Noname manuscript No.
(will be inserted by the editor)

Evaluation and Applications of Multi-Instrument Boundary-Layer Thermodynamic Retrievals

Elizabeth N. Smith[1] · Brian R. Greene[2,3] · Tyler M. Bell[2,4,1,3] · William G. Blumberg[5] · Ryann Wakefield[2] · Dylan Reif[2] · Qing Niu[2,4] · Qingyu Wang[2] · David D. Turner[6]

Received: DD Month YEAR / Accepted: DD Month YEAR

Abstract Recent reports have highlighted the need for improved observations of the boundary layer. In this study, we explore the combination of ground-based active and passive remote sensors deployed for thermodynamic profiling to analyze various boundary-layer observation strategies. Optimal-estimation retrievals of thermodynamic profiles from Atmospheric Emitted Radiance Interferometer (AERI) observed spectral radiance are compared with and without the addition of active sensor observations from a May–June 2017 observation period at the Atmospheric Radiation Measurement–Southern Great Plains Site. In all, three separate thermodynamic retrievals are considered here: retrievals including AERI data only, retrievals including AERI data and Vaisala water vapour differential absorption lidar data, and retrievals including AERI data and Raman lidar data. First, the three retrievals are compared to each other and to reference radiosonde data over the full observation period to get a bulk understanding of their differences and characterize the impact of clouds on these retrieved profiles. These analyses show that the most significant differences are in the water vapour field, where the active sensors are better able to represent the moisture gradient in the entrainment zone near boundary layer top. We also explore how differences in retrievals may impact results of applied analyses including land–atmosphere coupling, convection indices, and severe storm environmental characterization. Overall, adding active sensors to the optimal-estimation retrieval showed some added information,

[1] NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma ·
[2] School of Meteorology, University of Oklahoma, Norman, Oklahoma ·
[3] Center for Autonomous Sensing and Sampling, University of Oklahoma, Norman, Oklahoma ·
[4] Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma ·
[5] NASA Postdoctoral Program/NASA Goddard Space Flight Center, Greenbelt, Maryland ·
[6] NOAA/OAR/Global Systems Laboratory, Boulder, Colorado ·

particularly in the moisture field. Given the costs of such platforms, the value of that added information must be weighed for the application at hand.

Keywords Boundary-layer observation · Remote sensing · Thermodynamic retrievals

1 Introduction

Widely deployed operational observation networks in the United States routinely monitor near-surface conditions (e.g., automated surface observing system, or ASOS, networks and mesonets) and conditions a kilometre above the surface and further aloft (e.g., weather radar and satellite observations). In the intervening layer from the surface to a few kilometres above it—in other words, the boundary layer—routine observations are few and far between. One common observation dataset collected in this portion of the atmosphere comes from balloon-borne packages, or radiosondes. However, operational radiosonde stations are located 500-km apart and only launched twice a day (Melnikov et al. 2011). Another dataset is the aircraft meteorological data relay, or AMDAR, data, however these are primarily temperature and wind with few water vapour observations—only about 10% of aircraft have water vapour observation capabilities (Moninger et al. 2010; Zhang et al. 2019). These profiles are not collected at all airports, and are 'flights of opportunity', resulting in poor diurnal sampling even at these airports. An obvious gap exists.

In recent decades, the need for improved observations of the boundary layer to serve the growing needs of society has become apparent. Over the past ten years, the National Research Council has published multiple National Academies of Science reports that partially attribute limits of current knowledge of lower-atmospheric phenomena to limitations in observing capabilities and call for improved observations of temperature, humidity, wind, and cloud characteristics in and near the boundary layer. In particular, these reports call for a new ground-based network of these boundary-layer observations (National Research Council 2009, 2010). Wulfmeyer et al. (2015) made similar recommendations in a review of remote sensing of the lower-troposphere. More recently, the 2017–2027 Decadal Survey (National Academies of Sciences, Engineering, and Medicine 2018) has instigated interest in possible space-based solutions for observing the planetary boundary layer (PBL). Although such a solution currently has many physical and financial limitations (e.g., cloud cover, large satellite footprints, low signal-to-noise, expense), a space-based observing system would be informed and complemented by ground-based assets. The wide variety of solutions being pursued suggests that thorough knowledge of instrument synergy will be necessary to consider when investing in any PBL-oriented observing systems.

The literature suggests that platforms combining thermodynamic and wind observation capabilities may be most useful for many applications (e.g., Hartung et al. 2011; Otkin et al. 2011). Several such platforms have been operating in the U.S. for many years. These include fixed-site observatories such as the

Table 1 Mobile multi-instrument boundary layer profiling systems.

Platform	**Citation**
National Center for Atmospheric Research Integrated Sounding System NCAR ISS	Parsons et al. (1994)
University of Alabama in Huntsville Mobile Integrated Profiling System and Mobile Doppler Lidar System MIPS; MoDLS	Karan and Knupp (2006) Knupp et al. (2009) Wingo and Knupp (2015)
University of Wisconsin SSEC Portable Atmospheric Research Center SPARC	Wagner et al. (2019)
University of Oklahoma/NOAA – NSSL Collaborative Lower Atmospheric Mobile Profiling Systems CLAMPS-1/CLAMPS-2	Wagner et al. (2019)

Department of Energy Atmospheric Radiation Measurement Southern Great Plains site (Sisterson et al. 2016), as well as mobile platforms such as those listed in Table 1. Such boundary-layer profiling systems provide observations of thermodynamic and kinematic variables every few minutes that can inform understanding of boundary-layer structure, convection initiation, severe storm environments, and land–atmosphere interactions, which are all sensitive to profiles of wind, temperature, and moisture in the boundary layer and above. For example, standard CLAMPS (see Table 1) operating modes provide 5-minute resolution for temperature and moisture observations and 2-minute resolution for wind observations.

In recent years, weather sensing uncrewed aircraft systems (WxUAS) have emerged as a potential observation platform to study the boundary layer (e.g., Koch et al. 2018; Kral et al. 2020; de Boer et al. 2020; Segales et al. 2020). WxUAS have been shown to perform just as well or better than ground-based remote sensors in some scenarios, though improvements could still be made (Bell et al. 2020). Regulatory challenges hinder incorporation of autonomous systems into the National Airspace System. Ground-based profilers can provide long-term continuous observations both in harsh or remote environments (where maintenance of the WxUAS could be difficult) and highly populated areas (where operating WxUAS over people poses liability). Additionally, most ground-based profilers can already operate autonomously, which is at this point only a burgeoning capability of WxUAS in the United States. Even as WxUAS technology development continues, an effective solution to filling the boundary-layer data gap likely includes both WxUAS and ground-based profiling platforms.

In pursuing an observation framework upon which a national network could be designed, it is important to consider how various instruments may be able to work synergistically to maximize benefits while minimizing cost. This strategy was first explored in an observation simulation system experiment (OSSE) framework (Löhnert et al. 2009; Otkin et al. 2011; Hartung et al. 2011). Recently, improvements to convective-scale forecasts have been found from assimilating small network-style deployments of ground-based thermodynamic and kinematic profilers into mesoscale numerical weather prediction models (Degelia et al. 2019; Hu et al. 2019; Coniglio et al. 2019). Given an emerging market of active remote sensors to perform thermodynamic profiling, an important avenue to explore is multi-instrument retrievals. Variational-based physical retrievals, such as the AERIoe algorithm (Turner and Löhnert 2014), can integrate a variety of instruments with various strengths and weaknesses to produce a better retrieved atmospheric profile than one instrument alone (Turner and Blumberg 2019).

In this work, we explore the combination of active and passive remote sensors deployed for thermodynamic profiling with the intent of adding to a growing body of scientific literature analyzing various boundary-layer observation strategies. By adding active remote sensor observations into a framework commonly applied for passive profiling, we aim to understand how the resulting profiles change and what impacts those changes have. To explore these impacts, we conduct a variety of scientific analyses using these retrieved data to determine if changed profiles change results.

2 Data

This study utilizes data collected during an evaluation experiment at the Department of Energy's Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP), which is a long-operational site located in north-central Oklahoma, surrounded by mainly flat open pasture and rangeland (Sisterson et al. 2016; 36.605^{o}N, 97.486^{o}W). In addition to instrumentation typically located at the site, a differential absorption lidar (DIAL) was deployed for evaluation against ARM-SGP instrumentation from 15 May to 12 June 2017. A summary of this evaluation effort can be found in Newsom et al. (2020). Here, we evaluate the utility of each thermodynamic profiler used in a combined manner to produce more confident atmospheric profile retrievals.

2.1 Atmospheric Emitted Radiance Interferometer

The Atmospheric Emitted Radiance Interferometer (AERI) is a passive remote sensor similar to a microwave radiometer, except it observes downwelling radiation in the mid-infrared portion of the spectrum that is sensitive to the vertical thermodynamic structure of the atmosphere. The AERI measures downwelling infrared radiation every 20 s from 3.3 to 19 μm in wavelength (Knuteson et al. 2004). After applying a noise filter (Turner et al. 2006) and averaging the radiances to 2-min intervals, the spectral radiances are processed through an optimal-estimation-based retrieval algorithm, discussed below. Blumberg et al. (2017a) showed that the high temporal resolution of the system is useful in detecting rapid changes in stability (in Great Plains environments).

2.2 Raman Lidar

Since 1996, an automated Raman lidar (RLID) has been operated by the ARM program profiling atmospheric water vapour, aerosols, and clouds (Turner et al. 2016). RLID is an active remote sensor which transmits a 300 mJ pulse of laser energy (355 nm) vertically, and detects backscatter at the transmitted wavelength and at wavelengths associated with Raman scattering from water vapour (408 nm) and nitrogen (387 nm). Profiles of backscatter are collected with 7.5 m vertical resolution every 10 seconds (Goldsmith et al. 1998; Newsom et al. 2009). After some quality assurance measures are applied, the ratio of the water vapour to nitrogen signals is computed, which is expected to be proportional to the water vapour mixing ratio (Turner and Goldsmith 1999). This relationship and some calibration steps (employing collocated radiosonde launches) are used to produce value-added products containing atmospheric thermodynamic profiles. For the analyses herein, RLID data with 10 min temporal and 75 m vertical resolution are used[1].

2.3 Water Vapour Differential Absorption Lidar

Ground-based DIAL water vapour observations were made as early as the 1990s (Wulfmeyer 1999). In recent years, turnkey DIAL platforms have been developed for the purpose of ground-based profiling of boundary-layer thermodynamics, specifically water vapour. This development work has included efforts by Montana State University and the National Center of Atmospheric Research (Spuler et al. 2016; Weckwerth et al. 2016), Tokyo Metropolitan University (Le Hoai et al. 2016), and Vaisala (Roininen and Münkel 2016; Newsom et al. 2020). DIAL instruments provide measurements of the vertical profile of a trace gas concentration by transmitting two or more wavelengths of laser energy. Changes in molecular absorption at these different wavelengths (due to the spectroscopic properties of the gas) result in differences in attenuation at different laser frequencies. These laser wavelengths are typically chosen to be very near each other spectrally, so that other possible atmospheric properties that could lead to differences in the observed attenuated backscatter signal (e.g., aerosol optical properties) are assumed to be similar enough that they can be ignored. Narrowband DIAL systems wherein the output laser energy is monochromatic at each of the desired wavelengths, such as the Spuler et al. (2016) systems, are able to directly provide calibrated profiles of that trace gas (e.g., water vapour) without the need for external calibration. However, broadband DIAL systems, such as the system built by Vaisala and described in Newsom et al. (2020), transmit the laser energy over a finite spectral range at each 'characteristic' frequency and thus require an external calibration source. This study will use the Vaisala water vapour DIAL (hereafter wvDIAL), which

[1] While the RLID can provide partial profiles of temperature (Newsom et al. 2013) those data are not used in the AERIoe retrieval described in Sect. 2.4, since the comparison was relative to wvDIAL.

was calibrated using the built-in in situ humidity sensor at the surface (Newsom et al. 2020).

