4 that the 4th term (A_1) is in good approximation a straight line, the right hand side of Equation
 5 (7) forms itself in good approximation a straight line:

$$\log(\beta_{\text{raw}}(r)) + \log(o(r)) = A + B r \quad \forall r \in [0, R]. \quad (8)$$

6 Assuming further that $O_{\text{CHM}}(r)$ is correct in the range $[R_{\text{OK}}, R]$, i.e. $\log(o(r)) = 0 \quad \forall r \in$
 7 $[R_{\text{OK}}, R]$, the coefficients A and B are obtained from fitting Equation (8) to the data in this
 8 same range.

9 The correction function in the range $[0, R]$ is given by the difference between the fit (right
 10 hand side of Equation (8) and the data as follows:

$$o(r) = e^{-\left(\log(\beta_{\text{raw}}(r)) - (A + Br)\right)} \quad \forall r \in [0, R]. \quad (9)$$

11 3.2 Outline of the algorithm

12 While the approach presented in the previous section is quite straight forward, the
 13 implementation of an automatic algorithm is not. The most difficult parts are the selection of
 14 favourable atmospheric conditions and the quality control of the result. These two aspects are
 15 discussed in detail in the Appendix A, while only a brief description of the algorithm is given
 16 in the following.

17 The algorithm processes a swath of 24h of data for which one overlap correction function is
 18 derived. The swath is split into 282 intervals of length $\Delta T = 30\text{min}$ with starting times t_i
 19 every 5min from 00:00 to 23:30. For each time interval the mean profile is computed and the
 20 fitting interval $[R_{\text{OK}}, R_{\text{MAX}}]$ is determined, where Equation (8) can be fit to the mean profile.
 21 The lower boundary R_{OK} of the fitting interval represents the lowest range where the overlap
 22 function is known with satisfactory accuracy and the upper boundary R_{MAX} represents the
 23 maximum range where the atmosphere is homogeneous. While R_{OK} is instrument specific and
 24 constant throughout the processing, R_{MAX} has to be determined for each time interval as
 25 described in Appendix A. A series of fits is performed in the fitting interval $[R_{\text{OK}}, R_{\text{MAX}}]$ from
 26 which each one undergoes a sequence of quality checks to evaluate quality and plausibility of
 27 the fit itself and the obtained overlap correction functions. The final overlap correction



1 function for the entire swath is taken as the median of all overlap correction functions that
 2 passed the quality check.

3 **4 Results**

4 **4.1 Case Study: 16 June 2014**

5 An example of a successful correction of the overlap function is shown in Figure 1. This day
 6 is representative of a typical planetary boundary layer development (Stull, 1988). The residual
 7 layer is visible during night as well as the convective layer that developed during the day. An
 8 enhancement of the signal centered at 250 m is visible all day (Figure 1-a). This feature
 9 becomes very pronounced when plotting the gradient of the range corrected signal (Figure 1-
 10 b) and must be attributed to artefacts induced by inaccuracies in the overlap function provided
 11 by the manufacturer.

12 The algorithm described in Section 3.2 was applied for this day. The areas marked with
 13 dashed lines indicate the time and height intervals where Equation (8) could be fit to the data.
 14 For this day, 144 overlap correction functions were selected by the algorithm for 44 out of the
 15 282 time intervals of the swath (for details see Appendix A). The original and the corrected
 16 overlap functions are shown in Figure 2. The overlap function provided by the manufacturer
 17 agrees well down to 600 m. Below, the original overlap function underestimates overlap by
 18 up to 45% around 300 m (where the overlap value provided by the manufacturer is about 0.2).
 19 The median of the corrected overlap functions was applied to the range corrected signal
 20 (Figure 1-c) and the gradient was recalculated (Figure 1-d). The example demonstrates nicely
 21 that the artefact disappears when the overlap correction is applied.

22 **4.2 Long term variability**

23 The algorithm was applied to the ceilometer measurements taken in Payerne from the 08
 24 February 2013 to the 25 November 2014. The instrument was pointing vertically and
 25 achieved a data availability of 99.24%. It has not been moved during this time period. Out of
 26 the 651 days of operation an overlap correction could be derived for 153 days (23.50% of all
 27 the analyzed data). The success rate of the algorithm shows a strong seasonal cycle with a
 28 higher success rate in summer than in winter (see Figure 3). This is explained by the fact that
 29 in Winter, the site is often affected by low clouds and fog. Moreover the homogeneous



1 atmospheric conditions often do not reach the required height due to the shallow boundary
2 layer.

3 The obtained overlap functions (Figure 4) show a large variability and discrepancies up to
4 50% with respect to the values provided by the manufacturer. A seasonal cycle is present in
5 the overlap correction with higher values in Summer than in Winter (not shown).

6 Assuming that this seasonal cycle is caused by variations of the temperature of the
7 components, the daily overlap functions in Figure 4 are displayed as a function of the median
8 of the internal temperature measurements corresponding to the successful candidates (see
9 Section 3.2 and Appendix A). Figure 4 reveals a clear dependence of the overlap function on
10 the internal temperature with higher values for warmer temperatures. It can further be seen
11 that the overlap function provided by the manufacturer corresponds to corrected overlap
12 functions at low internal temperatures. This temperature dependence is further analyzed in the
13 following section and a model to correct for temperature effects is proposed.

14

15 **5 Effect of the internal temperature**

16 Fluctuations of the ambient temperature influence the temperature of the laser and the optical
17 and electronic components. According to the manufacturer, the most temperature dependent
18 part of the system is the spatial sensitivity of the photo detector (personal communication).
19 This in turn affects directly the overlap function.

20 The norm of the relative difference between corrected and uncorrected signal is represented as
21 function of the internal temperature (Figure 5) revealing a clear correlation. The difference
22 between the overlap function provided by the manufacturer and the overlap function
23 calculated by the algorithm is increasing with the temperature.

24 The impact of the temperature on the overlap function is now revealed and can be
25 investigated further on. Figure 6-a shows the relative difference between the corrected and
26 uncorrected signals at each altitude. The shape of the relative difference is in agreement with
27 the artefact described in Section 4.1. On this figure, the color of each line is given by the
28 temperature. The difference between corrected and uncorrected signal reached 45% at a range
29 of 250 m for the 07 June 2014 when the median internal temperatures was over 35°C. On the



1 opposite, when the internal temperature was below 20°C on the 11 March 2014, the difference
2 decreased to 20%.

