Improved method for temporally interpolating radiosonde profiles

Linus von Klitzing[1], David D. Turner[2], Diego Lange[1], and Volker Wulfmeyer[1]

[1]Institute of Physics and Meteorology, University of Hohenheim, Stuttgart, Germany
[2]NOAA Global Systems Laboratory, Boulder, CO, USA

Correspondence: Linus von Klitzing (linus.vonklitzing@uni-hohenheim.de)

Abstract. A significantly improved technique for temporally interpolating radiosonde profiles of potential temperature and water vapor mixing ratio (WVMR) during daytime is introduced. The key innovation of this technique is its operation on a height grid normalized with the planetary boundary layer height (PBLH). This study utilized a three-month dataset of three-hourly soundings from the Atmospheric Radiation Measurement Program's (ARM) Southern Great Plains (SGP) site. The technique
5   was evaluated for convective boundary layer (CBL) cases, with the necessary PBLH data obtained from a ground-based infrared spectrometer. A total of 79 comparisons were conducted between reference soundings and interpolated profiles that did and did not employ height normalization (HN). The results demonstrated a substantial improvement in the representation of interpolated profiles using the new technique, characterized by enhanced correlation, improved amplitude representation, and reduced bias for potential temperature, as well as improved correlation and reduced bias for WVMR.

1   Introduction

Atmospheric science relies heavily on radiosonde (RS) soundings, which provide an accurate, high vertical resolution way of profiling the atmosphere. From the surface to altitudes of 20 to 30 km, RSs conduct measurements of temperature, pressure, humidity, and wind during their ascent with high frequency and under all weather conditions (Jensen et al., 2016; Wulfmeyer et al., 2015).

15   This dependency on RSs has drawbacks, as the sounding locations are relatively scarce (Wulfmeyer et al., 2015), the soundings are not vertical, and the launches occur routinely only two to four times per day, even at larger facilities such as the Atmospheric Radiation Measurement Program (ARM) Southern Great Plains (SGP) site considered in this work (Jensen et al., 2016).

However, many applications require a higher temporal resolution than 6 or 12 h. Consequently, temporal interpolations of
20   RS profiles with high accuracy are very important. To address this need for the SGP site, ARM routinely provides a simple interpolation product, which linearly averages the nearest soundings on a fixed height grid (Fairless et al., 2021). An example of the application of this data product is the multisensor cloud retrieval product MICROBASE, which aims to produce macro- and microphysical properties of all clouds over the site. Its strength is its ability to apply to all clouds and conditions, enabled by various retrieval techniques, but more importantly for this work, it includes temporally interpolated RS data (Shupe et al.,
25   2016; Dunn et al., 2011; Troyan, 2010). This interpolated RS data is used right at the beginning of the corresponding algorithm

to determine the cloud water phase and, with that, the radar reflectivity (Dunn et al., 2011). More accurate data can be expected to improve the accuracy of this step and, consequently, the accuracy of the data product.

The improvement proposed in this work is straightforward. Rather than interpolating between two soundings on a fixed height grid, the new technique employs a normalized height grid, which is obtained by dividing the standard height grid by the planetary boundary layer height (PBLH) (Rosenberger et al., 2024). For the time being, the applicability of this new technique is limited to the case of a convective boundary layer (CBL). The CBL consists of the unstable surface layer, the mixed layer where turbulence brings potential temperature and humidity to a relatively constant level with height, and the interfacial layer that separates the free atmosphere (FA), characterized by gradients in temperature and moisture, which also signify the PBLH (Stull, 1988). The relative heights and corresponding temperature and humidity levels of these layers may change throughout the day, but the overall layering remains, which is exactly the fact that the new technique exploits. Therefore, knowledge of the PBLH enables scaling soundings to the same relative height levels and interpolating only physically related regimes.

The proposed interpolation technique should therefore apply to all atmospheric situations that exhibit a similar temporal behavior and vertical structure in the CBL with clearly identifiable layers. Another example is the nighttime boundary layer, consisting of a stable boundary layer and a residual layer, and with clear expectations for the profiles of temperature and humidity (Stull, 1988). Although being limited to specific atmospheric situations may seem like a drawback, the case of the investigated CBL plays a unique role in atmospheric science. Intensive research is conducted on the CBL as it is essential for understanding atmospheric processes and feedbacks (Wulfmeyer et al., 2016; Behrendt et al., 2020).

Applying this new interpolation technique requires a time-wise continuous and smooth PBLH estimate. Fortunately, this can be derived from comparably simple instruments, making it possible to acquire a relatively precise picture of the boundary layer using only basic equipment. It may not be feasible for smaller measurement sites, nor necessary, depending on the requirements, to employ complex, active remote sensing systems, which also depend on the weather situation (Wulfmeyer et al., 2015).

The primary scientific question of this study can be stated as follows: Does temporally interpolating RS profiles on a normalized height grid yield a more accurate representation of the vertical structure in the CBL compared to interpolating on a standard height grid?

This work focuses on the interpolation of potential temperature and water vapor mixing ratio (WVMR) profiles and begins with a concise overview of the data utilized and the underlying instrumentation in Sect. 2. The selection process for suitable cases and the interpolation technique are presented in Sect. 3. Thereafter, the results of this work are explained in Sect. 4, which includes an example case and a statistical analysis. Finally, the key findings are summarized in Sect. 5.

2 Observations

This study utilizes measurements and data products from the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement Program (ARM) Southern Great Plains (SGP) site in Oklahoma, USA (Sisterson et al., 2016). The ARM program aims to support atmospheric science by providing a broad spectrum of meteorological measurements collected in key climate regions. The SGP site covers a northern hemisphere, continental, mid-latitude area, providing a "continuous multivariable ob-

servationally based dataset" of the atmosphere (Sisterson et al., 2016). It is equipped with a wide range of surface-based and remote-sensing instruments, enabling synergies and intercomparisons of different data products. A detailed description of the site and the present atmospheric conditions is given by Krishnamurthy et al. (2021a).