2.4 AERIoe

As mentioned above, spectral radiances observed by the AERI (after noise filtering, see Turner et al. 2006, and 2-min averaging) are processed through an optimal-estimation-based retrieval algorithm called AERIoe, which is described in Turner and Löhnert (2014). AERIoe obtains estimates of the vertical profile of temperature, T, and water vapour mixing ratio, $WVMR$, as well as the cloud liquid water path and mean cloud effective radius in the column. The retrieval is constrained in the middle to upper troposphere by a first guess based on climatological mean conditions for the region derived from radiosonde archived data, but the final retrievals are thought to be insensitive to the particular first-guess profile that is used (Turner and Löhnert 2014). Alone, AERI spectra processed through AERIoe produce retrieved profiles that lose vertical resolution rapidly with height and contain far fewer independent pieces of information than what can be obtained from in situ methods such as radiosondes (see Turner and Löhnert 2014 Fig. 7d, f). However, the information content in the AERI observations, which may have 4–8 independent pieces of information depending on the environment, is much higher than for other platforms such as microwave radiometers with only 2–4 independent pieces of information (Löhnert et al. 2009; Turner and Löhnert 2014; Blumberg et al. 2015).

The retrieval itself is an ill-posed problem; many different thermodynamic solutions can produce the radiance observations that were measured. Recent improvements to AERIoe have allowed more types of observations to be provided as input to the retrieval, as long as there is a forward model that can convert between the state space—which describes the atmospheric state—and observation space–which is what the platform observes; in the case of the AERI, spectral radiances (Turner and Blumberg 2019; Turner and Löhnert 2020). In essence, this forces the retrieval to find a solution that not only agrees with the radiance observations, but is also within the uncertainty of the additional observations. Due to the rapid drop off of independent data points in the middle troposphere when only using AERI spectra in the retrieval, NOAA Rapid Refresh model analysis (Benjamin et al. 2016) profiles are used to constrain the retrieval above 4 km, given their hourly availability. Other numerical model output could be used here since generally we expect—due to modern data assimilation methods and less horizontal variability—reasonable accuracy in mid-troposphere model analyses. Since ground-based sensors have little sensitivity above the boundary layer, we rely on these analyses to improve the quality of the retrieved profile for integrated or otherwise profile-estimated quantities. Additionally, in situ surface observations are used to constrain the near surface part of the retrieval. When available, microwave radiometer brightness temperatures or other remote sensor observations can be included in the retrieval. This capability to include additional remote sensors is leveraged in this study.

Here we include additional observations from active thermodynamic remote sensors, or more specifically lidar water vapour profilers, as constraints in the AERIoe retrieval in order to evaluate changes in retrieved thermodynamic profiles and the resulting impacts on subsequent products and analyses. Improving the accuracy of the retrieved water vapour profile by adding lidar water vapour profiles as input into the retrieval algorithm allows the algorithm to use the temperature sensitivity of the water vapor bands to improve the temperature profile (Turner and Löhnert 2020). Observations were processed through the AERIoe algorithm for the entire period with AERI data only (hereafter noted as AERIonly), AERI data constrained by RLID observations (hereafter noted as AERIrLID), and AERI data constrained by wvDIAL (hereafter noted as AERIvDIAL). In each instance, the AERIoe retrieval was performed with the same settings. These retrievals have 5-min resolution, use a prior estimate based on a 30-year climatology of radiosondes released from the SGP site, and include NOAA Rapid Refresh temperature and humidity profiles as a constraint from 4–10km. The retrievals are also constrained by including nearby microwave radiometer brightness temperature observations, surface meteorology observations, and observed cloud base heights. The improvements to retrievals that include such data in the observation vector are detailed in Turner and Blumberg (2019).

2.5 Radiosondes

Balloon-borne radiosondes have been launched from the ARM-SGP site since 1992, providing in situ measurements along vertical profiles of both the thermodynamic state of the atmosphere, and the wind speed and direction. At present, radiosondes are typically launched from this location four times daily valid at 0600, 1200, 1800, and 0000 UTC with occasional additional releases during intensive field campaigns. During the period of interest for this work, 109 Vaisala RS41 model radiosondes were launched at the SGP site between 0532 UTC on 16 May 2017 and 1726 UTC on 12 June 2017. Assuming a nominal 5 m s^{-1} ascent rate of the balloon and noting that the radiosonde takes a measurement every 2 s, data should have a vertical resolution of approximately 10 m. $WVMR$ values are calculated from dew point temperature and pressure reported by the post-processed radiosonde observations using the empirical approximation for saturation vapour pressure in Bolton (1980). Radiosonde temperature and $WVMR$ in the range of 0–4 km a.g.l. are then linearly interpolated in the vertical to match the same altitude bins as the AERIoe retrievals. To ensure direct comparisons, the time stamp of each radiosonde altitude bin is iteratively matched with the nearest AERIoe profile time stamp. This is necessary as the post-processed AERIoe profiles are effectively instantaneous with 5 min time resolution, whereas the radiosonde can take anywhere from 10–15 min to traverse the same altitudes observed by the ground-based remote sensors.

3 Bulk Analysis

In order to understand what impacts the inclusion of active sensors may have on the retrieved thermodynamic profiles, we present a few sets of analyses considering the full 15 May to 12 June 2017 period. First, retrievals including different sensors will be compared to one another to understand when and where differences may be apparent. Next, all retrievals are evaluated against radiosondes as a common standard. Finally, retrieval-radiosonde comparisons are considered in cloudy and cloud-free conditions to evaluate if sensitivity to clouds becomes more or less significant with various sensors included.

3.1 Retrieval Intercomparisons

3.1.1 Relative Differences

In order to establish the overall impact of adding the RLID and wvDIAL to the retrieval, we will first examine the relative differences in T and $WVMR$ of retrievals including active sensors compared to the base passive-only retrieval. Figures 1 and 2 show the mean differences between the AERIonly retrieval and the AERIrLID and AERIvDIAL retrievals, respectively, for the full analysis period. Both active-inclusive retrievals result in small impacts on the T profile in a mean sense, with average differences at all levels being less than 0.5 °C (Figs. 1 and 2). The standard deviation of the differences grows with height up to 1.5 km, which is expected since the AERI still suffers from a lack of information at higher altitudes. Above 1.5 km, the standard deviation of the differences (especially in the water vapour field) is approximately constant. This is also expected due to the lack of AERI information at higher altitudes. However, there are some interesting features that can be seen between 100 and 300 m in the individual points (grey markers) in both the AERIrLID and AERIvDIAL temperature retrievals. Both show some sort of 'inflection' point at 200 m. The differences associated with the AERIrLID retrieval trend warm below this inflection point and cool above it, but no such trends are apparent in the AERIvDIAL retrieval. The source of these features is quite unclear and will require more detailed analysis in future work.

The largest differences in $WVMR$ occur from 1–1.5 km a.g.l. (Fig. 1 and 2). Given this is a fairly typical boundary-layer height (see, for example, the Krishnamurthy et al. 2020 analysis of SGP boundary-layer heights), this could suggest that the RLID and wvDIAL help the retrieval better capture the moisture gradient in or near the entrainment zone. The AERIrLID retrieval differs more from AERIonly retrieval than the AERIvDIAL retrieval, with mean differences of up to 0.75 g kg^{-1} occurring in the 1–1.5 km layer. In

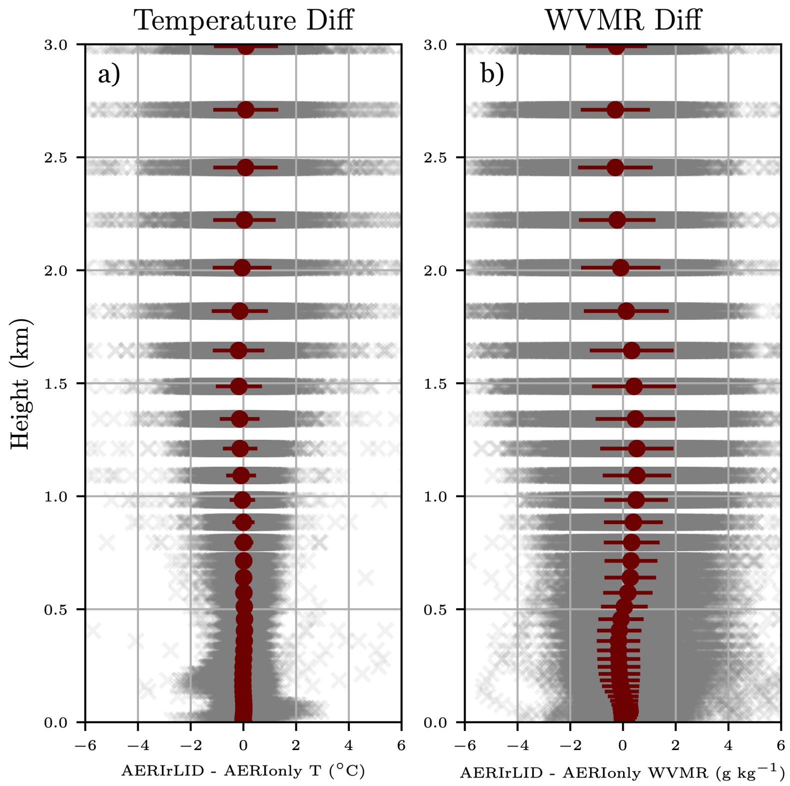

Fig. 1 This shows the profile of the differences between the AERIrLID retrieval and AERIonly retrievals for T (a) and $WVMR$ (b) at each retrieved level. The red points show the mean difference while the grey points are the individual differences. The errorbars indicate the standard deviation of the differences

comparison, the AERIvDIAL only differed in the mean by up to 0.25 g kg^{-1}. This makes some sense as the wvDIAL data were commonly limited to 1 km (Newsom et al. 2020; see Figs. 7 and 8). This absence of wvDIAL data means that AERIonly and AERIvDIAL are often using effectively identical information in that layer. The AERIrLID retrieval also tends to be drier below 1 km and above 2 km.

3.1.2 Time–Height Differences

Given the variable structure of the boundary layer through the diurnal cycle, it is important to evaluate how the differences change as a function of time. Figure 3 shows the mean difference of T and $WVMR$ in a time–height cross-section. While there is little signal in the T field (Fig. 3a, c), there is a clear

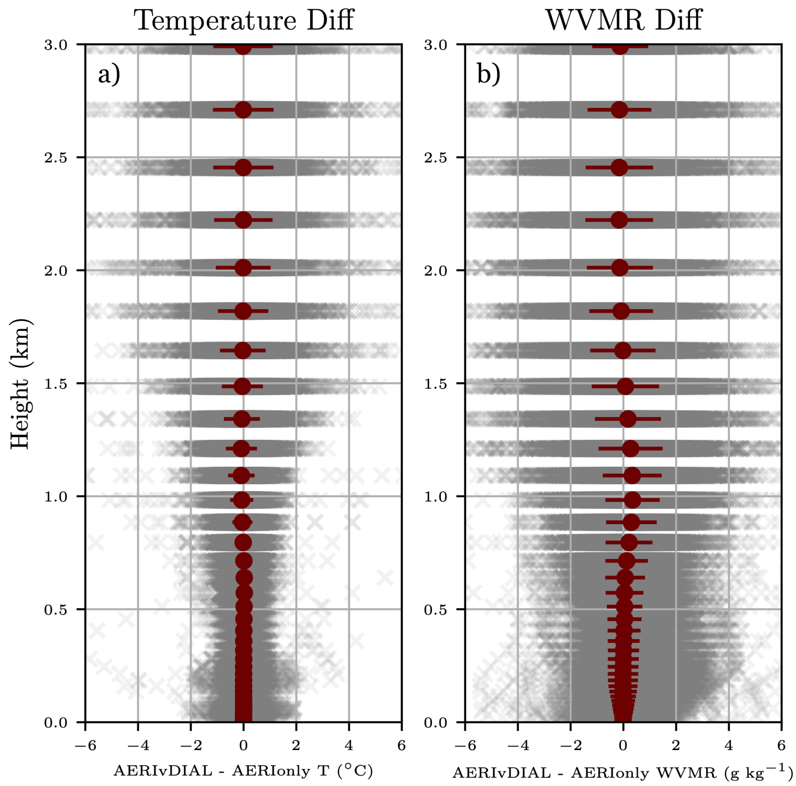

Fig. 2 As in Fig. 1, but for the AERIvDIAL retrieval

pattern to the differences in the $WVMR$ fields, especially in the AERIrLID retrieval. This pattern appears to follow the typical pattern of a growing atmospheric boundary layer (e.g., Stull 2012) during the 1400–2000 UTC period.