3 In the following a simple model to correct this temperature effect is described. At each range
4 the relative difference is assumed to depend linearly on the mean internal temperature. The
5 coefficients for each range are determined by linear fitting of the relative difference at this
6 range (Figure 6-a). The resulting model is presented in Figure 6-b. To better highlight the
7 temperature dependence in Figure 6-a, 23 outliers have been identified and discarded (out of
8 the 153 daily corrections). However, to calculate the model coefficients used throughout the
9 study all data points were considered.

10 The performance of the model to correct artefacts is assessed in the next section. Major
11 advantages of the model are the possibilities to correct for short term variations on scales of
12 hours (day/night) and to correct data in real time.

13 Unfortunately the coefficients of the temperature model are instrument specific and cannot be
14 used for other instruments or even for other optical modules. However, the algorithm
15 described in 3.2 can be used on any CHM15k to determine the appropriate overlap correction
16 if the data set is long enough and covers the entire range of internal temperatures that have to
17 be expected for the site.

18

19 **6 Effect of the overlap correction on edge detection**

20 In almost all boundary layer detection algorithms using aerosols as tracers, the detection of
21 edges or gradients in the backscatter data is the first step. More or less sophisticated
22 approaches are then chosen to attribute one of the detected edges or gradients to the planetary
23 boundary layer height. This attribution is a very important step in the detection of the
24 planetary boundary layer but is beyond the scope of this study. Therefore this section is
25 limited to demonstrate the effect of our overlap correction method on the detection of aerosol
26 gradients. It is obvious that removing false candidates will naturally improve also the
27 attribution procedure.



1 **6.1 Case Study: 15 July 2014**

2 In Figure 7, the performance of the temperature model is compared with corrections made
 3 with a single daily overlap function (like in Section 4.1). Figure 7-a, Figure 7-c and Figure 7-e
 4 show the logarithm of the range corrected signal (S) measured at Payerne on the 15 July 2014.
 5 For this day an aerosol layer up to roughly 1500m is clearly visible. Figure 7-b, Figure 7-d
 6 and Figure 7-f show the corresponding gradient calculated together with the time series of the
 7 three strongest gradients as well as the lowest gradient. The gradients were calculated every 5
 8 minutes from smoothed range corrected signals (below the cloud base height if any) and
 9 gradients of low magnitude were neglected.

10 If no correction is applied on CHM15k measurements, the strongest gradient is very often
 11 located at a constant altitude (Figure 7-a). By applying the algorithm described in Section 3.2,
 12 an overlap correction was determined using a homogeneous layer below 800m from 00:30 to
 13 01:30 (Figure 7-c and Figure 7-d). Using this overlap correction significantly improved the
 14 detection of the strongest gradient at the top of the aerosol layer around 1100m. For this day,
 15 the external temperature was varying between 11°C to 25°C and the internal temperature
 16 between 22°C to 30°C. This change in temperature had an impact on the overlap function,
 17 thus, the overlap correction retrieved around 01:00 does not perfectly correct the overlap
 18 artefact for the entire day. With the temperature model described in Section 5, the artefact can
 19 be almost perfectly removed for the entire day (Figure 7-f). Consequently, false candidates
 20 due to the artefact induced by inaccuracies in the overlap function could be almost completely
 21 removed (Figure 7-e).

22 **6.2 Long term variability**

23 The impact of the overlap correction on the detection of the strongest gradient was tested for
 24 the years 2013 and 2014. Like in Section 6.1, gradients were calculated every 5 minutes, and
 25 the strongest gradient at each time step was selected. The strongest gradient has been chosen
 26 since this can be considered as a simple attribution solution to the boundary layer (Haeffelin
 27 et al., 2012). Figure 8 represents the frequency distribution of the height of this strongest
 28 gradient. Uncorrected data are shown in red and the results after the correction with the model
 29 are shown in green. For the uncorrected data, a clear spike is visible around 360m. This spike
 30 corresponds to the artefact induced by the uncorrected overlap function described previously.



1 After the correction this spike almost disappears and permits more detections between 400
2 and 1000 m which are physically meaningful. These gradients were previously masked by
3 some erroneous detections at the altitude of the spike around 360m.

4 The presented correction method has thus the potential to significantly improve the detection
5 of the boundary layer using gradient based methods because it removes false candidates, e.g.
6 in situations of well-mixed convective boundary layer, and hence simplifies the attribution of
7 the detected gradients to the planetary boundary layer. A particularly high benefit can be
8 expected for the detection of shallow stable layers typical for nighttime situations.

9

10 **7 Summary and Conclusions**

11 Ceilometers are low-cost elastic lidars for unattended operation and state-of-the-art
12 instruments have the capability to perform aerosol profiling. This opens new applications like
13 alert systems in case of volcanic ash events, monitoring of long range transport of dust and
14 the determination of the planetary boundary layer height. However, the quality of the range
15 and overlap corrected signal, which is used in these applications, is often strongly degraded in
16 the first hundreds of meters because of imperfections in the specification of the overlap
17 function. Here a method has been presented to correct the overlap function, which is suited
18 for automatic use in large networks, since it does not require any manipulation of the
19 instrument. The method is based on the assumption that the atmosphere is homogeneous over
20 a given time and range interval, in which the overlap function is known with satisfying
21 quality. A straight line is fit to the data in this interval and a correction function can be
22 computed under the assumption that the atmosphere is also homogeneous from the ground up
23 to the lower boundary of the fitting range interval. The novelty of the method lies in the
24 implementation rather than in the approach itself, the latter being based on (Sasano et al.,
25 1979). A series of checks based on the spatio-temporal gradient is performed to identify
26 homogeneous conditions and the suited fitting interval. The obtained fits and the derived
27 correction functions for a 24 h swath of data undergo thorough quality checking using a
28 permutation scheme and severe tests for the homogeneity of the corrected data.

29 The analysis of 2 years of data revealed a distinct seasonal cycle in the corrected overlap
30 function. It was assumed that these variations are due to variations in the physical temperature
31 of the components. Therefore a model has been developed to compute the corrected overlap



1 function as a function of the internal temperature measured by the instrument, which is the
2 other novel aspect of the presented work. The temperature model has been used to correct
3 data and revealed that gradients related to artefacts induced by the overlap function can be
4 removed to greatest extent, even during cases where strong temperature differences between
5 day and night are present. The determination of the coefficients of the temperature model a
6 data set representative of a full seasonal cycle, i.e. of at least one year.. Once the coefficients
7 are determined, the temperature model allows correcting ceilometer data in real-time and
8 accounting for variations on short time scales and is therefore perfectly suited for the
9 application in large networks dedicated to real-time applications.
10



1 Appendix

2 A. Algorithm details

3 Parameters used in the following paragraphs are explained in Table 2 together with the values
 4 chosen for the implementation for a CHM15k lidar operated in the configuration as specified
 5 in Table 1. The algorithm processes a swath of 24h of data for which one overlap correction
 6 function is derived. First, the swath is split into 282 intervals of length $\Delta T = 30\text{min}$ with
 7 starting times t_i every 5min from 00:00 to 23:30.