This study uses data from 17 April to 21 July 2019, a period that is particularly well-suited for this research on RSs due to the frequent launches every three hours as part of a special field campaign (Weckwerth et al., 2020). Additionally, PBLH time series data provided by the Tropospheric Remotely Observed Profiling via Optimal Estimation (TROPoe) algorithm, which retrieved temperature and humidity profiles from the Atmospheric Emitted Radiance Interferometer (AERI) at the site, as well as Raman lidar data, are used for a quantitative and qualitative comparison. During the period under review, sunrise occurred between 11:00 and 12:00, and sunset occurred between 1:00 and 2:00 UTC.

2.1 Radiosondes

Since the end of 2017, the Vaisala RS41 has been the standard RS model at the ARM SGP site. During its flight, the RS41 sampled temperature, pressure, relative humidity, and wind at a rate of once per second, resulting in a height resolution of about 5 m for a nominal ascent rate of 5-5.5 m s^{-1} (Holdridge, 2020). The corresponding data product also includes secondary quantities, such as dew point temperature and height above sea level (Keeler et al., 2022), which allows for the calculation of potential temperature and WVMR. For more details and a comparison with the previous RS model, see the work by Jensen et al. (2016).

Typically, launches take place every six hours at 5:30, 11:30, 17:30, and 23:30 UTC. However, during the period under review, three-hourly launches were conducted, resulting in additional take-offs at 2:30, 8:30, 14:30, and 20:30 UTC (Weckwerth et al., 2020). This yields four to five soundings during daylight hours.

2.2 AERI and TROPoe

The required smooth time series of the PBLH is obtained through the TROPoe algorithm that retrieves thermodynamic profiles from the AERI. The AERI is an autonomous, passive, ground-based infrared spectrometer developed for the ARM program and deployed to all of its primary sites, including the SGP site (Turner et al., 2016b). It measures downwelling radiance from the atmosphere with a radiometric accuracy better than 1% of the ambient radiance, thanks to a precise calibration with two blackbodies (Knuteson et al., 2004).

Temperature and humidity profiles are retrieved from the AERI output via the TROPoe algorithm on a fixed vertical height grid with a resolution that exponentially decreases from 25 m near the surface to 800 m at a height of 3 km. The TROPoe algorithm uses observations and a Gaussian a priori probability density function that describes the mean atmospheric state to iteratively estimate the current atmospheric conditions via a forward radiative transfer model. It can provide a complete error covariance matrix for the solution and has a mean bias of less than 0.2 K and 0.3 g kg^{-1} for temperature and WVMR, respectively, under optimal conditions and atmospheric heights below 2 km (Turner and Löhnert, 2014). In addition to the primary variables, which include temperature and WVMR, the data product also includes other derived quantities, such as the required PBLH (Turner, 2021).

It should be noted that the use of AERI and TROPoe is not strictly necessary. The only essential requirement is the availability of a temporally smooth PBLH time series, which can originate from different active or passive remote sensing instruments (Duncan Jr. et al., 2022; Kotthaus et al., 2023), such as lidars, and can even include machine learning techniques (Krishnamurthy et al., 2021b). Any suitable data product could have been used for this purpose. However, the specific charm of AERI and TROPoe is their routine character, which makes them easily adoptable at other ARM facilities.

2.3 Raman lidar

It is desirable to compare the RS interpolation not only at a single point in time but throughout the entire daily cycle. For WVMR, this is achievable using the temporally and spatially high-resolution output of the SGP Raman lidar (Newsom et al., 2014; Turner and Goldsmith, 1999; Turner et al., 2016a). Its functional principle is based on Raman scattering, which involves detecting and analyzing the energetic shift in the backscattered signal of the emitted laser pulses. These pulses have a wavelength of 355 nm and a frequency of 30 HZ. The raw data has a spatial resolution of 7.5 m and a temporal resolution of 10 s, although in this study, a data product with 10 min averages is used (Newsom et al., 2022).

Important to note is the fact that this data product is only used for a qualitative visual comparison and is not required for the presented interpolation itself.

3 Case selection and interpolation

From the three available months, those RS launches were selected, which took place within a CBL six or nine hours apart from one another. This allows for one or two in-between soundings that can be compared to time-wise corresponding interpolations. Two different types of temporal interpolations were performed for these cases. First, the interpolation without an height normalization (HN), which represents the current standard procedure of the ARM program, and second, the one with such a normalization using the smooth PBLH estimate.

3.1 Case selection

The presented technique aims to improve the temporal RS interpolation for cases of a CBL. The temperature and humidity profiles exhibit a unique structure with clear layers, which provides the potential for an interpolation, as they can remain structurally similar through appropriate compression or stretching.

The corresponding cases from April to July 2019 were selected by manually inspecting the potential temperature profiles for characteristics indicative of a CBL. As a result, this study focuses on the daytime period, with launches available from 14:30 to 23:30 UTC. The 11:30 UTC launch, which coincides with sunrise, does not exhibit a clear CBL structure and is therefore excluded.

This selection process resulted in 79 available comparisons between the reference and interpolated soundings.

3.2 Interpolation without a height normalization

The "normal" or "old" temporal interpolation between two RSs, as routinely performed by the ARM program (Jensen et al., 1998), is straightforward. To interpolate to a time of interest, the two closest soundings in time are vertically interpolated to a common fixed height grid with a resolution of $\Delta z = 20$ m in the lowest 3.5 km. Then, for each height level, a linear interpolation in time is performed for the atmospheric state variables (Fairless et al., 2021).

However, ARM's routine data product is not used in this study, as it incorporates all three-hourly separated soundings, leaving no intermediate soundings for comparison. In other words, the ARM program interpolates soundings that are three hours apart from each other. Still, this study needs a product based on the interpolation of soundings that are six hours apart from one another to use the additional soundings for the presented quantitative comparisons. Therefore, the normal interpolation is implemented manually on a fixed height grid with a spatial resolution similar to ARM's.

3.3 Interpolation with a height normalization

The key distinction of the "new" interpolation method lies in its operation on a normalized height grid $\hat{z}$, which takes into account the daily development of the CBL. Rosenberger et al. (2024) demonstrated the potential benefits of this method by showing that autocovariance analysis of high temporal resolution lidar data of vastly changing PBLHs z_i leads to a better agreement of the higher moments close to z_i with reality, if it is executed on a normalized height grid. This normalized height grid is the standard height grid z divided by an appropriate temporal average of the PBLH z_i, resulting in the transformation $\hat{z} = z\,\overline{z_i}^{-1}$. Consequently, a temporally smooth estimate of the PBLH $\overline{z_i}$ is required.