Sunrise during the measurement period occurs at approximately 1100 UTC, while sunset occurs around 0100 UTC. Starting at 0900 UTC, both the AERIrLID and AERIvDIAL retrievals have a period where they are more moist than the AERIonly retrieval near the surface. This could be the moisture surge which has been documented in the early morning hours (e.g., Blumberg et al. 2017a; Chilson et al. 2019). As seen in the Sect. 3.1.1, later in the day the AERIrLID and AERIvDIAL retrievals are more moist in the layer from 0.5 km to 1.5 km, before becoming drier than the base retrieval above this layer. The overall shape is reminiscent of the classical idealized boundary-layer growth model, with the boundary layer growing with time after the sun rises. This further supports the suggestion that the AERIrLID and AERIvDIAL runs are representing the moisture gradient in the entrainment zone differently than

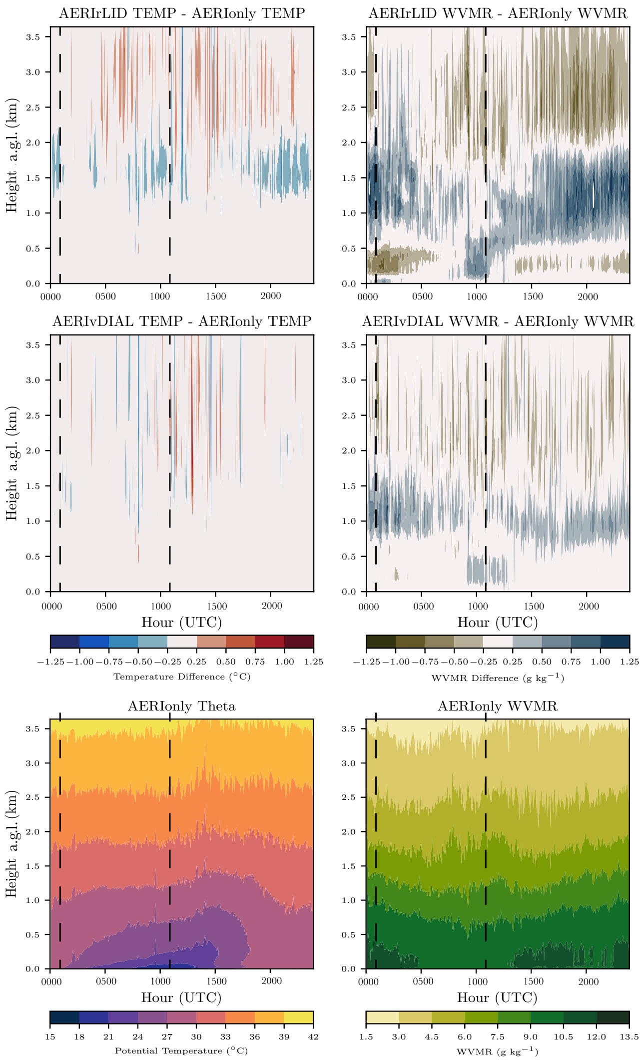

Fig. 3 Mean difference of T and $WVMR$ shown in a time-height cross-section comparing the AERIonly retrieval to AERIrLID (upper panels) and AERIvDIAL (middle panels). The mean potential temperature and WVMR from the same period are shown on the bottom panels. The composites use data from 15 May 2017 to 12 June 2017. The vertical dashed lines show the approximate sunset and sunrise times

the AERIonly version. It could be that the active sensors are better able to capture moisture gradients (since there is less smoothing due to the lack of information at higher altitudes) and this results in better represented moisture gradients in the retrieval.

3.1.3 Correlation Matrix Differences

While evaluating the derived T and $WVMR$ profiles is useful from a more operational standpoint, it is also beneficial to take advantage of the retrieval's posterior covariance matrix. The ideal posterior covariance matrix is one where all the off-diagonal components are zero. This implies that the retrieval has enough information in the observations for each level such that it does not have to rely on the prior covariance matrix to determine a solution (Turner and Blumberg 2019). The posterior covariance matrices from a selected retrieval time were converted to correlation matrices and are shown in Fig. 4.

The AERIoe-retrieved posterior correlation matrices of the AERIrLID and AERIvDIAL both exhibit improvements, namely by reducing the magnitude of the off-diagonal correlation values, over the AERIonly retrievals, most notably in the water vapour field. The addition of water vapour data from the lidar shows little impact on the temperature field in terms of the level-to-level covariance. Regarding the correlated error in the water vapour retrievals, the AERIrLID has the most improvement, with data below 2 km being mostly independent. The level-to-level correlations above 2 km are similar in shape to those in Turner and Blumberg (2019), though slightly larger in magnitude. While not as drastic as the AERIrLID retrieval, the AERIvDIAL retrieval also shows improvement in the posterior correlations, especially below 1 km.

In Sect. 3.1, we examined the relative differences between the three retrievals in different ways: the bulk differences with height (Sect. 3.1.1), the differences in time and height (Sect. 3.1.2), and the relative differences in the posterior correlation matrices (Sect. 3.1.3). These sections show that adding other measurement types into the retrieval does produce differences that may be related to physical phenomena. However, these differences do not provide information about 'truth' or accuracy.

3.2 Comparison with Radiosondes

To evaluate the performance of the AERIoe retrievals relative to a common standard, we consider all 109 SGP radiosonde profiles described in Sect. 2.5 as a baseline. Comparisons include data below 4 km a.g.l., with statistics summarized in Table 2[2]. In general, there is robust statistical agreement between the radiosonde observations and AERIoe retrieval profiles for both T (Fig. 5a–c; Pearson correlation coefficient $R^2 > 0.98$ for all) and $WVMR$ (Fig. 5d–e; $R^2 > 0.92$ for all). The AERIrLID retrieval (Fig. 5b) performed the closest to

[2] Cloudy scenes were not controlled for in this analysis as those comparisons are reserved for Sect. 3.3.

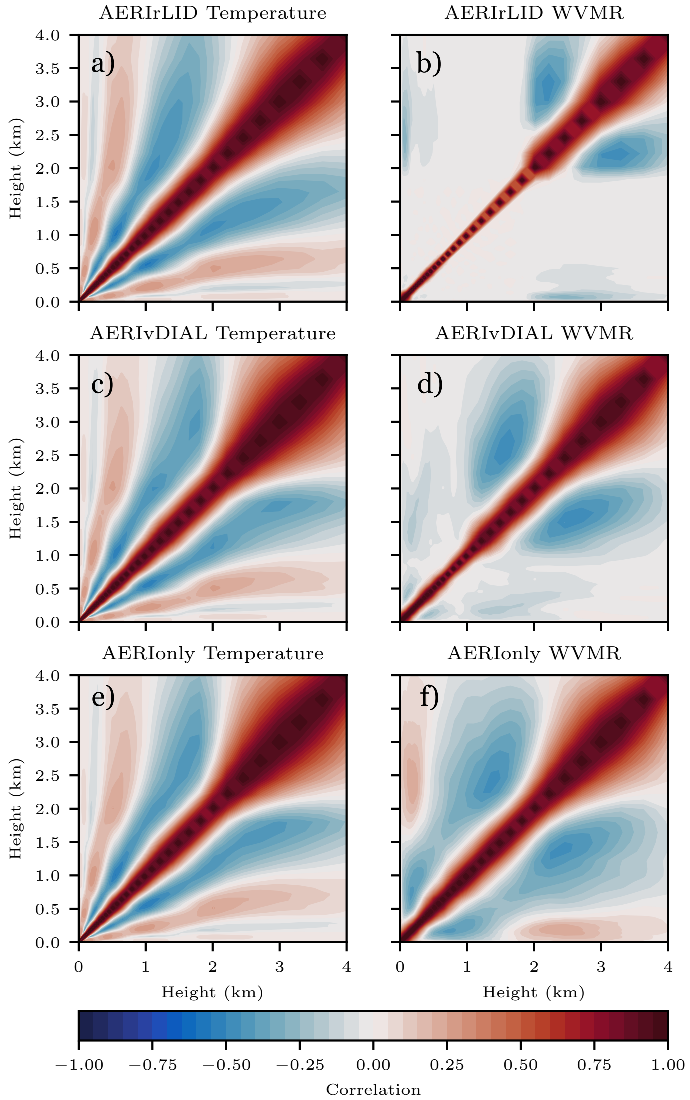

Fig. 4 Representative posterior correlation matrices for the AERIrLID (a–b), AERIvDIAL(c–d), and AERIonly (e–f) retrievals. The first column (a, c, e) contains correlation matrices for T while the second column (b, d, f) contains correlation matrices for $WVMR$. These matrices are from 31 May 2017 at 0245 UTC

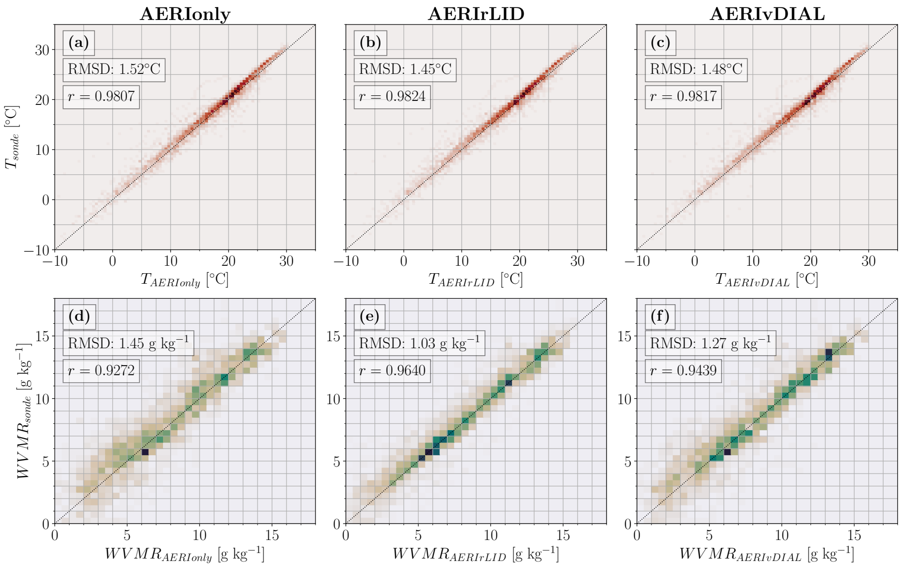

Fig. 5 Two-dimensional histograms comparing AERIonly (a and d), AERIrLID (b and e), and AERIvDIAL (c and f) temperature (a, b, c) and water vapour mixing ratio (d, e, f) retrievals to collocated radiosonde observations at levels below 4 km a.g.l. Temperature and $WVMR$ are binned by 0.5°C and 0.5 g kg^{-1}, respectively. Also included on each panel are the root mean square differences ($RMSD$; in respective units) and Pearson correlation coefficient (R^2) between each observational technique. These values are reproduced in Table 2 for clarity

Table 2 Root mean square differences ($RMSD$; in respective units) and Pearson correlation coefficient (R^2) for each retrieval relative to contemporaneous radiosonde observations in the 0–4 km a.g.l. layer.

Retrieval	**Temperature**		$WVMR$	
	$RMSD$ (°C)	R^2	$RMSD$ (g kg^{-1})	R^2
AERIrLID	1.45	0.9824	1.03	0.9640
AERIvDIAL	1.48	0.9817	1.27	0.9439
AERIonly	1.52	0.9807	1.45	0.9272

the radiosonde temperature observations with a root mean squared difference ($RMSD$) of 1.45°C and correlation coefficient of 0.9824 being the lowest and highest, respectively, of the three set-ups. The AERIvDIAL and AERIonly retrievals (Fig. 5a, c) follow closely behind. Comparisons for $WVMR$ follow the same order of similarity as for T: AERIrLID (Fig. 5e) performed the closest, with $RMSD$ of 1.03 g kg^{-1} and correlation coefficient of 0.9640. AERIvDIAL (Fig. 5f). AERIvDIAL and AERIonly followed in that order, with increasing $RMSD$ and decreasing R^2 as shown in Table 2.