8 Determination of the fitting intervals

9 During this step, it is determined if during the considered time interval $[t_i, t_i + \Delta T]$, $i \in$
 10 $1 \dots 282$, there is a range R_{MAX} below which the atmospheric conditions satisfy the
 11 assumptions of homogeneity and thus where fitting intervals $[R_1, R_2] \in [R_{OK}, R_{MAX}]$ can be
 12 constructed and tested.

13 Note that the R_{MAX} value may change from one time interval to another, and is limited by
 14 $R_{MAX,MAX}$, usually inside the boundary layer. The R_{OK} and $R_{MAX,MAX}$ values depend on the
 15 instrument and on the general atmospheric conditions and are fixed for all calculations.
 16 $R_{MAX,MAX}$ determines the maximum range below which homogeneous conditions can be
 17 expected. This parameter is not critical for the results but allows to save computational time.

18 In order to calculate R_{MAX} , the following series of checks are applied:

- 19 1. **Data availability and bad weather:** Data availability must be 100%, i.e. the time interval
 20 must consist here of 60 non-erroneous profiles, and within the time interval no
 21 precipitation or fog events (bad weather events) should occur, because these events result
 22 in saturated, inhomogeneous signals. Weather information is taken here on a profile-by-
 23 profile basis directly from the ceilometer's output (sky condition index), but it could be
 24 taken also from surface station measurements.
- 25 2. **Cloud and signal-to-noise limitation:** The fitting interval should not contain clouds
 26 (which result in peaks in the signal), and should not be too noisy. Therefore, the range
 27 R_{CLOUD} of the lowest cloud base height during the whole time interval is identified as well
 28 as the range of the lowest maximum detection height, R_{SNR} . Cloud base heights and
 29 maximum detection heights are taken here on a profile-by-profile basis directly from the
 30 ceilometer's output, but they could be calculated as well.
- 31 3. **Test for homogeneity:** Here we check if characteristic properties of a homogeneous
 32 atmosphere are present. The 60 profiles of $\log_{10}(abs(\beta_{raw}))$ are considered. For
 33 brevity, $\log_{10}(abs(\beta_{raw}))$ is hereafter referred to as S . At 1064 nm, a homogenous



atmosphere yields a profile of S close to being a line with a small negative slope, because of the small molecular influence. Therefore, almost vanishing spatial fluctuations of S are expected. These spatial fluctuations can however only be checked starting from the range R_{OK} where the overlap function is known with satisfactory accuracy, because below this range, artificial gradients may appear due to the incorrect manufacturer's overlap correction. Temporal fluctuations in S , which should remain small, are from the ground up, where the ground R_{GROUND} is taken here as the lowest range where the overlap function is larger than 0.05, because below this range the signal is usually too noisy to be processed. The interval $[t_i, t_i + \Delta T]$ is split into sub-intervals of duration $\Delta T_s = 10$ min starting every 30 s from t_i until $t_i + \Delta T - \Delta T_s$. All statistical variables and temporal gradients in the following are derived from these sub-intervals.

3.i. *Temporal homogeneity:*

- 3.i.i. For each range between R_{GROUND} and $R_{MAX,MAX}$, the ratio of the standard deviation over the median of S is calculated and the maximum value is kept in memory. The lowest range R_{STD} where this maximum value becomes greater than κ_1 is derived.
- 3.i.ii. For each range between R_{GROUND} and $R_{MAX,MAX}$ the norm of the temporal relative gradient:

$$\nabla_x^* S = \frac{|\nabla_x S|}{|S|} \quad (\text{A.1})$$

is calculated, with ∇_x being calculated with a Sobel operator. The lowest range R_{GRADX} where

$$\max(\nabla_x^* S(r, t)) \geq \kappa_2 \quad (\text{A.2})$$

with $(r, t) \in [R_{GROUND}, R_{GRADX}] \times [t_i, t_i + \Delta T]$, is derived.

- 3.ii. *Spatial homogeneity:* For each range between R_{OK} and $R_{MAX,MAX}$ the norm of the spatial relative gradient

$$\nabla_y^* S = \frac{|\nabla_y S|}{|S|} \quad (\text{A.3})$$

is calculated, with ∇_y being calculated with a Sobel operator. The lowest range R_{GRADY} where

$$\max(\nabla_y^* S(r, t)) \geq \kappa_2 \quad (\text{A.4})$$

with $(r, t) \in [R_{OK}, R_{GRADY}] \times [t_i, t_i + \Delta T]$ is derived.

- 3.iii. *Spatial and temporal homogeneity:* For each range between R_{OK} and $R_{MAX,MAX}$ the norm of the two dimensional relative gradient is calculated with the following equation:



$$\nabla_{XY}^* S = \sqrt{\left| \frac{\nabla_X S}{S} \right|^2 + \left| \frac{\nabla_Y S}{S} \right|^2} \quad (\text{A.5})$$

1 The lowest range R_{GRADXY} is derived, where

$$\max(\nabla_{XY}^* S(r, t)) \geq \kappa_2 \quad (\text{A.6})$$

2 or where:

$$\text{mean}(\nabla_{XY}^* S(r, t)) \geq \kappa_3 \quad (\text{A.7})$$

3 with $(r, t) \in [R_{OK}, R_{GRADXY}] \times [t_i, t_i + \Delta T]$.

4

5 Once these bad weather, cloud, noise and homogeneity tests are done, the upper boundary of
 6 the fitting interval is set to

$$R_{MAX} = \min(R_{CLOUD}, R_{SNR}, R_{STD}, R_{GRADX}, R_{GRADY}, R_{GRADXY}). \quad (\text{A.8})$$

7 If R_{MAX} is smaller than $R_{OK} + \Delta R_{MIN}$ the time interval $[t_i, t_i + \Delta T]$ is rejected. If
 8 $R_{MAX} > R_{MAX, MAX}$, we set its value to $R_{MAX, MAX}$, because the fitting part and subsequent quality
 9 check in the following are computationally costly.