Thus, the normalization ensures that only equivalent atmospheric layers are interpolated, thereby preventing nonphysical gradients. For instance, the center of a mixed layer is interpolated by averaging the centers of the mixed layers at the times used for the interpolation. This principle applies to all height levels. Crucially, there is no nonphysical interpolation between an earlier interfacial layer and a later mixed layer at the same height level.

For the interpolation procedure, potential temperature and WVMR are calculated for the input soundings. The corresponding PBLHs are identified by automatically searching for the height between 100 and 2500 m where the potential temperature gradient peaks. This simple method can lead to false identifications, but the improvement of this is left for future work. Here, the main interest lies in the statistical investigation of many cases, and not in the quality of a single interpolation.

As the temporal interpolation will be performed on a normalized height grid, the identified PBLHs are used to normalize the corresponding profiles. The PBLH at the time of interest is extracted from the TROPoe time series data, which is smoothed using a 30 min moving average. When compared to high-resolution Raman lidar data (not shown), the retrieved estimate is not always accurate. To address this, a bias correction is applied, which linearly interpolates the bias between the PBLH estimate from TROPoe and the gradient peak of the RSs at the two launch times to the time of interest. The normalized "target" height grid at this time can then be defined by dividing a fixed height grid with a $\Delta z = 20$ m resolution by the corresponding smoothed and bias-corrected PBLH $\overline{z_i}$.

Next, the two normalized RS profiles are spatially linearly interpolated to the normalized height grid of the time of interest. Now, on the same normalized height level, the temporal interpolation is performed. The subsequent quantitative analysis
155 operates on the normalized height grid, as only this allows for meaningful height-related comparisons between different cases.

4 Results

First, the presented interpolation technique is qualitatively evaluated by comparing time-height cross-sections of WVMR for a day with a well-developed CBL. Both interpolation techniques, as well as the "true" state of the atmosphere represented by the temporally and spatially high-resolution measurements of a Raman lidar, are compared. This qualitative assessment is
160 followed by a quantitative analysis, in which for each interpolation technique, the same 79 interpolated profiles are compared to the corresponding reference profiles in terms of their correlation, their ratio of standard deviations, and their bias.

4.1 Example case

Figure 1 presents time-height cross-sections of WVMR m for 24 June 2019, between 14:29 and 23:33 UTC from the two interpolations (without an HN in (a), (c), and (e); with an HN in (b), (d), and (f)) and the Raman lidar (same cross-section in
165 (g) and (h)). The interpolations were performed between soundings (dashed lines) that launched six (first two rows) or nine (third row) hours apart from each other. Compared is a cross-section from the continuously measuring Raman lidar with 10 min average values. Consequently, the same temporal resolution was applied for the interpolation.

It is immediately evident that the new interpolation technique provides a more accurate representation of atmospheric conditions in the CBL, as confirmed by the Raman lidar. While the Raman lidar is not perfect, as the noise in the lower parts
170 of the mixed layer demonstrates, it continuously observes the overall structure (layering) of the CBL. This structure is much better represented by the interpolation that utilizes an HN, which holds for all three covered time intervals, even the ones that span nine hours. In contrast, the normal interpolation method fails to approximate the observed CBL structure even roughly, a deficiency that becomes particularly evident at the nine-hour intervals between soundings.

In this specific instance, the proposed interpolation technique demonstrates clear superiority over the method routinely
175 used by the ARM program. Consequently, the question arises whether this is an isolated case or a statistically significant improvement.

4.2 Statistics

The quantitative analysis presented here leverages the additional RS launches during the selected months to address this question. For this period, interpolations between launches that are six or more hours apart are performed, which is consistent with
180 the standard case at the SGP site (Holdridge, 2020). However, these launches are now interpolated to the time of the intermediate soundings, enabling a quantitative comparison. For example, the soundings at 17:30 and 23:30 are interpolated so that the time of the additional sounding at 20:30 UTC is covered.
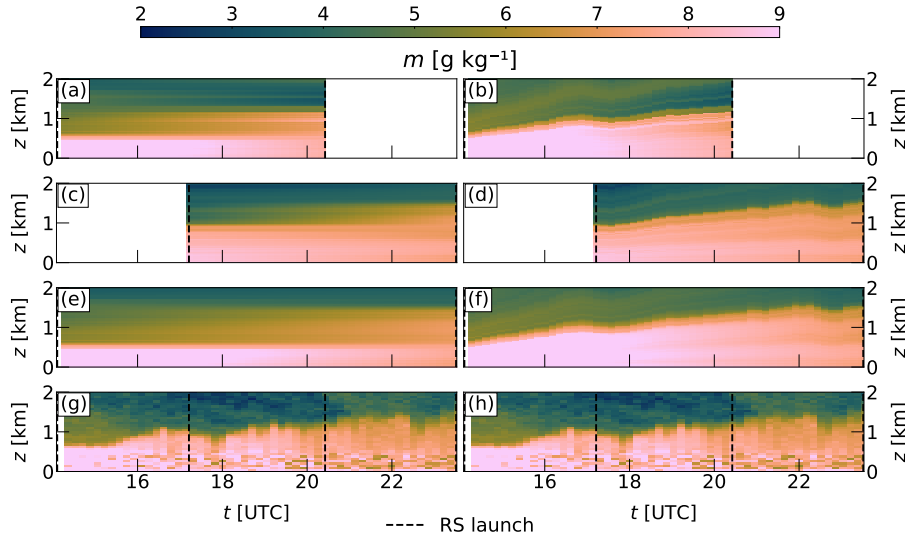
Figure 1. Time-height cross-sections of WVMR m for 24 June 2019, between 14:29 and 23:33 UTC are shown for the RS interpolation on a standard height grid ((a), (c), and (e)), on a normalized height grid ((b), (d), and (f)), and from a Raman lidar ((g) and (h)). The RS interpolations were performed between the following soundings (dashed lines): (a), (b) between 14:30 and 20:34, (c), (d) between 17:29 and 23:32, and (e), (f) between 14:30 and 23:32 UTC. The utilized color map comes from the Python package of Crameri et al. (2020).