Since the spread in the bulk comparison statistics for the three retrievals in both T and $WVMR$ is relatively small, it is insightful to examine performance as a function of height (Fig. 6). For example, the spread in T (Fig. 6a) as indicated by the interquartile range (IQR) is relatively large for all retrievals

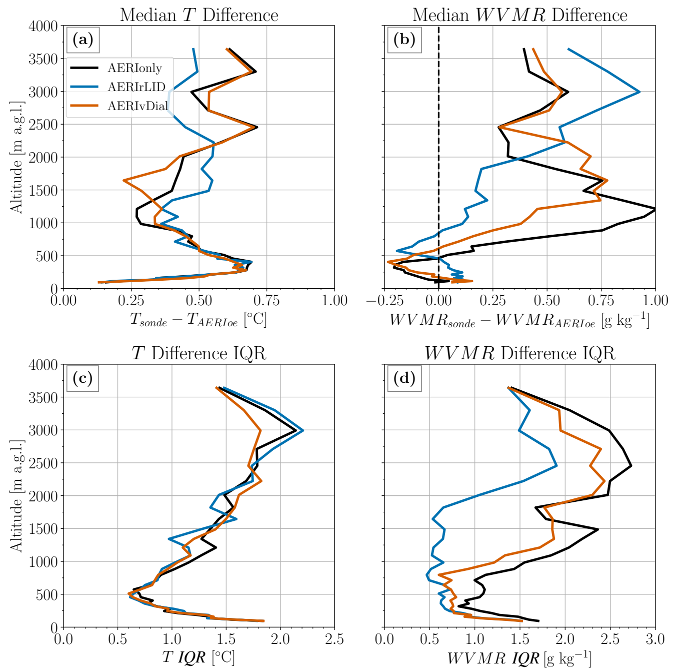

Fig. 6 Median differences from the radiosonde observations in T (a) and $WVMR$ (b) versus altitude a.g.l. for the three different retrievals (the legend in (a) is valid for all four panels). The interquartile ranges of T (c) and $WVMR$ (d) are also included to emphasize the variability in each retrieval

close to the ground, reaches a minimum around 500 m, and generally increases with height, again reaching a maximum around 3000 m. While the median differences for all three retrievals track closely with altitude, it is apparent that the AERIrLID specifically compares the best above 2250 m, which is likely the basis for its leading performance in a bulk sense. Radiosonde-retrieval comparisons binned by radiosonde launch time (not shown) suggest relative maxima in median differences and IQR near the surface may be related to nocturnal boundary layers (0060 UTC and 0012 UTC median differences are largest near the surface). However, given radiosondes can also be imperfect sensors and require surface input data, it is hard to draw conclusions from this dataset alone.

The comparisons versus height for $WVMR$ (Fig. 6b) are more pronounced than those for T. There is again a pronounced spread between the differences for each retrieval as compared to the radiosondes in the lowest 300 m that decreases vertically until around 500 m. In this surface to 500 m layer, the median differences for all three retrievals are within 0.25 g kg^{-1} in magnitude. Between 500 and 2000 m, the AERIonly and AERIvDIAL retrievals increase in median difference and IQR spread with height, whereas the AERIrLID remains relatively small for both. This layer is likely the predominant cause for the AERIrLID performing the strongest in the bulk analysis (Fig. 5e). This makes sense as the 10-min RLID $WVMR$ product has very good signal-to-noise ratio in this layer. Above 2500 m, all three retrievals maintain a roughly constant $WVMR$ bias with height compared to the radiosondes, although the IQR tends to decrease.

3.3 Sensitivity to Clouds

The presence of clouds has impacts on thermodynamic and radiative properties in the boundary layer and in the atmosphere more generally. While sensitivities to clouds may be understood for each individual measurement platform considered in this work, it is additionally important to understand how cloudiness might impact retrievals combining active and passive sensors. It is important to note that here we are referring to clouds near the top of the boundary layer or above. Since most clouds are opaque to these instruments, low cloud would prevent observation over the depth of the boundary layer. Lidars, such as vDIAL or RLID, are able to profile into a cloud until about an optical depth of 1; thus, it is possible to get a partial profile into a cloud. However, for liquid water clouds this vertical distance is usually pretty small — O(10 m), which is about 1 range gate — thus we tend to ignore it. The AERIoe-retrieved values start to get affected by the cloud presence at a height equal to cloud-base height minus one half of the vertical resolution of the retrieval at cloud base (see Turner and Blumberg 2019, Fig. 13). If additional information (e.g., lidar profiles) are added to the AERI retrieval, then the vertical resolution improves, and the region not impacted by the cloud gets closer to the cloud base.

This analysis follows a similar method to the analysis presented in Sect. 3.2, but in this case, data are classified into either overcast or clear periods. Using cloud-base height detected by RLID, this classification uses a two-hour rolling window to classify the period as overcast (continuous cloud-base height detected during the period) or clear (no cloud-base height detected during the period). Both rectangular and Gaussian rolling windows were tested for application in this method, but results were quite similar. Periods with inconsistent detection of cloud-base height were classified as unclear and not considered in this work. Overcast periods include 35 samples, while clear periods include 45 samples.

Comparisons of each retrieval under overcast and clear conditions are shown in Fig. 7. Generally, these mean profiles and spreads, as indicated by the IQR, show similar results, as shown in Fig. 6, as expected. In some instances, the mean retrieved profiles of T and $WVMR$ have slightly larger differences from radiosonde profiles under overcast conditions for each considered retrieval. These results are consistent with those shown in Wulfmeyer et al. (2015), where AERIonly retrievals were compared in a similar way. To understand if any of these differences between overcast and clear conditions are statistically significant, a student t-test is used (Fig. 8). In this case, larger

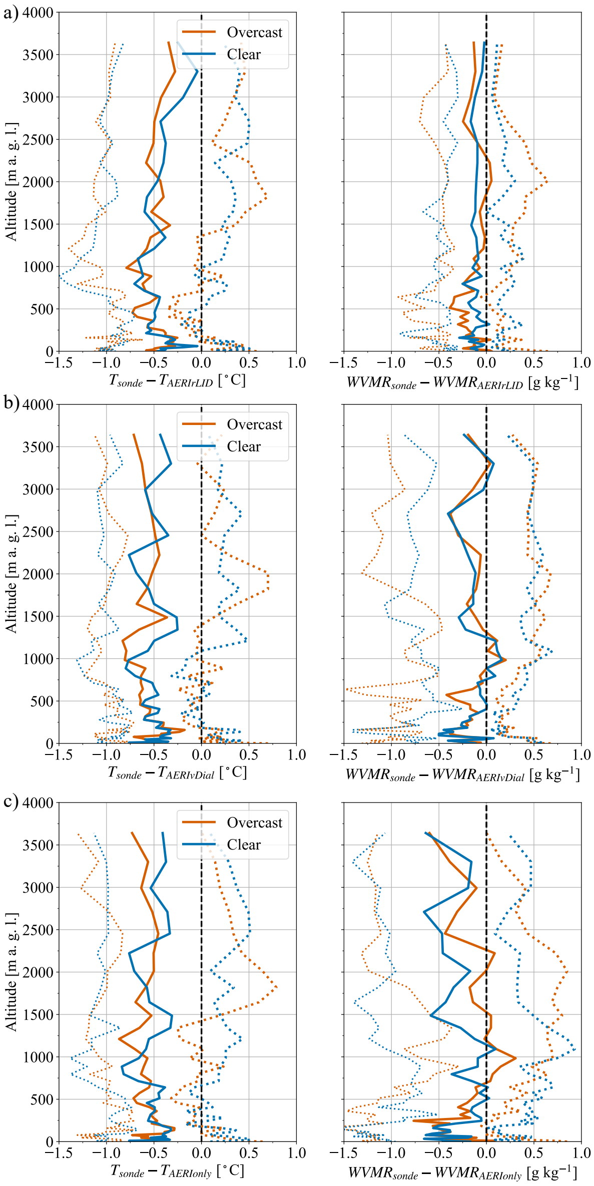

Fig. 7 Median differences from the radiosonde observations in T (left) and $WVMR$ (right) versus altitude a.g.l. for AERIvDIAL (a–b), AERIonly (c–d), and AERIrlid (e–f). Blue curves represent clear conditions, while orange curves represent overcast conditions. Also included are the 25^{th} and 75^{th} percentile differences (dotted lines) to emphasize the variability in each retrieval

magnitude t-values would indicate that differences between mean retrieval profiles under overcast and clear conditions are large. Statistical significance is not found at any level for any retrieved temperature profile, suggesting while slight differences are present in the profiles, T retrieval performance is not significantly sensitive to overcast conditions. Similarly, statistical significance is not found at any level for AERIvDIAL $WVMR$ retrieval. Differences between mean $WVMR$ retrieval profiles under overcast and clear conditions are significant at 300 m a.g.l. for AERIonly and just below 500 m a.g.l. for AERIrLID.

It should be noted that such thin layers of significance may not hold much meaning given the vertical resolution of the retrieved profiles. This may be especially true since the reference data (i.e., radiosondes) were not convolved with the averaging kernel provided in the retrieval (Löhnert et al. 2009) to match the effective vertical resolution of the retrieved profiles, which varies based on the atmospheric conditions[3]. Still, these levels correspond with interesting features in the difference profiles (Fig. 7). Under clear conditions, AERIonly $WVMR$ retrievals tend to depict drier profiles at 300 m than under overcast conditions (see Fig. 7d). This level shows diverging or mirrored difference profile shapes, which is unlike elsewhere in the profile where differences are largely related to a shifted profile with similar shape. A similarly drier clear profile with a diverging or mirrored shape compared to the overcast profile is apparent near 500 m in the AERIrLID comparisons (see Fig. 7d). It is not clear why this is the case in either retrieval. It is worth noting that 300 m is the level above which the thermodynamic retrieval prevents lapse rates from becoming steeper than superadiabatic[4]. However, the 500m level bears no particular significance to the retrieval or any of the constraints applied to it, so perhaps the differences at 300 m are simply coincidental.

4 Applied Analysis

In addition to the bulk analyses presented in Sect. 3, we also evaluated these data in more applied settings to showcase how these retrieved observations might be useful in various applications, and how added information in the retrievals may thus be important. First, we evaluate how various versions of the retrieval impact land–atmosphere coupling metrics important to understand-

[3] In these applications, convolving the radiosonde data with the averaging kernel would act to minimize the vertical representativeness error in the comparison of the AERIoe retrievals and the radiosonde profiles. The authors purposefully chose not to take this step. In this sort of analysis we feel it is important to evaluate the data as most users would encounter it. This does mean that our results may make the retrieval appear to fare less well than it may if the reference data were convolved with the averaging kernel. See Turner and Löhnert (2014).

[4] This is one of two physical constraints added to the retrieval, and the level below which it is applied is configurable by the user. The other constraint requires relative humidity be less than 100% (Turner and Blumberg 2019). Metadata about these settings can always be found in retrieval output.