10

11 **Quality check of the fits and determination of a set of overlap correction candidates**

12 The range interval $[R_{OK}, R_{MAX}]$ is now split into all possible intervals $[R_1, R_2]$ on the discrete
 13 range grid and of length equal to or larger than ΔR_{MIN} that fit into $[R_{OK}, R_{MAX}]$. In each such
 14 range interval $[R_1, R_2]$ the mean profile of S for the time interval $[t_i, t_i + \Delta T]$ is fit with a
 15 straight line according to Equation (8) and the obtained linear fits undergo the following
 16 series of checks:

- 17 4. **Plausibility of slope and ground value:** Under homogeneous conditions, the slope of the
 18 fit is approximately $\frac{-2}{\log(10)} \alpha_p$ and the y-axis offset is approximately $\log_{10}(C_L) +$
 19 $\log_{10}(\frac{\alpha_p}{L})$. Note that the factor $\log(10)$ is needed because S is calculated with the log with
 20 base 10. Bounds based on estimations of reasonable values for α_p , C_L and L can be set,
 21 such that the slope must lie between κ_4 and κ_5 and the y-axis offset must lie between κ_6
 22 and κ_7 .
- 23 5. **Goodness of fit:** The RMSE of the fit divided by the mean of the fit must be smaller than
 24 κ_8 .



1
 2 The linear fits that successfully passed these checks form a set of candidates to be used to
 3 derive the overlap correction.

4

5 **Quality check of the overlap correction candidates**

6 For each such candidate, with its fitting range $[R_1, R_2]$ as unique identifier, the corrected
 7 overlap function, O_{corr} , is computed using Equation (2) and (9) where $O_{corr(R \geq R_2)} =$
 8 $O_{CHM}(R \geq R_2)$. The corrected overlap function is checked for plausibility with the following
 9 series of checks:

- 10 6. **Maximum value:** Corrected overlap functions showing unphysically high values are
 11 discarded. Therefore, $\max(O_{corr}) / \max(O_{CHM})$ must be smaller than $\kappa_9 = 1.01$.
- 12 7. **Small relative error with respect to the manufacturer's overlap in the full overlap**
 13 **region:** The relative error $\frac{|O_{corr}(R) - O_{CHM}(R)|}{|O_{CHM}(R)|}$ must be smaller than $\kappa_{10} = 0.01$ for the ranges
 14 $R \geq R_{O_{CHM}=1}$ (range of full overlap, where it is assumed that the manufacturer's overlap
 15 is exact). For the CHM15k, $R_{O_{CHM}=1}$ can vary from instrument to instrument between 500
 16 and 2000 m.
- 17 8. **Temporal and spatial homogeneity:** The 60 profiles of
 18 $S_{corr} = \log_{10}(\text{abs}(\beta_{rawcorrected}))$ obtained from Equation (3) with the corrected
 19 overlap function (Equation 2) are now considered. The relative spatio-temporal gradients
 20 $\nabla_{XY}^* S_{corr}$ are calculated as in test 3.iii "Spatial and temporal homogeneity". Temporal and
 21 spatial fluctuations are expected to be small for all ranges from R_{GROUND} to R_2 . Therefore
 22 the following conditions must be satisfied:

$$\max(\nabla_{XY}^* S_{corr}(r, t)) < \kappa_2 \quad (\text{A.9})$$

$$\text{mean}(\nabla_{XY}^* S_{corr}(r, t)) < \kappa_3 \quad (\text{A.10})$$

23

24 with $(r, t) \in [R_{GROUND}, R_2] \times [t_i, t_i + \Delta T]$.

- 25 9. **Monotonic increase:** An overlap function should be monotonically increasing until the
 26 range of full overlap, therefore only a small negative slope (resulting from small
 27 inhomogeneities in the correction) should be allowed. The slope of O_{corr} , computed with
 28 a Savitzky-Golay filter of width 5 and order 3, must be larger than $\kappa_{11} = -0.00025 \text{ m}^{-1}$
 29 between 0 and R_2 , i.e. a decrease of maximum $0.015\% \text{ m}^{-1}$ is allowed.

30

31 **Final Selection**



1 All successful candidates obtained from each time interval $[t_i, t_i + \Delta T]$ are kept in a global
2 list of candidates for the entire swath (24 h). For the entire swath a minimum of 15 candidates
3 must be obtained, otherwise the swath is rejected for the calculation of an overlap correction.
4 Under the assumption that the overlap function does not change much within one swath, each
5 candidate is checked in the time interval of all other candidates with test 8 and test 3.i.i from
6 range R_{GROUND} to their ranges R_2 . From the successful candidates outliers are removed,
7 whereas an outlier lies outside of 3 interquartile ranges from the median with respect to both
8 slope and y-axis offset. If the final set contains more than 10 candidates, the final overlap
9 correction is the median overlap correction. Otherwise, the swath is rejected.

10 Note that checking each candidate in the time interval of all other candidates is omitted if the
11 number of candidates exceeds 100, in order to save computational time and because it is
12 likely that for such a large number of candidates, the incorrect ones will be filtered out during
13 the outlier removal step.



1 Acknowledgments

2 This study has been financially supported by [ICOS-CH](#) and [E-PROFILE \(EUMETNET\)](#). The
 3 authors would further like to thank Gianni Martucci, Robert J. Sica and Martial Haeffelin for
 4 their constructive remarks. The authors would like to acknowledge the contribution of the
 5 COST Action ES1303 (TOPROF). The authors are grateful to Kornelia Pönitz and Holger
 6 Wille (Lufft) for technical information about the CHM15k.