By focusing exclusively on CBL cases and normalizing with the PBLH, the new interpolation technique is specifically designed to improve the structural representation within the daytime planetary boundary layer (PBL). As a result, this analysis focuses on height levels from the ground up to 1.2 times the PBLH, ensuring coverage of the interfacial layer. Both interpolated profiles, as well as the reference profile, are spatially interpolated onto a fixed normalized height grid with a spacing of $\Delta \hat{z} = 0.05$, resulting in 25 height levels that can be compared for each case. This requires transforming the interpolation without an HN and the reference profile to this normalized height grid, which is achieved using a manually validated PBLH estimate to guarantee a meaningful comparison between the different profiles.

4.2.1 Correlation and amplitude

To assess the performance of the two interpolation methods, the structure of the reference profile is first examined to determine how well it is approximated by the interpolations within the specified height interval. For this purpose, Pearson's correlation coefficient r and the ratio of the standard deviations of the interpolated and the reference profile $\sigma_{\mathrm{interp}}/\sigma_{\mathrm{ref}}$ are calculated for every selected comparison, respectively. r indicates how well the compared profiles vary together (Stull, 2017), or in other words, how well the interpolation can recreate the structure (here primarily the layering) of the actual profile. The ratio $\sigma_{\mathrm{interp}}/\sigma_{\mathrm{ref}}$ describes how well the amplitude of the structural variations is represented (Taylor, 2001). In the context of a CBL with a constant mixed layer and a sharp gradient at the top, a well-reproduced structure would imply that $\sigma_{\mathrm{interp}}/\sigma_{\mathrm{ref}}$ correctly captures the amplitude of the gradient if this ratio is equal to unity. However, to ensure that over- and underestimations of

the amplitude of the same absolute factor $2^{(-)x}$ with $x \in \mathbb{R}_0^+$ are weighted in the same way, the binary logarithm of this ratio

200    $\log_2\left(\sigma_{\text{interp}}/\sigma_{\text{ref}}\right)$ is considered because $\sigma_{\text{interp}} = 2^{(-)x}\sigma_{\text{ref}} \Leftrightarrow (-)x = \log_2(\sigma_{\text{interp}}/\sigma_{\text{ref}})$. Hereafter, this logarithmic ratio is also referred to as "the amplitude".

     It is essential to note that both r and $\sigma_{\text{interp}}/\sigma_{\text{ref}}$ represent the entire selected height profile, extending up to 1.2 times the PBLH. Figure 2 illustrates the distributions of these metrics for the 79 cases, distinguishing between interpolations with (red) and without (blue) an HN, for (a) potential temperature θ and (b) WVMR m. The figure also differentiates between cases

205   interpolated over six-hour intervals (dots) or nine-hour intervals (crosses).

     The distribution of cases reveals that the interpolation interval (six or nine hours) does not appear to have a significant impact, so that any potential differences related to this are not investigated further. It is evident that the interpolation with an HN leads on average to higher correlations, as most cases cluster closely around $r = 1$, whereas the interpolation without an HN exhibits a larger spread to lower correlation coefficients. This is observed for both variables, but is particularly pronounced

210   for potential temperature. Additionally, the interpolation without an HN appears to produce more cases with larger deviations in amplitude.

     Notably, there is a cloud of cases for potential temperature interpolated without an HN that have lower correlation coefficients between 0.6 and 0.9 and a systematic overestimation of the actual amplitude. These cases can be readily explained and highlight the limitations of the normal interpolation method. Upon closer examination of individual interpolations (not shown), it turns

215   out that these low correlations are often associated with the introduction of nonphysical atmospheric layers as artifacts of insufficient interpolation, which are then incorporated into the analyzed height levels. However, additional layers with a change in potential temperature increase the profile's variability compared to a single mixed layer capped by a gradient.

     The results presented in Fig. 2 are summarized in Table 1, which provides the median, mean, and standard deviation of the correlation and the amplitude for the profile comparisons of the two variables (potential temperature and WVMR) and

220   interpolation techniques (with and without an HN). To determine whether the interpolation with an HN yields significant improvements over the interpolation without an HN, a two-sided paired permutation test was conducted. This statistical test was chosen because it is well-suited for comparing two samples consisting of matched pairs drawn from unknown distributions. This test function is available in the widely used SciPy Python package (Haberland, 2023). The test statistic is based on the difference between the sample means, medians, or standard deviations. A total of 10000 permutations were performed. The

225   null hypothesis states that the data are randomly assigned to samples. If the null hypothesis is accepted, it implies that the interpolation techniques do not produce significantly different results. It is rejected when the probability of obtaining the observed result under the null hypothesis is less than 5 %. More information on SciPy can be found in Virtanen et al. (2020). A comprehensive introduction to permutation tests can be found in LaFleur and Greevy (2009). Values that show significant improvement over the other interpolation technique are bolded.