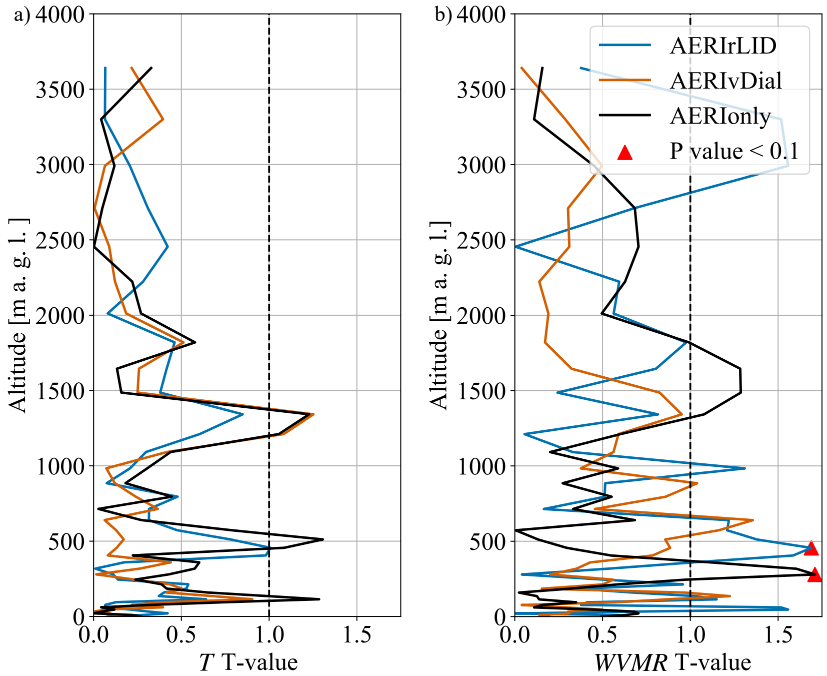

Fig. 8 Profiles of student t-test values (values are shown as magnitudes) from the comparison of each retrieval's median profile under overcast and clear conditions are shown for a) T and b) $WVMR$. Red triangles mark levels where $p > 0.1$, indicating confidence that differences are significant

ing how the underlying land surface interacts with and modifies the atmosphere. Next, the derivation of common convection indices and the impacts of including data from active sensors on them is explored. Finally, we introduce a case of severe convection near the observation site to evaluate how retrieved boundary-layer information may be valuable on short time scales preceding severe weather.

4.1 Land–Atmosphere Coupling Metrics

Land–atmosphere coupling metrics describe the degree of covariability between the land surface and atmosphere. In the absence of larger scale atmospheric forcing, soil-moisture driven changes to surface flux partitioning can influence the development of clouds and precipitation. The degree of atmospheric sensitivity to these changes varies based on climate; however, semi-arid regions such as the Southern Great Plains have been shown to display greater sensitivity to changes in evapotranspiration (Trenberth 1999; Guo et al. 2006; Koster et al. 2011; Wei et al. 2016). The Convective Triggering Potential and Low-level Humidity Index ($CTP - HI_{low}$) framework (Findell and Eltahir 2003a,b) uses vertical profiles of temperature and moisture taken in the early morning before the convective boundary layer begins to develop — 1200 UTC in the U.S. — to diagnose the atmosphere's preconditioning toward land–atmosphere coupling. In other words, the framework determines whether locally triggered convection is more likely over dry or wet soils based upon atmospheric instability and moisture within the lower troposphere. CTP is computed by integrating the area between the temperature profile and the moist adiabat drawn upward from the temperature observed 100 mb above the surface to a point 300 mb above the surface, while HI_{low} is defined as the sum of dew point depressions (the difference between temperature and dew point) at 50 and 100 mb above the surface. Traditionally, observational applications of the framework rely on vertical profiles derived from radiosondes, which leads to undersampling of the boundary layer in time and in horizontal space. Ground-based remote sensing techniques can provide boundary-layer profiles where radiosonde data is sparse, and at a much finer temporal resolution. As such, estimates of land–atmosphere coupling from metrics such as the $CTP - HI_{low}$ framework can be obtained for multiple profiles in time and space using ground-based remote sensors.

Using the $CTP{-}HI_{low}$ (hereafter $CTP{-}HI$) framework, we classified each day during the observation period using AERIonly, AERIrLID, and AERIvDIAL thermodynamic retrievals within the hour corresponding to the time of the 1200 UTC radiosonde observation, which may be as early as 1100 UTC. First, we identified days in which the retrieval CTP and HI values produced the same classification for atmospheric pre-conditioning as was identified by radiosonde profiles. All three retrievals were able to produce the same classification as the radiosonde over 75% of the time (Fig. 9).

The rigid nature of the categorical thresholds to characterize atmospheric preconditioning can result in two platforms having nearly identical CTP and HI values, but different classifications. Small differences in CTP or HI values may be within the observational range of uncertainty. Therefore, we produced $CTP-HI$ classifications for all $CTP-HI$ combinations within a 1-standard-deviation range of uncertainty. If a combination within this uncertainty range produced the same classification as the radiosonde data, then it was counted as matching only within the range of uncertainty. Introducing the range of uncertainty resulted in an additional 10% of AERIrLID and AERIvDIAL days that matched radiosonde classifications while for AERIonly data this percentage was slightly lower. Consequently, when we included a range of uncertainty all three retrievals were able to produce the same classification as that obtained from radiosonde data, nearly 90% of the time (Fig. 9). AERIrLID retrievals performed best at producing the same classification, followed by AERIvDIAL and AERIonly.

When CTP and HI obtained from each retrieval were compared to radiosonde values (Table 3), covariability between radiosonde and retrieval observations of these quantities was strong. All three retrievals displayed similar R^2 values at or above 0.65 for CTP. AERIvDIAL produced the smallest median difference and IQR in CTP differences, while AERIonly had the largest

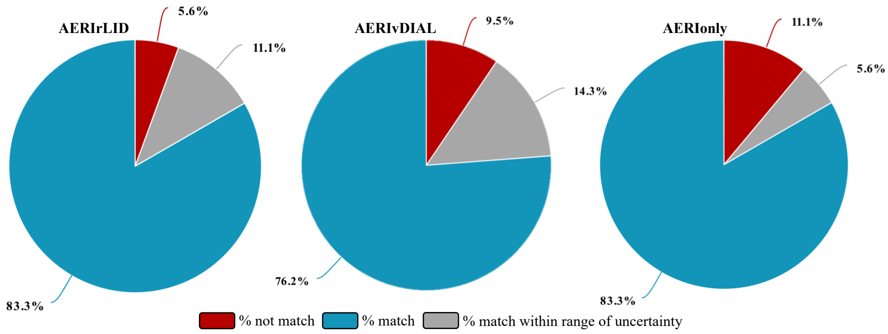

Fig. 9 Percentage of days in which CTP and HI obtained from retrievals produce the same classification of atmospheric conditions as CTP and HI obtained from radiosondes (teal). Grey shading indicates the days in which retrieval CTP and HI only produce the same classification as radiosonde data within a window corresponding to the retrieval range of uncertainty (within one standard deviation). Red shading corresponds to days in which neither the observed CTP and HI values or values within a range of uncertainty produced the same classification as radiosonde CTP and HI

Table 3 Median difference, difference IQR and R^2 statistics for retrieval minus radiosonde CTP and HI values. Bold values denote most favourable values (smallest differences or highest R^2)

CTP	**median difference**	**difference IQR**	R^2
AERIrLID–sonde	-27.18	109.78	**0.70**
AERIvDIAL–sonde	**-19.39**	**86.69**	**0.70**
AERIonly–sonde	-34.43	148.45	0.65
HI	**median difference**	**difference IQR**	R^2
AERIrLID–sonde	**0.11**	**3.22**	**0.92**
AERIvDIAL–sonde	0.32	5.38	0.82
AERIonly–sonde	-0.95	5.33	0.68

median difference, the greatest IQR and the lowest R^2 values. AERIrLID had the highest R^2 value (0.92) for HI as well as the lowest median difference and difference IQR. Median difference was greatest in magnitude for AERIonly, but difference IQR was nearly the same for AERIonly and AERIvDIAL. All three retrievals, however, had R^2 values above 0.65 indicating good agreement between retrieval and radiosonde HI values.

Introducing active sensors into the AERI retrievals does appear to improve estimation of the two quantities used in this land–atmosphere coupling metric. The most pronounced benefit, as demonstrated by the best linear relationship between retrieval- and sonde-derived values, was realized in the AERIrLID observations of HI (Fig. 10a), though AERIvDIAL(Fig. 10b) also performs noticeably better when compared to AERIonly observations (Fig. 10c).

Improvements were less obvious for CTP (Fig. 10d–f). CTP is an integrated metric, and AERI observations of integrated quantities such as convective available potential energy (CAPE) have been shown to have greater uncer-

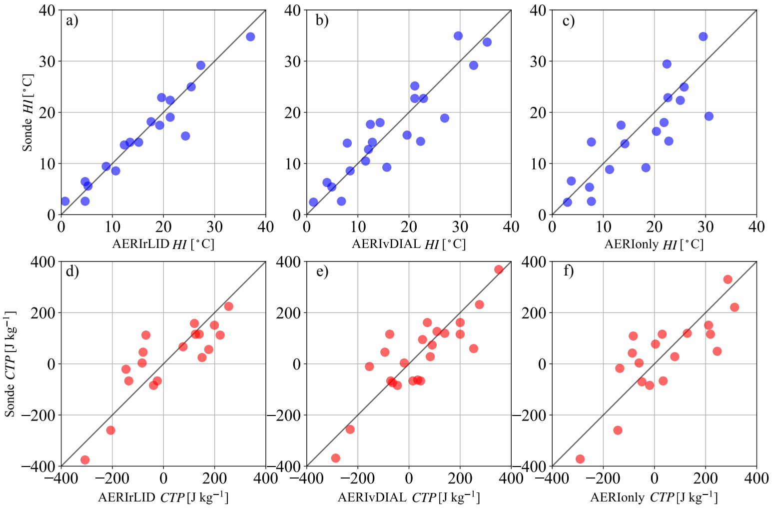

Fig. 10 Scatter plots of retrieval observations (x-axis) versus radiosonde observations (y-axis) of HI (a–c) and CTP (d–f)

tainty than non-integrated quantities (Blumberg et al. 2017a). Also, CTP is obtained at higher levels (within a 20-mb deep layer from 100 mb AGL to 300 mb AGL) in the atmosphere than HI (levels 50 to 150 mb above the surface). As vertical resolution decreases with height in the AERI retrievals, there is inherently greater uncertainty associated with the CTP observations at higher levels. The impact of vertical resolution (see footnote 3) is explored in Wakefield et al. (2021), where comparing CTP and HI obtained from radiosonde profiles with the same vertical resolution as AERI retrievals does show some improvement to the agreement between platforms. Even so, the limitations associated with using AERI are minor, and are outweighed by the ability to observe these and other coupling metrics at a high temporal resolution and outside of the commonly available radiosonde observation times and locations. This particular utility of the AERI retrievals is further addressed in Wakefield et al. (2021).

4.2 Retrieved Convection Indices

Because the retrieval provides a full covariance matrix for each retrieved solution, the uncertainties of convection indices from that profile can be derived.

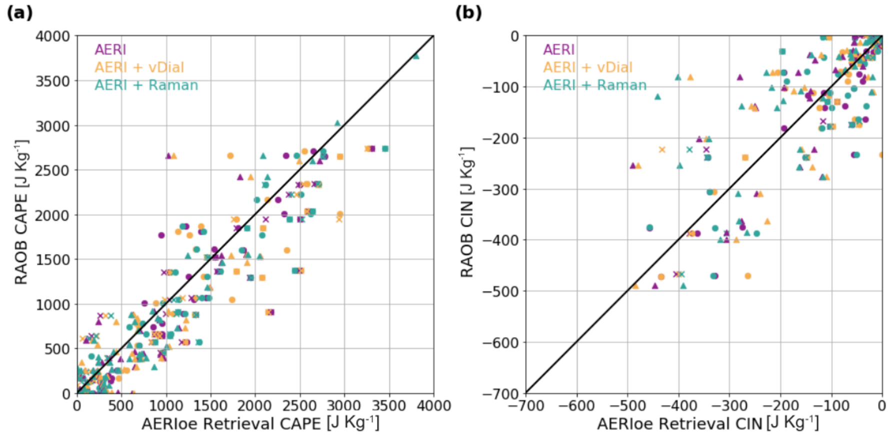

Fig. 11 Scatter plots showing comparisons for CAPE (a) and CIN (b) indices derived from the different clear-sky AERIoe retrievals and radiosonde observations. Marker styles indicate the parcel type used, where circles are the most-unstable parcel, x is the surface-based parcel, and triangles are the 100-mb mixed-layer parcel. Marker colours denote the different AERIoe retrievals

Following Blumberg et al. (2017a), Monte Carlo sampling is performed to generate 500 profiles for each retrieval and radiosonde profile. For each profile out of the 500, a set of convection indices (e.g., convection available potential energy, or CAPE, and convection inhibition, or CIN, etc.) are generated. For each index, an estimate of that index's uncertainty is derived using non-Gaussian statistics (median, interquartile range) since Gaussian statistics sometimes do not well describe the distribution of variables with bounds. Convection indices are derived using the Sounding and Hodograph Analysis and Research Program in Python (SHARPpy; Blumberg et al. 2017b). By comparing the convection indices derived from the different retrievals to the radiosondes, the influence of the active sensors in the AERIoe retrieval relative to the AERIonly retrievals can be understood.