8 References

- 9 Biavati, G., Donfrancesco, G. D., Cairo, F. and Feist, D. G.: Correction scheme for close-
 10 range lidar returns, *Appl. Opt.*, 50(30), 5872, doi:10.1364/AO.50.005872, 2011.
- 11 Campbell, J. R., Hlavka, D. L., Welton, E. J., Flynn, C. J., Turner, D. D., Spinhirne, J. D.,
 12 Scott, V. S. and Hwang, I. H.: Full-Time, Eye-Safe Cloud and Aerosol Lidar Observation at
 13 Atmospheric Radiation Measurement Program Sites: Instruments and Data Processing, *J.*
 14 *Atmospheric Ocean. Technol.*, 19(4), 431–442, doi:10.1175/1520-
 15 0426(2002)019<0431:FTESCA>2.0.CO;2, 2002.
- 16 Flentje, H., Claude, H., Elste, T., Gilge, S., Köhler, U., Plass-Dülmer, C., Steinbrecht, W.,
 17 Thomas, W., Werner, A. and Fricke, W.: The Eyjafjallajökull eruption in April 2010 –
 18 detection of volcanic plume using in-situ measurements, ozone sondes and lidar-ceilometer
 19 profiles, *Atmos Chem Phys*, 10, 10085–10092, doi:10.5194/acp-10-10085-2010, 2010.
- 20 Guerrero-Rascado, J. L., Costa, M. J., Bortoli, D., Silva, A. M., Lyamani, H. and Alados-
 21 Arboledas, L.: Infrared lidar overlap function: an experimental determination, *Opt. Express*,
 22 18(19), 20350, doi:10.1364/OE.18.020350, 2010.
- 23 Haeffelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G. P., Lolli, S.,
 24 O'Dowd, C. D., Sauvage, L., Xueref-Rémy, I., Wastine, B. and Feist, D. G.: Evaluation of
 25 Mixing-Height Retrievals from Automatic Profiling Lidars and Ceilometers in View of Future
 26 Integrated Networks in Europe, *Bound.-Layer Meteorol.*, 143(1), 49–75, doi:10.1007/s10546-
 27 011-9643-z, 2012.
- 28 Kuze, H., Kinjo, H., Sakurada, Y. and Takeuchi, N.: Field-of-View Dependence of Lidar
 29 Signals by Use of Newtonian and Cassegrainian Telescopes, *Appl. Opt.*, 37(15), 3128–3132,
 30 doi:10.1364/AO.37.003128, 1998.
- 31 Reichardt, J., Wandinger, U., Klein, V., Mattis, I., Hilber, B. and Begbie, R.: RAMSES:
 32 German Meteorological Service autonomous Raman lidar for water vapor, temperature,
 33 aerosol, and cloud measurements, *Appl. Opt.*, 51(34), 8111, doi:10.1364/AO.51.008111,
 34 2012.
- 35 Roberts, D. W. and Gimmestad, G. G.: Optimizing lidar dynamic range by engineering the
 36 crossover region, vol. 4723, pp. 120–129., 2002.
- 37 Sasano, Y., Shimizu, H., Takeuchi, N. and Okuda, M.: Geometrical form factor in the laser
 38 radar equation: an experimental determination, *Appl. Opt.*, 18(23), 3908,
 39 doi:10.1364/AO.18.003908, 1979.



- 1 Stelmaszczyk, K., Dell’Aglia, M., Chudzyński, S., Stacewicz, T. and Wöste, L.: Analytical
 2 function for lidar geometrical compression form-factor calculations, *Appl. Opt.*, 44(7), 1323,
 3 doi:10.1364/AO.44.001323, 2005.
- 4 Stull, R. B.: *An Introduction to Boundary Layer Meteorology*, Springer Science & Business
 5 Media., 1988.
- 6 Vande Hey, J., Coupland, J., Foo, M. H., Richards, J. and Sandford, A.: Determination of
 7 overlap in lidar systems, *Appl. Opt.*, 50(30), 5791, doi:10.1364/AO.50.005791, 2011.
- 8 Wandinger, U. and Ansmann, A.: Experimental Determination of the Lidar Overlap Profile
 9 with Raman Lidar, *Appl. Opt.*, 41(3), 511–514, doi:10.1364/AO.41.000511, 2002.
- 10 Welton, E. J. and Campbell, J. R.: Micropulse Lidar Signals: Uncertainty Analysis, *J.*
 11 *Atmospheric Ocean. Technol.*, 19(12), 2089–2094, doi:10.1175/1520-
 12 0426(2002)019<2089:MLSUA>2.0.CO;2, 2002.
- 13 Welton, E. J., Voss, K. J., Gordon, H. R., Maring, H., Smirnov, A., Holben, B., Schmid, B.,
 14 Livingston, J. M., Russell, P. B., Durkee, P. A., Formenti, P. and Andreae, M. O.: Ground-
 15 based lidar measurements of aerosols during ACE-2: instrument description, results, and
 16 comparisons with other ground-based and airborne measurements, *Tellus B*, 52(2), 636–651,
 17 doi:10.1034/j.1600-0889.2000.00025.x, 2000.
- 18 Wiegner, M. and Geiß, A.: Aerosol profiling with the Jenoptik ceilometer CHM15kx, *Atmos*
 19 *Meas Tech*, 5(8), 1953–1964, doi:10.5194/amt-5-1953-2012, 2012.

20



1 Tables

Parameter	Value
Integration time	30 s
Bin size	15 m
Maximum range	15 km
Overlap corrected by manufacturer	Yes, TUB120011_20121112_1024.cfg
Station	Payerne (Switzerland, 6.9417°N/46.8117°E)
Altitude	490m
Azimuth / zenith angles	0° / 0°
Wavelength	1064nm
Average repetition rate	6.5KHz
Average pulse energy	8μJ
Full overlap range	800m