230   As already illustrated in Fig. 2, the improvement of the interpolation with an HN is particularly pronounced for potential temperature. Its median and mean values for the correlation and the profile amplitude are significantly improved. The median correlation increases from 0.850 to 0.963, and the median amplitude decreases from 0.401 to -0.004, bringing both values
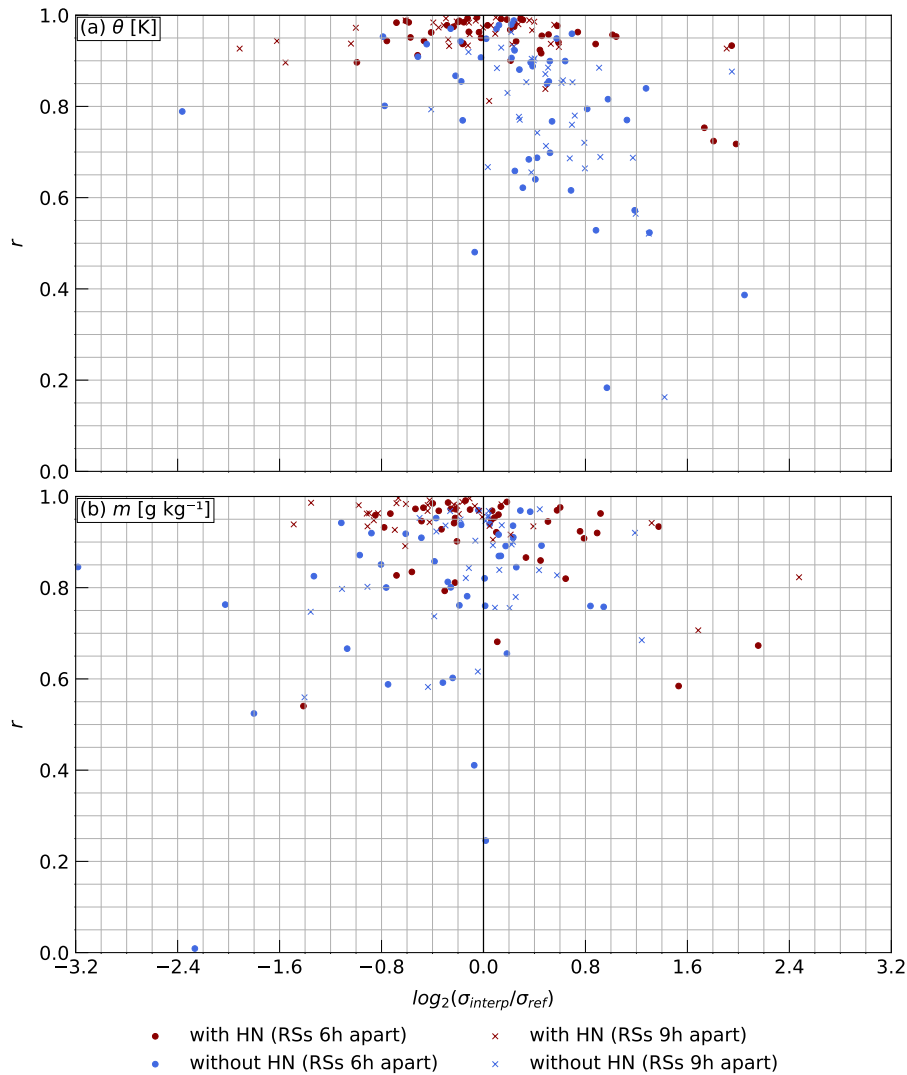
Figure 2. Distributions of the 79 profiles are shown for the interpolation with (red) and without (blue) an HN. Each interpolation is characterized by its correlation (r) to the reference profile and the binary logarithm of the normalized standard deviation ($\log_2\left(\sigma_{\text{interp}}/\sigma_{\text{ref}}\right)$) of the whole profile for (a) potential temperature θ and (b) WVMR m. Points represent interpolations where the used RSs were launched 6 h apart, while crosses mark interpolations where this time difference was 9 h. For better visualization, seven data points (five without and two with an HN) between $r = 0$ and $r = -1$ are not shown in (b).

much closer to the ideal values of $r = 1$ and $\log_2\left(\sigma_{\text{interp}}/\sigma_{\text{ref}}\right) = 0$. Furthermore, the standard deviation for the correlation is also improved, with a significantly smaller value of 0.055 compared to 0.167.

235   This picture is less clear for WVMR. Although the correlation is significantly improved, with a median of 0.954 compared to 0.839, the standard deviations are much higher and of comparable magnitude, at 0.312 and 0.352. The amplitude of the

Table 1. Comparison of the median (Median), mean (Mean), and standard deviation (StdDev) of Pearson's correlation coefficient (r) and the binary logarithm of the normalized standard deviation ($\log_2 (\sigma_{\text{interp}}/\sigma_{\text{ref}})$) between the interpolation on a normalized height grid and a standard height grid (with and without an HN respectively) are shown for potential temperature θ and WVMR m between 0.0 and 1.2 times the PBLH. This statistic is based on the 79 cases selected. Bolded values indicate a permutation-tested significant improvement of this metric for the corresponding interpolation technique over the other one.

Variable	Technique	r			$\log_2 (\sigma_{\text{interp}}/\sigma_{\text{ref}})$		
		Median	Mean	StdDev	Median	Mean	StdDev
Pot. Temp. θ [K]	with HN	**0.963**	**0.950**	**0.055**	**-0.004**	**0.061**	0.717
	without HN	0.850	0.787	0.167	0.401	0.417	0.612
WVMR m [g kg^{-1}]	with HN	**0.954**	**0.875**	0.312	-0.209	-0.098	0.774
	without HN	0.839	0.731	0.352	-0.064	-0.229	0.737

Table 2. Analog to Table 1 but for the height interval between 0.0 and 1.0 PBLH.

Variable	Technique	r			$\log_2 (\sigma_{\text{interp}}/\sigma_{\text{ref}})$		
		Median	Mean	StdDev	Median	Mean	StdDev
Pot. Temp. θ [K]	with HN	**0.828**	**0.770**	**0.184**	**-0.216**	**-0.219**	0.883
	without HN	0.566	0.528	0.302	1.286	1.165	0.902
WVMR m [g kg^{-1}]	with HN	**0.876**	**0.793**	0.326	**-0.204**	**-0.238**	0.870
	without HN	0.780	0.654	0.357	0.461	0.321	0.872

interpolated WVMR profiles is not improved (but also not worsened), as the median value of the interpolation without an HN is closer to the ideal value of 0, while for the mean, it is the value of the interpolation with an HN that is closer. The standard deviations are comparable, at around 0.75.

240   Overall, the new interpolation represents a clear improvement in approximating the structure and amplitude of potential temperature profiles in the convective PBL. The structure of WVMR profiles is also better represented with the new technique, although the picture is less clear for its amplitude. It is worth noting that WVMR profiles are generally more challenging to interpolate accurately, as moisture is significantly affected by different sources at the surface, advection, and cloud formation, exhibiting more visible variations in individual profiles (Stull, 1988). Please note that the new interpolation technique is

245   proposed to improve profiles within the convective PBL. Table 2 therefore shows the same statistics for the correlation and amplitude representation for the height interval between 0.0 and 1.0 PBLH.

Although the absolute values for correlation and amplitude seem degraded, as they lie for both techniques farther away from the ideal values of $r = 1$ and $\log_2 (\sigma_{\text{interp}}/\sigma_{\text{ref}}) = 0$, the significant improvements of the new method compared to the old one

stay the same. Now, even the median and mean values of the amplitude representation for WVMR are significantly improved
from 0.461 to -0.204 in the case of the median. The reduced absolute values are no surprise, as this adjusted height interval
does not encompass the (full) gradient in potential temperature or WVMR, and with that has much less structure that can be
captured and compared.