The CAPE indices derived from the different AERIoe retrievals and radiosondes were first compared. Figure 11 shows scatter plots from these comparisons for different parcel types (surface-based, most unstable, 100-mb mixed layer). For CAPE, the scatter plot displays a very strong relationship between the CAPE values measured between the two techniques (radiosonde and AERIoe) with no noticeable differences between the different AERIoe

Table 4 Comparison statistics between different clear-sky AERIoe retrieval configurations and radiosondes for the CAPE index using the surface-based (SB), 100 mb mixed-layer (ML), and most-unstable (MU) parcels. Statistics shown are the number of cases (n), the bias, the 1-sigma standard deviation of the errors (*Std.Dev*), and the correlation coefficient (r). The median values from the computed convection index distribution are used in this comparison

SBCAPE	*Retrieval*	n	*Bias*	*Std. Dev.*	r
	AERIrLID	74	99.4	275.9	0.98
	AERIvDIAL	71	103.4	315.1	0.97
	AERIonly	71	95.3	294.9	0.97
MLCAPE	*Retrieval*	n	*Bias*	*Std. Dev.*	r
	AERIrLID	74	42.7	196.9	0.97
	AERIvDIAL	71	53.1	294.8	0.92
	AERIonly	71	50.4	299.8	0.92
MUCAPE	*Retrieval*	n	*Bias*	*Std. Dev.*	r
	AERIrLID	74	110.1	331.1	0.96
	AERIvDIAL	71	93.2	498.9	0.90
	AERIonly	71	102.1	472.2	0.91

configurations (Fig. 11a). Table 4 further echoes this result, as for all parcel and retrieval types, the correlation coefficient is above or equal to 0.9. One noticeable difference is that the AERIrLID retrievals display the lowest standard deviation of the errors for all parcels relative to the other retrievals. This result is likely a consequence of the improved signal-to-noise ratio of the RLID instrument being used in the retrieval. All of the AERIoe retrievals also exhibit a slight positive bias for CAPE. This bias is small, with an average value of less than 84 J kg^{-1}. This is certainly within the uncertainty range of the retrieval, and different CAPE calculation methods can result in even larger differences, so this bias is likely not all that meaningful.

It appears that unlike CAPE, CIN derived from the retrievals using active sensors does not compare better to those observed by the radiosonde. Figure 11b indicates that although there is noticeable scatter along the 1-to-1 line, there still is a visible relationship between the CIN values of the two measurements. One such reason for this is that the extra information provided by the wvDIAL and RLID instruments describes the structure of water vapour, whereas the CIN calculation is strongly dependent upon the retrieval's ability to resolve temperature inversions above the parcel source height. Although the a priori dataset in the retrieval does describe cross-correlations between temperature and humidity, it does not appear that information provided by the active sensors is sufficient to reliably depict the inversion at the top of the PBL. These poor CIN comparisons were also seen in Blumberg et al. (2017a), and better comparisons may require modifications to the retrieval to leverage other information from active sensors (e.g., vertical backscatter gradients to identify the PBL top) to help resolve the elevated inversions better.

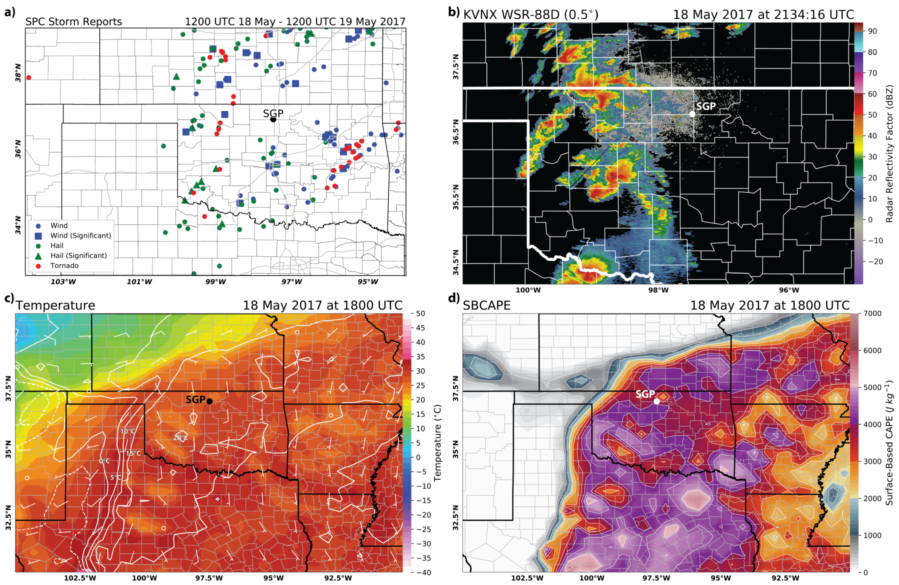

Fig. 12 18 May 2017 a) severe storm reports sourced from the NWS *Storm Data* report, b) Vance Air Force Base (KVNX) WSR-88D radar reflectivity at 2134 UTC, c) 1800 UTC surface temperature (°C, colour fill), dew point temperature (°C, contour), and winds (barbs), and d) 1800 UTC surface based CAPE. Surface variables and CAPE come from the SPC SFCOA dataset. The ARM-SGP site is marked on all panels

4.3 Severe Convection Case Study

On 18 May 2017, 135 severe weather reports (severe wind, severe hail, and tornado) were documented in Oklahoma in the National Weather Service (NWS) *Storm Data* record (NCEI 2020). While there was enough certainty in the forecast for severe weather to lead to a high risk in the Day 1 Convective Outlook from the Storm Prediction Center (SPC), uncertainty remained regarding storm coverage and timing. Specifically, if too many cells were to initiate too early in the day, the full potential of the regional instability and shear would not be realized, which could act to limit the severity of the day's weather. Figure 12 summarizes this event, depicting the day's storm reports and snapshots of convection morphology and environmental conditions.

In cases like this one, it is important to understand how the atmosphere evolves *after* the 1200 UTC operational radiosonde observation is collected—a benefit ground-based sensors can provide. At present, numerical tools are often relied on to provide some understanding of lower-atmospheric evolution. One such example is the SPC SurFaCe Objective Analysis (SFCOA; Bothwell et al. 2002), which is a comprehensive surface objective analysis scheme designed to assimilate the various real-time observational datasets using hourly mesoscale model output as first-guess fields. Since this event occurred near the ARM SGP site during the special observation period, we have a unique opportunity to compare environmental convective parameters as derived from boundary-layer profiler thermodynamic retrievals, available radiosondes, and the SFCOA. This enables intercomparison between each retrieval and exploration of potential benefits associated with high temporal resolution boundary-layer profiling for environmental characterization. Additionally, the high temporal resolution thermodynamic retrievals offer a dataset against which the SFCOA can be compared as conditions evolve.

After the 1200 UTC radiosonde observation, surface-based CAPE (SBCAPE) represented by SFCOA and all retrievals rapidly increases with the onset of daytime heating (Fig. 13). Prior to approximately 1500 UTC, SBCAPE in the SFCOA data is consistently larger than the SBCAPE in any of the retrievals. The SFCOA uses Rapid Refresh (RAP) model profiles as a first guess for the objective analysis. Upon comparison of RAP and retrieval profiles, it becomes apparent that prior to sunrise (between 1100 and 1200 UTC) all retrieved profiles depict the warm nose near 900 mb as too weak and too smooth (Fig. 14), thus representing it as too deep. The AERIoe retrieval includes a constraint which prevents the T profile from becoming superadiabatic above a specified height (in this configuration 300 m a.g.l.; see footnote 4). This results in the retrieved profile remaining too warm above the warm nose. This warmer temperature aloft can result in lower CAPE and increased CIN.

From 1500–1800 UTC, SBCAPE values from the SFCOA and all retrievals remain in approximate agreement. Retreival values vary from 1615–1730 UTC, which was related to broken cloud (ARM Total Sky Imager (TSI) observations; not shown). After 1730 UTC, retrieval values become much less variable, and the general value of surface-based CAPE decreases by a small amount. The same TSI observed consistent cloud cover from 1800–1900 UTC. While clouds do impact the profile-to-profile variability for all retrievals (i.e., intermittent clouds result in more variability), there does not appear to be strong sensitivity of the general mean value of SBCAPE to clouds, consistent with findings in Sect. 3.3.

In this case, differences in between AERIonly, AERIrLID, and AERIvDIAL retrieved SBCAPE time series were small and intermittent. Adding active sensors made little impact on these derived values. Several other common convective parametres were also explored (i.e., most unstable CAPE, surface-based and most unstable CIN, level of free convection, various boundary-layer lapse rates, all not shown) and results were generally similar. In the absence of retrieval or SFCOA data, a time series based on radiosonde-observed SBCAPE may not have been very accurate in this case[5]. Given the uncertainty about timing and thus utilization of the available instability in the region in this case, such observations may be quite misleading. As noted in Sect. 2.5, the ARM-SGP site typically collects radiosonde four times per day instead of

[5] It is of note that different methods of computing convection indices, in this case SBCAPE, can result in widely varied results, as is apparent from comparing values derived from radiosonde data by the University of Wyoming archive (orange dots on Fig. 13) and by SHARPpy (blue dots and error bars on Fig 13).

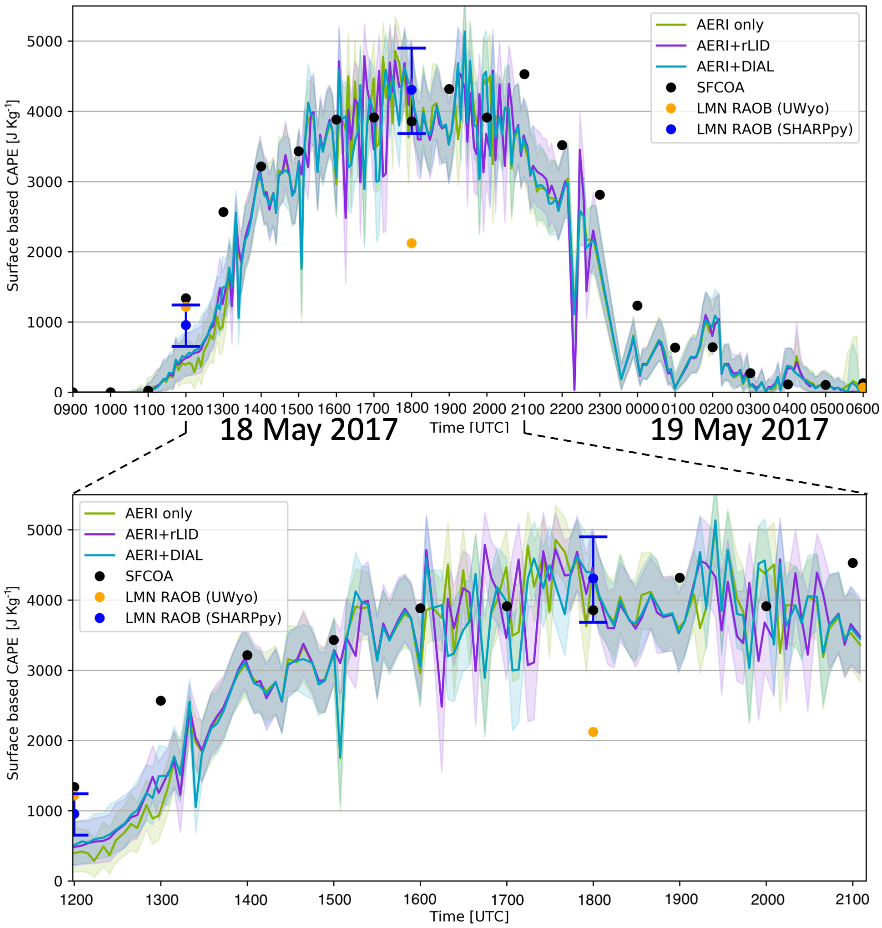

Fig. 13 Time series of SBCAPE on 18 May 2017 at the SGP site are shown for the three considered retrievals (AERIonly in green, AERIrLID in purple, and AERIvDIAL in blue). The solid coloured lines indicate the 50th percentile value of derived surface-based CAPE, while filled regions represent the spread between the 10th and 90th percentiles. Hourly SFCOA surface-based CAPE values are plotted as black dots. Any available radiosonde observation (RAOB) values from the SGP site are plotted as blue dots (processed via SHARPpy, error bars represent 10th and 90th percentiles) and orange dots (SBCAPE value recorded in the University of Wyoming archive). The upper panel includes data from 0900 UTC on 18 May to 0600 UTC on 19 May. The lower panel shows a subset of those data from 1200 to 2100 UTC on 18 May

the more typical 1200 and 0000 UTC synoptic times, meaning the 1800 UTC observation is more data than most locations collect (note that when severe risks are moderate to high, NWS operations often include special soundings beyond synoptic times). This demonstrates the importance of tools like the SFCOA and the potential benefit of profile observations of boundary-layer characteristics. Boundary-layer profilers offer the added benefit of high tem-