2 Table 1: Instrument parameters.

3

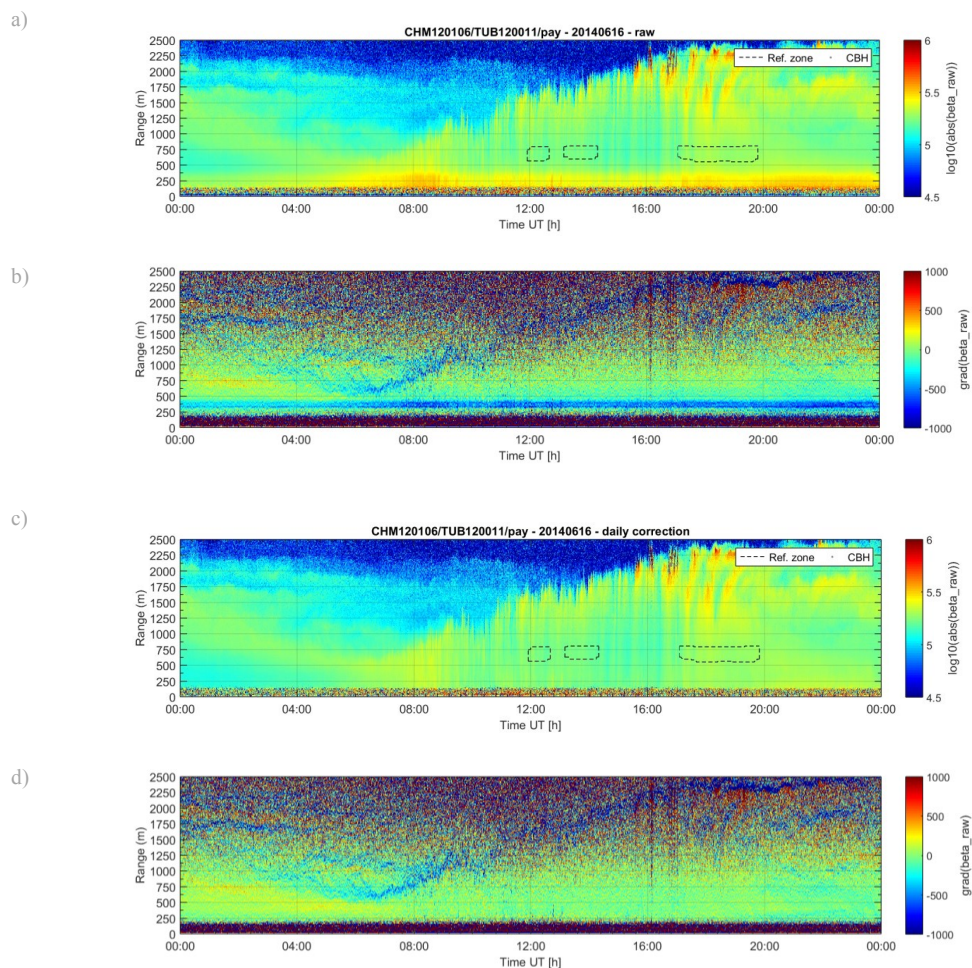


Parameter	Description	Value
R_{GROUND}	Lowest measurement range.	Lowest range where the overlap function provided by the manufacturer $\geq 5\%$
R_{OK}	Range above which the manufacturer's overlap function is accurate.	Lowest range where the overlap function provided by the manufacturer $\geq 80\%$
$R_{MAX,MAX}$	Highest allowed range for the fitting.	1200 m
$R_{OCHM=1}$	Lowest range where the manufacturer's overlap function reaches 1 (full overlap).	Lowest range where the overlap function provided by the manufacturer $\geq 100\%$
ΔR_{MIN}	Minimum length of the fitting intervals.	150 m
κ_1	Upper threshold for the ratio of the standard deviation over the median.	0.01
κ_2	Upper threshold for the relative gradient.	0.05
κ_3	Upper threshold for the mean relative gradient.	0.015
κ_4	Lower threshold for the slope of the linear fit.	$\frac{-2}{\log(10)} 10^{-5}$
κ_5	Upper threshold for the slope of the linear fit.	$\frac{-2}{\log(10)} 10^{-7}$
κ_6	Lower threshold for the y-axis offset of the linear fit.	4.75
κ_7	Upper threshold for the y-axis offset of the linear fit.	6
κ_8	Upper threshold for the relative RMSE of the linear fit.	0.0005
κ_9	Upper threshold for the ratio between the maximum values of the corrected overlap function and the manufacturer's overlap function .	1.01
κ_{10}	Upper threshold for the relative error of the corrected overlap function w.r.t. the manufacturer's overlap function in the full overlap region.	0.01
κ_{11}	Lower threshold for the slope of the corrected overlap function.	-0.00025

Table 2: Algorithm parameters.

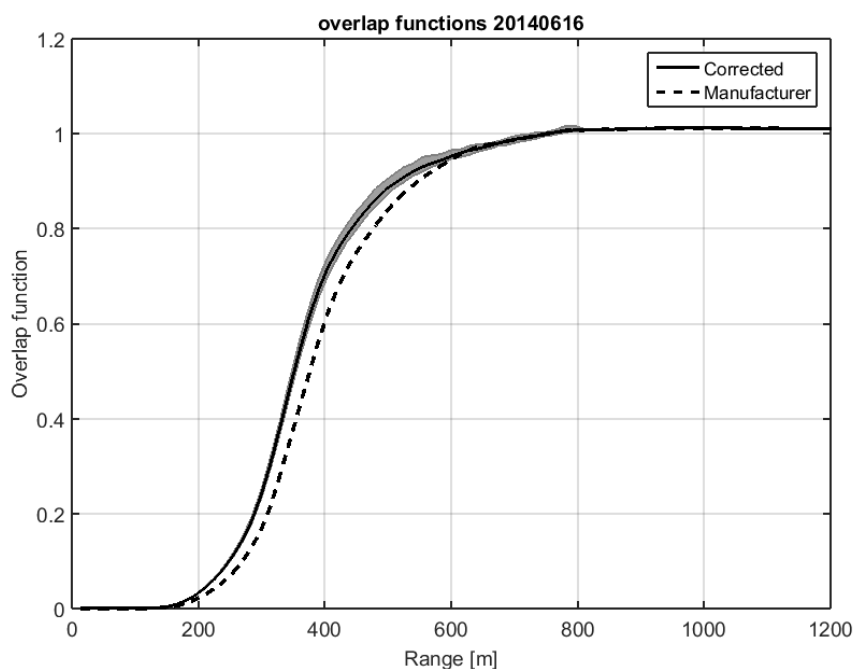


1 Figures



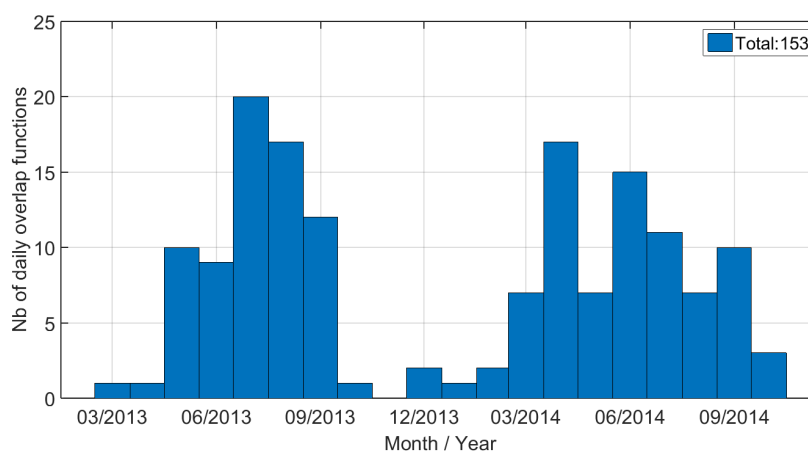
2 Figure 1: CHM15k Measurements at Payerne for the 16 June 2014. a) and c): logarithm of the range corrected signal. b)
 3 and d): Gradient of the range corrected signal. a) and b): without correction. c) and d): with overlap correction. The
 4 reference zones from which the overlap correction was calculated are circled with black dash lines.