4.2.2 Bias

After determining how well the two interpolation techniques perform in terms of their ability to recreate the reference profile's
structure and amplitude, the bias must be quantified. This is necessary because the previously discussed metrics may indicate a
more accurate representation of the shape of the reference profile. However, they are blind to the possibility that the interpolated
profile may result in a completely different temperature or humidity level.

The following analysis of the bias is conducted in a height-resolved manner, meaning that for each of the 25 available height
levels, the bias is calculated for all 79 cases. Figure 3 presents the height profiles of the mean bias (first column) and its standard
deviation (second column) for the interpolation with (red) and without (blue) an HN for potential temperature θ (first row) and
WVMR m (second row).

Upon examining the results, it is readily apparent that neither interpolation technique is perfect, as both exhibit a non-zero
bias. However, it is also clear that the interpolation with an HN displays a relatively constant bias and standard deviation
profile throughout the investigated height interval, facilitating a more meaningful correction if the corresponding bias can be
estimated. This is a further argument for the good structural representation of the reference profile. The interpolation without
an HN does not achieve this.

Returning to the actual meaning of the bias, it is visible that the new interpolation exhibits a mean cold bias of approximately
1 K and a mean wet bias of approximately 0.3 g kg^{-1}, with corresponding standard deviations of about 1 K and 1 g kg^{-1} in
the mixed layer. This bias arises from underestimating potential temperature and overestimating WVMR. The presence of a
bias is not surprising, given that the temporal interpolation is still linear and thus unable to capture mesoscale, or generally
non-linear, changes in overall temperature and humidity levels that occur over time due to the sun's varying energy input. A
concrete example is the sounding at 14:30 UTC in the morning with generally cooler conditions than the reference conditions
later in the day. Cases that were interpolated using this sounding have an especially pronounced cold bias (not shown).

The mean bias of the interpolation without an HN changes roughly linearly with height, reaching a negative (potential
temperature) or positive (WVMR) value at the PBLH of comparable or even greater magnitude than the new interpolation. In
other words, the two bias profiles diverge increasingly with height. The standard deviation also varies more and takes on larger
values in the upper part of the mixed layer, between 0.6 and 0.95 PBLH. This limits a general scientific application of the old
interpolation technique.

Focusing on potential temperature, this behavior can be understood by considering the mentioned cold bias and the fact
that the interpolation technique does not use an HN. This causes nonphysical artifact layers in the resulting profile. For the
height levels close to the surface, it is evident that the choice of the technique has a minimal impact, as corresponding physical
regimes are averaged. For instance, parts of the mixed layer from the earlier and later soundings are averaged, resulting in a
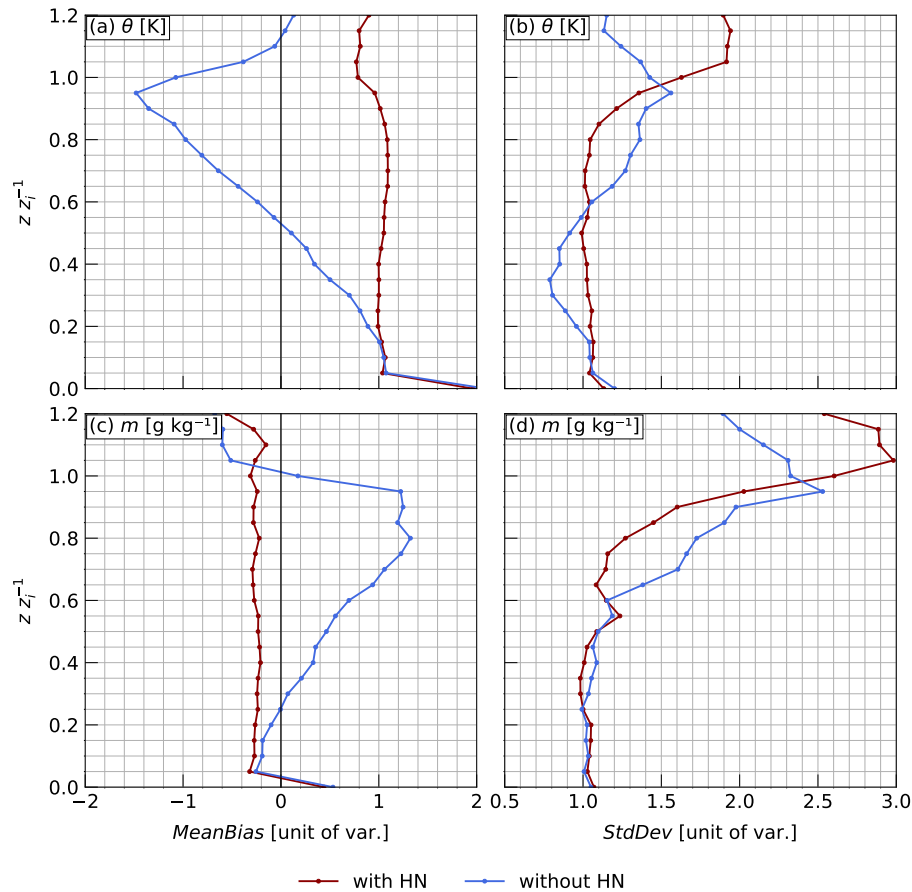
Figure 3. Profiles of (a), (c) the mean bias $MeanBias$, calculated as the reference value minus the interpolated value, and (b), (d) its standard deviation $StdDev$ are shown for different normalized height levels between 0.0 and 1.2 times the PBLH with $z_i =$PBLH for (a), (b) potential temperature θ and (c), (d) WVMR m with (red) or without (blue) an HN. Bias and standard deviation have the same units as the corresponding variable.

mixed layer that is likely to be present at the time of interest. The new interpolation technique can be superior, as it consistently averages corresponding fractions of the boundary layer, but this effect is small for a mixed layer with a nearly constant potential
285 temperature profile.