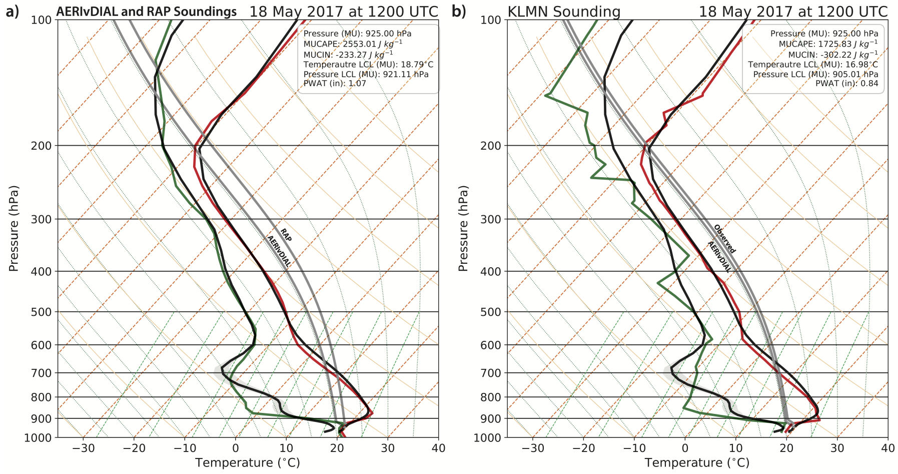

Fig. 14 1200 UTC AERIvDIAL and a) RAP, b) radiosonde profiles are compared on skew-T log-P diagrams. Retrieval profiles are depicted as dark curves with shading to represent profile variability over a 30-minute window centred on the sounding time. The dark curve on the right is the temperature (°C) and the dark curve on the left is the dew point (°C). RAP and radiosonde profiles are shown in red (temperature, °C) and green (dew point, °C). Parcel paths are labeled for each sounding. For clarity only one retrieval is shown, but results were consistent between the AERIvDIAL, AERIrLID, and AERIonly retrievals

poral resolution observed profiles. Such information can be powerful in cases where subtle changes in boundary-layer thermodynamics are important.

5 Summary

Filling the observational gap in the boundary layer is a challenge. As various technologies are evaluated and continue to emerge, it seems increasingly likely that viable observation solutions will include multiple instrument platforms. In such configurations, an additional challenge emerges: bringing multiple platforms and datastreams together to provide high quality observations and value added products. To address this challenge, we explored the combination of active and passive remote sensors deployed for thermodynamic profiling.

An experiment conducted at the ARM-SGP site in May–June 2017 (Newsom et al. 2020) provided several weeks of data for this comparison and evaluation effort. From 15 May to 12 June 2017, an AERI, wvDIAL, and RLID all operated continuously. From these data, thermodynamic profiles were retrieved via the AERIoe algorithm (Turner and Blumberg 2019; Turner and Löhnert 2020). Three sets of retrieved profiles were considered in this work: retrievals including AERI observations, retrievals including both AERI and wvDIAL observations, and retrievals including both AERI and RLID observations. The first set of analyses in this work focused on comparison and evaluation of the entire dataset in a bulk sense. The second set of analyses focused instead on some specific applications of retrieved thermodynamic profiles. We specifically aimed to highlight the differences that resulted from including active sensors in this retrieval framework, and explore the impact those differences might have in scientific applications.

The three versions of the retrieval were first compared with one another. Results showed that active-inclusive retrievals (i.e., AERIrLID and AERIvDIAL) were not very different than the passive-only (i.e., AERIonly) retrieval in terms of T (average differences less than 0.5 °C). On the other hand, active-inclusive retrievals did show mean differences in $WVMR$ of 0.25 to 0.75 g kg^{-1}, especially in the layer between 1 and 1.5 km a.g.l.. Differences in this layer and further evaluation of these differences as a function of time suggest that the active sensors help the retrieval to represent the moisture gradient across the entrainment zone near the top of the boundary layer.

Generally all retrievals agreed fairly well with a common standard—ARM-SGP radiosondes—with Pearson correlation coefficient $\mathrm{R}^2 > 0.98$ for all T retrievals, and $\mathrm{R}^2 > 0.92$ for all $WVMR$ retrievals. For both T and $WVMR$, AERIrLID performed closest to radiosonde observations. AERIoly and AERIvDIAL were quite similar in terms of T performance, but AERIvDIAL outperformed AERIonly in the $WVMR$ retrieval. Additionally, active-inclusive retrievals showed less overall spread in differences between retrieved and radiosonde observed $WVMR$ profiles. This reduced spread implies that including active sensors produces more consistently accurate profiles, at least in terms of $WVMR$.

The last set of bulk analyses compared overcast and clear periods to evaluate the impact of broad cloudiness on retrievals combining active and passive sensors. This set of retrieval-radiosonde comparisons showed similar results to the analogous comparison for the full dataset. There were some instances where retrieval-radiosonde differences were larger under overcast conditions. However, these differences were generally not found to be significant.

Land–atmosphere coupling metrics were the first application of the retrievals explored in this work. Retrieved thermodynamic profiles were used in the $CTP - HI_{low}$ framework, which determines whether locally triggered convection is more likely over dry or wet soils based upon atmospheric instability and moisture within the lower troposphere. The use of thermodynamic retrievals in this application can extend the framework to periods and locations where soundings, which are the typical input observations, are not regularly available. On days that were not atmospherically controlled, all three retrievals result in the same classification over 75% of the time; when they differed, the differences in CTP and HI often fell within the one standard deviation uncertainty range of the retrieval. Adding active sensors as constraints in AERIoe does have appear to have a positive impact on the estimation of HI. Improvements in CTP estimates were less clear. In any case, active sensors improve estimation of land–atmosphere coupling in this framework, but AERIonly retrievals can still produce desirable and applicable results.

We also evaluated the different retrievals by comparing derived convection indices against radiosonde values. As the retrievals provide a full error covariance matrix for each retrieved profile, we used Monte Carlo sampling of this matrix to provide uncertainty estimates for the convection indices. CAPE computed from radiosondes and all retrieval configurations agreed quite well with correlations of 0.9 or better. The AERIrLID retrieval did show the smallest standard deviation of errors among the configurations, likely a result of the improved signal-to-noise ratio of the RLID instrument. CIN computed from all retrieval configurations did not compare as well to CIN computed from radiosondes, consistent with prior findings (Blumberg et al. 2017a).

Lastly, the retrieved profiles were evaluated in the context of a severe convection case to understand what information is or is not currently available to forecasters in real time. We compared environmental convective parametres in Oklahoma on 18 May 2017 as derived from boundary-layer profiler thermodynamic retrievals, available radiosondes, and the SFCOA. Generally SFCOA and retrievals showed similar environments, though this comparison did highlight a propensity of the retrieval algorithm to depict overly smoothed, weak warm nose profiles. Differences between the three considered retrievals were small and intermittent, suggesting that the addition of active-sensors make small enough adjustments to the profiles to not result in large differences in derived indices. Though not shown, several other common convective parameters were also explored, and results were generally similar.

Overall, we find the addition of active sensors as a constraint in AERI-based retrievals do not make large impacts to the resulting thermodynamic profiles or indices derived from them. The same may not be true in other retrieval frameworks. There are perhaps specific applications for which gaining information about moisture at the boundary-layer top would be crucial, in which case the small changes seen in the AERIrLID and AERIvDIAL retrievals may be helpful. This suggests that for many applications, passive infrared remote sensor (e.g., AERI) profiling may provide sufficient information on the thermodynamic profile. This is an important finding given the costs associated with operating and maintaining multiple sensors. However, one important application that was not explored here is data assimilation, where quantification of information content and observation error is critical. As noted in Sect. 1, positive impacts have been noted in several studies for convection-scale forecasts when assimilating AERI-retrieved thermodynamic profiles (e.g., Degelia et al. 2019; Hu et al. 2019; Coniglio et al. 2019; Chipilski et al. 2020). More evaluation is needed to understand how to best use these observations; however, the benefits associated with the reduction in uncertainty and added information content when including active sensors cannot be overlooked in the context of data assimilation (e.g., Sect. 4.2, Turner and Löhnert 2020). While this work demonstrates that a one-size-fits-all optimal ground-based solution for boundary-layer profiling does not exist at present, we do show that active remote sensors are not necessarily a requirement for suitable thermodynamic profiles in all scenarios when passive sensors are available.

Acknowledgements E. N. Smith would like to acknowledge that most co-authors of this study are graduate students that chose to collaborate on this work as a voluntary side project related to their shared Boundary Layer, Urban Meteorology, and Land-Surface Processes Seminar course, bringing together varied expertise and offering a new opportunity to learn for everyone involved. Dr. Michael Coniglio provided useful internal review to the paper. This work was partially supported by the DOE Atmospheric System Research (ASR) program via grants DE-SC0014375 and 89243019SSC000034, and by the NOAA Atmospheric Science for Renewable Energy (ASRE) program. This work was prepared by the authors with support from the NSSL Forecast Research and Development Division (ENS) and the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce (TMB, QN). The contents of this paper do not necessarily reflect the views or official position of any organization of the U.S. Government.