5



1
 2 **Figure 2: Overlap functions for 16 June 2014. The thick black line is the median overlap function for this day. The dash**
 3 **line represent the overlap function provided by the manufacturer.**

4



5
 6 **Figure 3: Success rate of the algorithm for 2 years of data.**

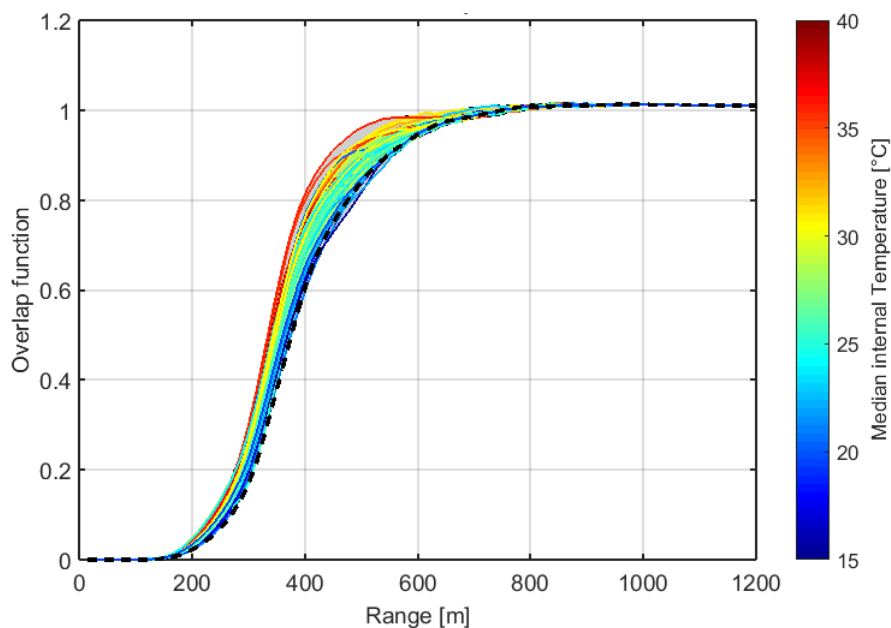


Figure 4: Overlap functions retrieved for payenne ceilometer in 2013 and 2014. The colors represent the ceilometer internal temperature when the overlap functions were calculated.

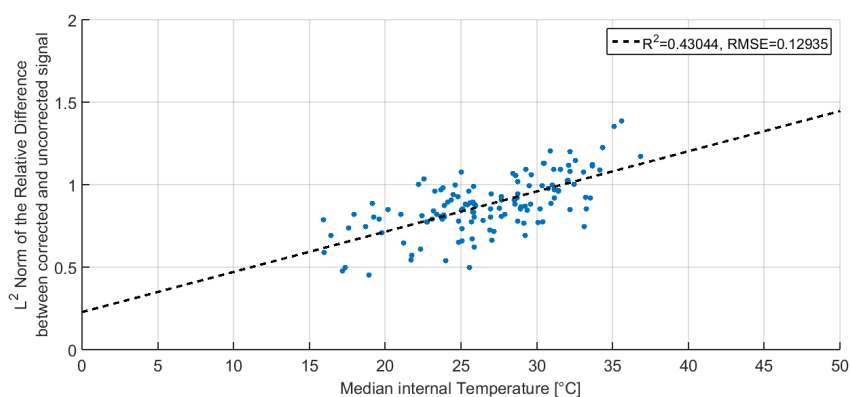
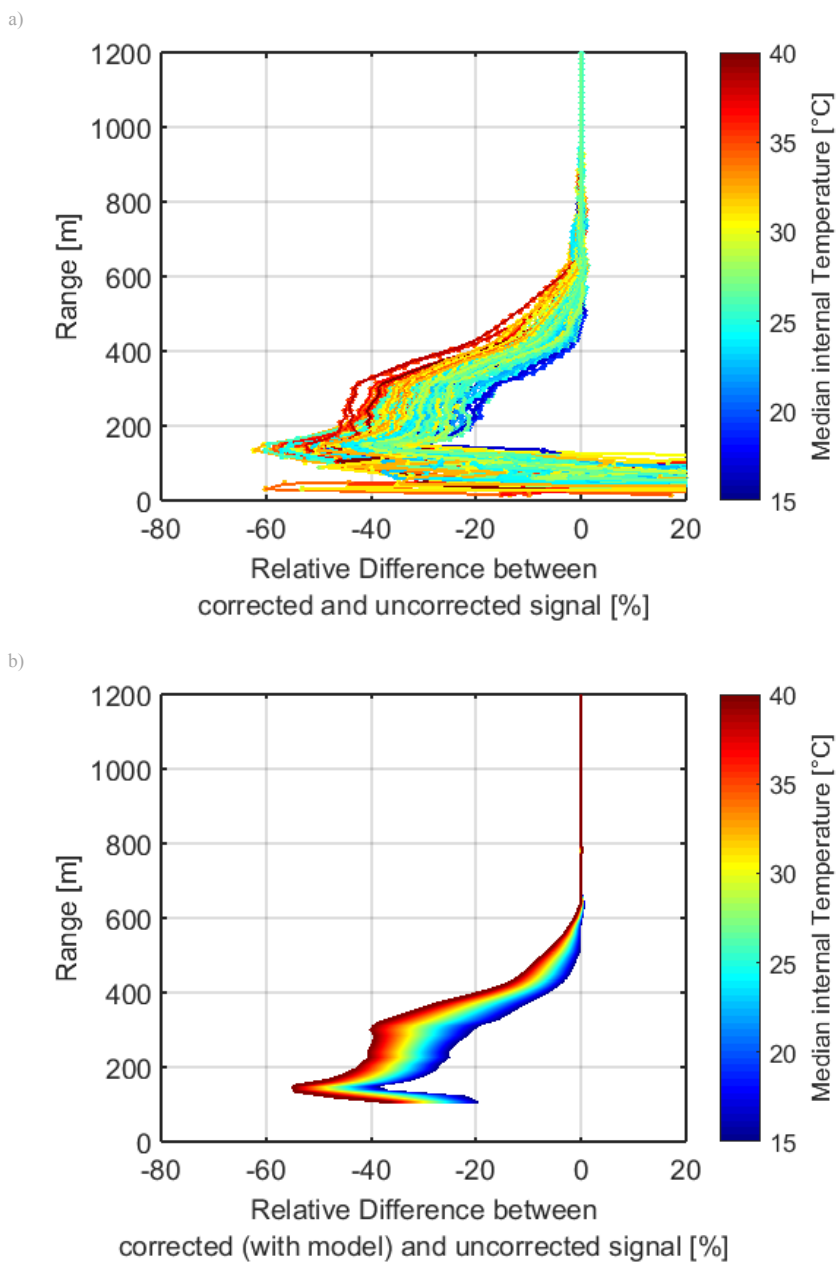
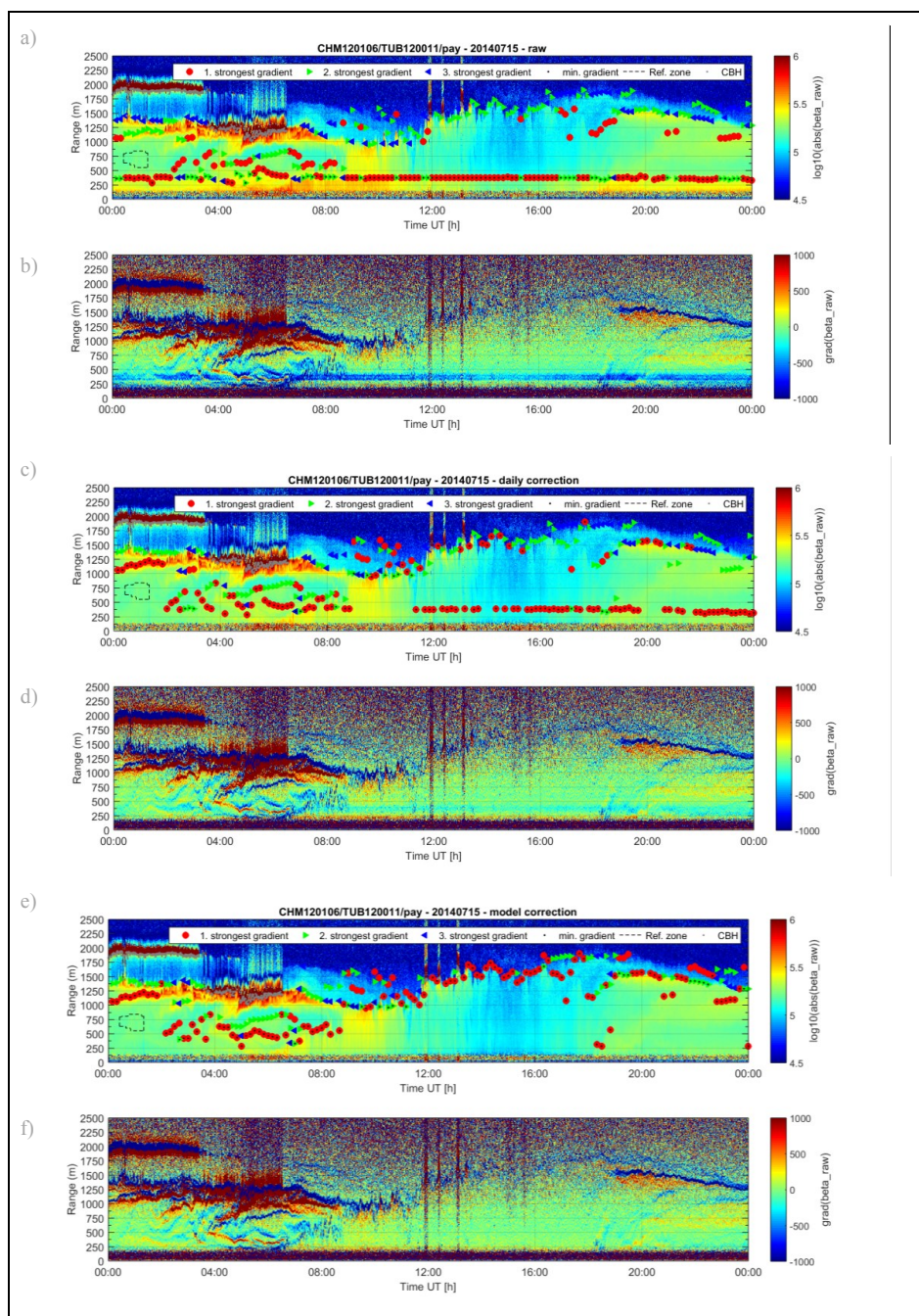