However, as the height increases to around 0.2 PBLH, there will be the first cases for which the interpolation without an HN averages different physical regimes, as the earlier sounding has a shallower PBLH (Stull, 1988). In this study, this comes down to the soundings at 14:30 UTC, which introduce a pronounced cold bias into the resulting profile. At these height levels, the interpolation without an HN averages the FA of the earlier sounding with the mixed layer of the latter. Since the FA is warmer
290 than the underlying mixed layer (Stull, 1988), the potential temperature of the introduced artifact layer can be closer to the

actual temperature or exhibit a warm bias resulting in a statistically smaller mean bias and standard deviation for this height compared to the interpolation that utilizes an HN.

This explanation is supported by the fact that the divergence of the bias profiles begins at higher height levels, and the standard deviation remains roughly constant when cases using the sounding at 14:30 UTC are excluded (not shown). Going up
295 in height, more and more cases of the non-HN interpolation introduce an artificial layer, shifting the mean bias towards larger and larger warm biases and increasing its standard deviation. At a height level around the PBLH, approximately 0.95 times the PBLH, the absolute mean bias decreases again, as the reference profile has reached the FA.

5 Conclusions

This work introduced a new and simple technique for temporally interpolating RSs in the atmospheric boundary layer, specif-
300 ically in the CBL. The new method operates on a normalized height grid, unlike the standard height grid commonly used by institutions such as the ARM program. Normalization is achieved using the PBLH, which enables the technique to average only physically related height levels during interpolation, thereby preventing the formation of unphysical artifact layers.

The required PBLH estimates were derived from the processing of radiance interferometer data, although they could have been obtained from any suitable source. A bias correction was applied to ensure that these estimates were consistent with the
305 PBLH values derived from the radiosondes, which was necessary to achieve the improvements presented in this study.

The improvements of the new technique were quantified by comparing interpolated profiles of potential temperature and WVMR with actual soundings as well as with the routine interpolation method that does not employ a height normalization. The comparison examined the correlation, amplitude representation, and bias between the two interpolation techniques and the reference sounding.
310 For the potential temperature, all metrics indicate a significant improvement of the new technique over the old one. The median correlation of 0.963 and the amplitude of -0.004 are closer to the ideal values of 1 and 0, respectively, than the interpolation without height normalization. The bias remains approximately constant at 1 K throughout the boundary layer height interval, whereas the old interpolation method's bias varies with height, suggesting a poor structural representation that manifests itself as biases that appear statistically improved at certain height levels. Similar significant improvements in
315 correlation and bias are observed for WVMR, although the representation of amplitude is only improved when changing the top of the investigated height interval to 1.0 times the PBLH.

In the future, it is important to investigate the performance of the new technique in various atmospheric situations beyond the CBL. Notably, time-wise structurally constant settings, such as a nighttime boundary layer, appear to be particularly promising. Additionally, a future focus should lie on exploring possibilities to simultaneously get the best results for the CBL, as well as
320 the FA.

In conclusion, the initial question and objective of whether temporally interpolating on a normalized height grid yields an improved representation of reality in the lower atmosphere compared to interpolating on a standard height grid can be answered

affirmatively. The presented method is recommended over the currently used one when feasible, as it benefits research and more sophisticated data products.

325 *Data availability.* Data were obtained from the Atmospheric Radiation Measurement (ARM) user facility, a U.S. Department of Energy (DOE) Office of Science user facility managed by the Biological and Environmental Research program.

Author contributions. DDT conceived of this project with input from LvK and VW. LvK developed the utilized code and performed the presented tests and comparisons with continuous support by DDT and DL. LvK prepared the manuscript, which was improved by all co-authors.

330 *Competing interests.* The authors declare that they have no conflict of interest.

Acknowledgements. The authors express their gratitude to Kirsten Warrach-Sagi and Andreas Behrendt from the Institute of Physics and Meteorology. Kirsten helped throughout this work with valuable input, while Andreas gave much-appreciated feedback for the finalization.

References

Behrendt, A., Wulfmeyer, V., Senff, C., Muppa, S. K., Späth, F., Lange, D., Kalthoff, N., and Wieser, A.: Observation of sensible and latent heat flux profiles with lidar, Atmos. Meas. Tech., 13, 3221–3233, https://doi.org/10.5194/amt-13-3221-2020, 2020.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nat. Commun., 11, 5444, https://doi.org/10.1038/s41467-020-19160-7, 2020.

Duncan Jr., J. B., Bianco, L., Adler, B., Bell, T., Djalalova, I. V., Riihimaki, L., Sedlar, J., Smith, E. N., Turner, D. D., Wagner, T. J., and Wilczak, J. M.: Evaluating convective planetary boundary layer height estimations resolved by both active and passive remote sensing instruments during the CHEESEHEAD19 field campaign, Atmos. Meas. Tech., 15, 2479–2502, https://doi.org/10.5194/amt-15-2479-2022, 2022.

Dunn, M., Johnson, K., and Jensen, M.: The Microbase Value-Added Product: A Baseline Retrieval of Cloud Microphysical Properties, US Department of Energy, Office of Science, Office of Biological and Environmental Research, https://doi.org/10.2172/1015189, 2011.

Fairless, T., Jensen, M., Zhou, A., and Giangrande, S.: Interpolated Sounding and Gridded Sounding Value-Added Products, U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington, https://doi.org/10.2172/1248938, 2021.

Haberland, M.: Fast Resampling and Monte Carlo Methods in Python, Journal of Open Source Software, 8, 5092, https://doi.org/10.21105/joss.05092, 2023.

Holdridge, D.: Balloon-Borne Sounding System (SONDE) Instrument Handbook, US Department of Energy, Office of Science, Office of Biological and Environmental Research, https://doi.org/10.2172/1020712, 2020.

Jensen, M., Giangrande, S., Fairless, T., and Zhou, A.: interpolatedsonde, Atmospheric Radiation Measurement (ARM) Archive, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (US) and ARM Data Center, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), https://doi.org/10.5439/1095316, 1998.

Jensen, M. P., Holdridge, D. J., Survo, P., Lehtinen, R., Baxter, S., Toto, T., and Johnson, K. L.: Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains site, Atmos. Meas. Tech., 9, 3115–3129, https://doi.org/10.5194/amt-9-3115-2016, 2016.

Keeler, E., Burk, K., and Kyrouac, J.: Balloon-borne sounding system (BBSS), WNPN output data, Atmospheric Radiation Measurement (ARM) Archive, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (US) and ARM Data Center, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), https://doi.org/10.5439/1595321, 2022.