References

Bell TM, Greene BR, Klein PM, Carney M, Chilson PB (2020) Confronting the boundary layer data gap: evaluating new and existing methodologies of probing the lower atmosphere. Atmos Meas Tech 13(7):3855–3872, DOI 10.5194/amt-13-3855-2020

Benjamin SG, Weygandt SS, Brown JM, Hu M, Alexander CR, Smirnova TG, Olson JB, James EP, Dowell DC, Grell GA, Lin H, Peckham SE, Smith TL, Moninger WR, Kenyon JS, Manikin GS (2016) A North American hourly assimilation and model forecast cycle: The Rapid Refresh. Mon Wea Rev 144(4):1669–1694, DOI 10.1175/MWR-D-15-0242.1

Blumberg W, Turner D, Löhnert U, Castleberry S (2015) Ground-based temperature and humidity profiling using spectral infrared and microwave observations. part II: Actual retrieval performance in clear-sky and cloudy conditions. J Appl Meteorol Climat 54(11):2305–2319

Blumberg W, Wagner T, Turner D, Correia Jr J (2017a) Quantifying the accuracy and uncertainty of diurnal thermodynamic profiles and convection indices derived from the atmospheric emitted radiance interferometer. J Appl Meteorol Climat 56(10):2747–2766

Blumberg WG, Halbert KT, Supinie TA, Marsh PT, Thompson RL, Hart JA (2017b) SHARPpy: An open-source sounding analysis toolkit for the atmospheric sciences. Bull Amer Meteorol Soc 98(8):1625–1636

de Boer G, Diehl C, Jacob J, Houston A, Smith SW, Chilson P, Schmale III DG, Intrieri J, Pinto J, Elston J, et al. (2020) Development of community, capabilities, and understanding through unmanned aircraft-based atmospheric research: The LAPSE-RATE campaign. Bull Amer Meteorol Soc 101(5):E684–E699

Bolton D (1980) The computation of equivalent potential temperature. Mon Wea Rev 108(7):1046–1053

Bothwell PD, Hart JA, L TR (2002) An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. In: 21st Conf. on Severe Local Storms, Amer. Meteor. Soc., vol JP3.1

Chilson PB, Bell TM, Brewster KA, Britto Hupsel de Azevedo G, Carr FH, Carson K, Doyle W, Fiebrich CA, Greene BR, Grimsley JL, Kanneganti ST, Martin J, Moore A, Palmer RD, Pillar-Little EA, Salazar-Cerreno JL, Segales AR, Weber ME, Yeary M, Droegemeier KK (2019) Moving towards a network of autonomous UAS atmospheric profiling stations for observations in the Earth's lower atmosphere: The 3D mesonet concept. Sensors 19(12), DOI 10.3390/s19122720

Chipilski HG, Wang X, Parsons DB (2020) Impact of assimilating PECAN profilers on the prediction of bore-driven nocturnal convection: A multi-scale forecast evaluation for the 6 July 2015 case study. Mon Wea Rev 148(3):1147–1175

Coniglio MC, Romine GS, Turner DD, Torn RD (2019) Impacts of targeted AERI and Doppler lidar wind retrievals on short-term forecasts of the initiation and early evolution of thunderstorms. Mon Wea Rev 147(4):1149–1170

Degelia SK, Wang X, Stensrud DJ (2019) An evaluation of the impact of assimilating AERI retrievals, kinematic profilers, rawinsondes, and surface observations on a forecast of a nocturnal convection initiation event during the PECAN field campaign. Mon Wea Rev 147(8):2739–2764

Findell KL, Eltahir EA (2003a) Atmospheric controls on soil moisture–boundary layer interactions. part I: Framework development. J Hydrometeorology 4(3):552–569

Findell KL, Eltahir EA (2003b) Atmospheric controls on soil moisture–boundary layer interactions. part II: Feedbacks within the continental united states. J Hydrometeorology 4(3):570–583

Goldsmith J, Blair FH, Bisson SE, Turner DD (1998) Turn-key Raman lidar for profiling atmospheric water vapor, clouds, and aerosols. Appl Opt 37(21):4979–4990

Guo Z, Dirmeyer PA, Koster RD, Sud Y, Bonan G, Oleson KW, Chan E, Verseghy D, Cox P, Gordon C, et al. (2006) GLACE: the global land–atmosphere coupling experiment. part II: analysis. J Hydrometeorology 7(4):611–625

Hartung DC, Otkin JA, Petersen RA, Turner DD, Feltz WF (2011) Assimilation of surface-based boundary layer profiler observations during a cool-season weather event using an observing system simulation experiment. part II: Forecast assessment. Mon Wea Rev 139(8):2327–2346

Hu J, Yussouf N, Turner DD, Jones TA, Wang X (2019) Impact of ground-based remote sensing boundary layer observations on short-term probabilistic forecasts of a tornadic supercell event. Wea Forecasting 34(5):1453–1476

Karan H, Knupp K (2006) Mobile Integrated Profiler System (MIPS) observations of low-level convergent boundaries during IHOP. Mon Wea Rev 134(1):92–112, DOI 10.1175/MWR3058.1

Knupp KR, Coleman T, Phillips D, Ware R, Cimini D, Vandenberghe F, Vivekanandan J, Westwater E (2009) Ground-based passive microwave profiling during dynamic weather conditions. J Atmos Ocean Tech 26(6):1057–1073, DOI 10.1175/2008JTECHA1150.1

Knuteson R, Revercomb H, Best F, Ciganovich N, Dedecker R, Dirkx T, Ellington S, Feltz W, Garcia R, Howell H, et al. (2004) Atmospheric emitted radiance interferometer. part I: Instrument design. J Atmos Ocean Tech 21(12):1763–1776

Koch SE, Fengler M, Chilson PB, Elmore KL, Argrow B, Andra Jr DL, Lindley T (2018) On the use of unmanned aircraft for sampling mesoscale phenomena in the preconvective boundary layer. J Atmos and Ocean Tech 35(11):2265–2288

Koster R, Mahanama S, Yamada T, Balsamo G, Berg A, Boisserie M, Dirmeyer P, Doblas-Reyes F, Drewitt G, Gordon C, et al. (2011) The second phase of the global land–atmosphere coupling experiment: soil moisture contributions to subseasonal forecast skill. J Hydrometeorology 12(5):805–822

Kral ST, Reuder J, Vihma T, Suomi I, Haualand KF, Urbancic GH, Greene BR, Steeneveld GJ, Lorenz T, Maronga B, Jonassen MO, Ajosenpää H, Båserud L, Chilson PB, Holtslag AAM, Jenkins AD, Kouznetsov R, Mayer S, Pillar-Little EA, Rautenberg A, Schwenkel J, Seidl AW, Wrenger B (2020) The Innovative Strategies for Observations in the Arctic Atmospheric Boundary Layer Project (ISOBAR) —Unique fine-scale observations under stable and very stable conditions. Bull Amer Meteorol Soc pp 1–64, DOI 10.1175/BAMS-D-19-0212.1

Krishnamurthy R, Newsom RK, Berg LK, Xiao H, Ma PL, Turner DD (2020) On the estimation of boundary layer heights: A machine learning approach. Atmos Meas Tech Discuss:in review, DOI 10.5194/amt-2020-439

Le Hoai PP, Abo M, Sakai T (2016) Development of field-deployable diode-laser-based water vapor DIAL. In: EPJ Web of Conferences, EDP Sciences, vol 119, p 05011

Löhnert U, Turner D, Crewell S (2009) Ground-based temperature and humidity profiling using spectral infrared and microwave observations. part I: Simulated retrieval performance in clear-sky conditions. J Appl Meteorol and Climat (5):1017–1032

Melnikov VM, Doviak RJ, Zrnić DS, Stensrud DJ (2011) Mapping Bragg scatter with a polarimetric WSR-88D. J Atmos Ocean Tech 28(10):1273–1285, DOI 10.1175/JTECH-D-10-05048.1

Moninger WR, Benjamin SG, Jamison BD, Schlatter TW, Smith TL, Szoke EJ (2010) Evaluation of regional aircraft observations using TAMDAR. Wea Forecasting 25(2):627–645, DOI 10.1175/2009WAF2222321.1

National Academies of Sciences, Engineering, and Medicine (2018) Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. The National Academies Press, Washington, DC

National Research Council (2009) Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks. The National Academies Press, Washington, DC

National Research Council (2010) When weather matters: Science and services to meet critical societal needs. National Academies Press

NCEI (2020) Storm Data. National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce

Newsom R, Turner DD, Lehtinen R, Münkel C, Kallio J, Roininen R (2020) Evaluation of a compact broadband differential absorption lidar for routine water vapor profiling in the atmospheric boundary layer. J Atmos Ocean Tech 37(1):47–65

Newsom RK, Turner DD, Mielke B, Clayton M, Ferrare R, Sivaraman C (2009) Simultaneous analog and photon counting detection for Raman lidar. Appl Opt 48(20):3903–3914

Newsom RK, Turner DD, Goldsmith JEM (2013) Long-term evaluation of temperature profiles measured by an operational Raman lidar. J Atmos Ocean Tech 30(8):1616–1634, DOI 10.1175/JTECH-D-12-00138.1

Otkin JA, Hartung DC, Turner DD, Petersen RA, Feltz WF, Janzon E (2011) Assimilation of surface-based boundary layer profiler observations during a cool-season weather event using an observing system simulation experiment. part I: Analysis impact. Mon Wea Rev 139(8):2309–2326

Parsons D, Dabberdt W, Cole H, Hock T, Martin C, Barrett AL, Miller E, Spowart M, Howard M, Ecklund W, et al. (1994) The integrated sounding system: Description and preliminary observations from TOGA COARE. Bulletin of the American Meteorological Society 75(4):553–568

Roininen R, Münkel C (2016) Results from continuous atmospheric boundary layer humidity profiling with a compact DIAL instrument. In: Proceedings of the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation, Madrid, Spain, pp 27–30

Segales AR, Greene BR, Bell TM, Doyle W, Martin JJ, Pillar-Little EA, Chilson PB (2020) The coptersonde: an insight into the development of a smart unmanned aircraft system for atmospheric boundary layer research. Atmos Meas Tech 13(5):2833–2848, DOI 10.5194/amt-13-2833-2020

Sisterson D, Peppler R, Cress T, Lamb P, Turner D (2016) The ARM southern great plains (SGP) site. Meteorological Monographs 57:6–1

Spuler S, Weckwerth T, Repasky K, Hayman M, Nehrir A (2016) Testing and validation of a micro-pulse, differential absorption lidar (DIAL) for measuring the spatial and temporal distribution of water vapor in the lower atmosphere. In: Optics and Photonics for Energy and the Environment, Optical Society of America, pp EW3A–5

Stull RB (2012) An introduction to boundary layer meteorology, vol 13. Springer Science & Business Media

Trenberth KE (1999) Atmospheric moisture recycling: Role of advection and local evaporation. J Climate 12(5):1368–1381

Turner D, Löhnert U (2014) Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based atmospheric emitted radiance interferometer (AERI). J of Appl Meteorol Climat 53(3):752–771

Turner D, Goldsmith J, Ferrare R (2016) Development and applications of the ARM Raman lidar. Meteorological Monographs 57:18–1

Turner DD, Blumberg WG (2019) Improvements to the AERIoe thermodynamic profile retrieval algorithm. IEEE J Sel Top Appl Earth Obs Remote Sense 12(5):1339–1354

Turner DD, Goldsmith J (1999) Twenty-four-hour Raman lidar water vapor measurements during the Atmospheric Radiation Measurement Program's 1996 and 1997 water vapor intensive observation periods. J Atmos Ocean Tech 16(8):1062–1076

Turner DD, Löhnert U (2020) Ground-based temperature and humidity profiling: Combining active and passive remote sensors. Atmos Meas Techn submitted

Turner DD, Knuteson RO, Revercomb HE, Lo C, Dedecker RG (2006) Noise reduction of atmospheric emitted radiance interferometer (AERI) observations using principal component analysis. J Atmos Ocean Tech 23(9):1223–1238, DOI 10.1175/JTECH1906.1

Wagner TJ, Klein PM, Turner DD (2019) A new generation of ground-based mobile platforms for active and passive profiling of the boundary layer. Bull Amer Meteorol Soc 100(1):137–153

Wakefield RA, Turner DD, Basara JB (2021) Evaluation of a land-atmosphere coupling metric computed from a ground-based infrared interferometer. Journal of Hydrometeorology DOI 10.1175/JHM-D-20-0303.1

Weckwerth TM, Weber KJ, Turner DD, Spuler SM (2016) Validation of a water vapor micropulse differential absorption lidar (DIAL). J Atmos Ocean Tech 33(11):2353–2372

Wei J, Su H, Yang ZL (2016) Impact of moisture flux convergence and soil moisture on precipitation: a case study for the southern United States with implications for the globe. Climate Dynamics 46(1-2):467–481

Wingo SM, Knupp KR (2015) Multi-platform observations characterizing the afternoon-to-evening transition of the planetary boundary layer in northern Alabama, USA. Boundary-Layer Meteorology 155(1):29–53

Wulfmeyer V (1999) Investigation of turbulent processes in the lower troposphere with water vapor DIAL and radar–RASS. J Atmos Sci 56(8):1055–1076

Wulfmeyer V, Hardesty RM, Turner DD, Behrendt A, Cadeddu MP, Di Girolamo P, Schlüssel P, Van Baelen J, Zus F (2015) 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. RRev Geophys 53(3):819–895, DOI 10.1002/2014RG000476

Zhang Y, Li D, Lin Z, Santanello Jr JA, Gao Z (2019) Development and evaluation of a long-term data record of planetary boundary layer profiles from aircraft meteorological reports. J Geophys Res Atmos 124(4):2008–2030