Figure 5: Relative difference between corrected and uncorrected signal against internal temperature



1 Figure 6: Relative difference between corrected and uncorrected signal. Upper Panel: From measurements. Lower panel:
 2 with model. The color is representing the internal temperature of the instrument.

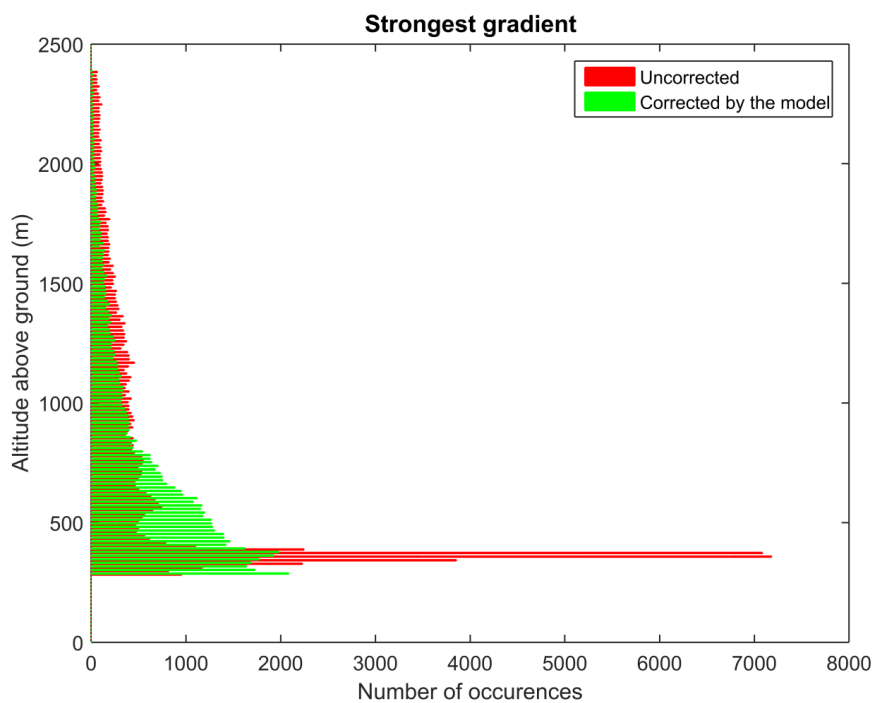
3



1 Figure 7: Times series PBL retrievals for the 15 July 2014. The red markers show the strongest gradient detected. Before
 2 correction (a and b), with daily correction (b and c) and with temperature model correction (c and d).



1



2

3 **Figure 8: Histogram of the altitude of the 5-min strongest gradients calculated in 2013 and 2014. Uncorrected data are**
 4 **represented in red. Data corrected with the temperature model are represented in green.**

5