Knuteson, R. O., Revercomb, H. E., Best, F. A., Ciganovich, N. C., Dedecker, R. G., Dirkx, T. P., Ellington, S. C., Feltz, W. F., Garcia, R. K., Howell, H. B., Smith, W. L., Short, J. F., and Tobin, D. C.: Atmospheric Emitted Radiance Interferometer. Part II: Instrument Performance, J. Atmos. Ocean. Tech., 21, 1777–1789, https://doi.org/10.1175/JTECH-1663.1, 2004.

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmos. Meas. Tech., 16, 433–479, https://doi.org/10.5194/amt-16-433-2023, 2023.

Krishnamurthy, R., Newsom, R., Chand, D., and Shaw, W.: Boundary Layer Climatology at ARM Southern Great Plains, Pacific Northwest National Lab. (PNNL), Richland, WA (United States), https://doi.org/10.2172/1778833, 2021a.

Krishnamurthy, R., Newsom, R. K., Berg, L. K., Xiao, H., Ma, P.-L., and Turner, D. D.: On the estimation of boundary layer heights: a machine learning approach, Atmos. Meas. Tech., 14, 4403–4424, https://doi.org/10.5194/amt-14-4403-2021, 2021b.

LaFleur, B. J. and Greevy, R. A.: Introduction to permutation and resampling-based hypothesis tests, J. Clin. Child. Adolesc., 38, 286–294,
370   https://doi.org/10.1080/15374410902740411, 2009.

Newsom, R., Sivaraman, C., and Zhang, D.: Water vapor mixing ratio from Raman Lidar, Atmospheric Radiation Measurement (ARM)
    Archive, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (US) and ARM Data Center, Oak Ridge National Laboratory (ORNL),
    Oak Ridge, TN (United States), https://doi.org/10.5439/1415137, 2014.

Newsom, R., Bambha, R., and Chand, D.: Raman Lidar (RL) Instrument Handbook, U.S. Department of Energy, Atmospheric Radiation
375   Measurement user facility, Richland, Washington, https://doi.org/10.2172/1020561, 2022.

Rosenberger, T. E., Turner, D. D., Heus, T., Raghunathan, G. N., Wagner, T. J., and Simonson, J.: Improving the estimate of
    higher-order moments from lidar observations near the top of the convective boundary layer, Atmos. Meas. Tech., 17, 6595–6602,
    https://doi.org/10.5194/amt-17-6595-2024, 2024.

Shupe, M. D., Comstock, J. M., Turner, D. D., and Mace, G. G.: Cloud Property Retrievals in the ARM Program, Meteor. Mon., 57, 19.1–
380   19.20, https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0030.1, 2016.

Sisterson, D. L., Peppler, R. A., Cress, T. S., Lamb, P. J., and Turner, D. D.: The ARM Southern Great Plains (SGP) Site, Meteor. Mon., 57,
    6.1–6.14, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1, 2016.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-009-3027-8,
    1988.

385 Stull, R. B.: Practical Meteorology: An Algebra-Based Survey of Atmospheric Science; 1.02 b, The University of British Columbia, Van-
    couver, Canada, ISBN 978-0-88865-283-6, https://www.eoas.ubc.ca/books/Practical_Meteorology/, 2017.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res.- Atmos., 106, 7183–7192,
    https://doi.org/10.1029/2000JD900719, 2001.

Troyan, D.: Merged Sounding Value-Added Product, US Department of Energy, Office of Science, Office of Biological and Environmental
390   Research, https://doi.org/10.2172/973815, 2010.

Turner, D.: tropoe (c1), Atmospheric Radiation Measurement (ARM) Archive, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN
    (US) and ARM Data Center, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), https://doi.org/10.5439/1996977,
    2021.

Turner, D. D. and Goldsmith, J. E. M.: Twenty-Four-Hour Raman Lidar Water Vapor Measurements during the Atmospheric Radia-
395   tion Measurement Program's 1996 and 1997 Water Vapor Intensive Observation Periods, J. Atmos. Ocean. Tech., 16, 1062–1076,
    https://doi.org/10.1175/1520-0426(1999)016<1062:TFHRLW>2.0.CO;2, 1999.

Turner, D. D. and Löhnert, U.: Information Content and Uncertainties in Thermodynamic Profiles and Liquid Cloud Properties Re-
    trieved from the Ground-Based Atmospheric Emitted Radiance Interferometer (AERI), J. Appl. Meteorol. Clim., 53, 752–771,
    https://doi.org/10.1175/JAMC-D-13-0126.1, 2014.

400 Turner, D. D., Goldsmith, J. E. M., and Ferrare, R. A.: Development and Applications of the ARM Raman Lidar, Meteor. Mon., 57, 18.1–
    18.15, https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0026.1, 2016a.

Turner, D. D., Mlawer, E. J., and Revercomb, H. E.: Water Vapor Observations in the ARM Program, Meteor. Mon., 57, 13.1–13.18,
    https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0025.1, 2016b.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J.,
405   van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat,
    İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald,

A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Weckwerth, T., Spuler, S., and Turner, D.: Micropulse Differential Absorption Lidar (MPD) Network Demonstration Field Campaign Report,
410   US Department of Energy, Office of Science, Office of Biological and Environmental Research, https://doi.org/10.2172/1595264, 2020.

Wulfmeyer, V., Hardesty, R. M., Turner, D. D., Behrendt, A., Cadeddu, M. P., Di Girolamo, P., Schlüssel, P., van Baelen, J., and Zus, F.: A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles, Rev. Geophys., 53, 819–895, https://doi.org/10.1002/2014RG000476, 2015.

Wulfmeyer, V., Muppa, S. K., Behrendt, A., Hammann, E., Späth, F., Sorbjan, Z., Turner, D. D., and Hardesty, R. M.: Determination of
415   Convective Boundary Layer Entrainment Fluxes, Dissipation Rates, and the Molecular Destruction of Variances: Theoretical Description and a Strategy for Its Confirmation with a Novel Lidar System Synergy, J. Atmos. Sci., 73, 667–692, https://doi.org/10.1175/JAS-D-14-0392.1, 2016.