Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2022-235-RC1, 2022

Comment on amt-2022-235

Anonymous Referee #2

Referee comment on "Estimates of the spatially complete, observational-data-driven planetary boundary layer height over the contiguous United States" by Zolal Ayazpour et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-235-RC1, 2022

Overall Recomendation

This is a well written and presented manuscript that describes a novel approach to estimate PBL height (PBLH) over the CONUS region that extends discrete aircraft (AMDAR) measurements spatially using machine learning and the inherent relationships between PBLH and other land-atmosphere variables within reanalysis products (ERA5). There is a strong community need for improved PBL products across space and time, including PBLH, that extend beyond traditional (synoptic radiosonde, field campaign, or limited spaceborne) observations. Capturing the diurnal cycle in terms of PBL evolution is also a high priority for the scientific community with many potential applications (air quality, NWP, etc.). The approach seems to work well overall, from a large-scale, seasonal perspective when compared with reanalysis and observations (spaceborne and aircraft). This manuscript will ultimately be valuable to the community, and the work is appropriate for this journal and audience. I do have some major concerns and suggestions related to addressing the uncertainties in some of the PBLH products, and also including a focus on sub-seasonal variability (day-day) in the profiles and PBLH estimates.

Major Comments

The main issue and limitation of this work is in the validation of this new PBLH product. Because it is novel, and there is such a need, there are not any observationally-based 'truth' products to wok with that don't each have their own large uncertainties and limitaitons. CALIPSO is used here as 'truth', but has significant issues in terms of estimating an automated PBLH products (dependent on PBL regime, signal to noise, elevated aerosol gradients, clouds, etc.). So the anlayses and intercomparisons are more relative than absolute, in terms of comparing the new AMDAR/ML PBLH vs. models (reanalyses) vs. observations (CALIPSO and aircraft-based lidar). As a sanity check, this comparison is useful, but it does not sufficiently reflect whether the new PBLH product is accurate on day-day or diurnal timescales.

Another concern is related to the diversity in the PBLH estimation techniques used in all these products that are being evaluated. L45 (paragraph). Can the authors say something about the methods used in these comparisons? Using CALIPSO implies that PBLH is based on aersol backscatter gradients, which is quite distinct from what each model PBLH is based upon thermodynamically (not to mention their respective PBLH approach that is mentioned earlier). Neither is actually 'wrong' or 'right', as they are looking for the top of the mixed layer or the top of the PBL turbulence, or T and RH gradients. The conclusions presented here suggest that CALIPSO aerosol-based PBLH is the 'truth' but also AMDAR-based thermodynamic PBLH is the 'truth', and both cannot be the case. These are relative intercomparisons that show that the models deviate from other estimates, but has it been shown when looking at T and q profiles that the models actually do 'overestimate' the true PBLH?

To reiterate the comment about the diversity of PBLH methods in the models and observations, what are the implicatons of comparing PBLH derived in 6 different ways (AMDAR, ERA, NARR, M2, plus CALIPSO, and HSRL)? There are tendencies from each method (TKE, RiB, etc.) in terms of the PBLH they capture and under what regimes they perform well/poorly. None is perfect, but comparing across all of them is problematic in a blanket sense.

Fig. 6: It would be nice to see a more detailed, nuanced analysis zooming into regions, and sub-seasons (even day-day). This generally looks like a good result for XGB, but as we know a lot of important variability and biases can be masked out on the seasonal timescale. Could the authors provide this even if in supplemental form?

Section 5: Given the diversity in models, observations, and PBLH estimation approaches discussed above, it would be helpful to include a more direct analysis of individual profiles (not just seasonal composites, or CONUS evaluations). Examining individual profiles would enable a direct comparison across all of these, and a visual aid to actually look at T, q, and aerosol profiles and their estimated PBLH. This would provide insight as to their behavior, and also provide to the reader a 2D vertical perspective of what this paper is all about and how much these can differ. The challenge is in selecting/sampling the locations and times, but that could be done in a single figure with multiple panels, regions, etc. after the authors perform a search of different locations and regimes that they feel are representative. No overall conclusions would be made based on these, but it likely would provide the insight and demonstrate the variability to the reader. Might also expose what CALIPSO is doing.

Minor Comments

Intro: Strong aerosol and AQ motivation here in terms of PBLH. The authors could also mention importance of PBLH for convection, shallow to deep convection, LCL deficit, etc. in terms of the thermodynamic pathways and feedbacks (ultimately on precipitation). Also the PBL mediation of surface fluxes (H and LE), soil moisture, vegetation, entrainment feedbacks. There is a big component of PBLH importance that isn't discussed here and would make the general impact of this work and dataset more robust.

L36: Another reference that looked at PBLH in M2, NARR, and CFSR: https://doi.org/10.1175/JCLI-D-14-00680.1

L79: Also potential applicability to drifting orbits of existing satellites (e.g. EOS) that cover other parts of the diurnal cycle.

L99: Many models and applicaitons have used 0.25. Can the authors justify using 0.5 here? A direct comparison of the results when using 0.5 vs. 0.25 seems important, given the sensitivty to the assumption and the impact on the ultimate product of PBLH.

Section 2.3: For CALIPSO, as McGrath-Spangler has shown, it is very difficult to automate PBLH retrieval due to elevated aerosols, clouds, signal to noise, etc. such that they were only able to confidenly produce seasonal climatologies. Because you are estimating your own CALIPSO-based PBLH here on a daily basis, there is likely much greater uncertainty due to these factors on the day-day level.

Fig. 5: I'm glad the authors included this detailed assessment of the variables more influential in the ML training. This is where the L-A interaction component becomes scientifically interesting and can be learned from. Fig. 5 makes sense (to me) overall in terms of which variables most impact/drive CBL growth, focused on surface heating and buoyancy (as well as the diurnal pattern/memory of the PBL growth itself). Identifying which is most land-driven (vs. water/ocean) makes sense as well. Given the focus of soil moisture in L-A interaction research, It would have been interesting to include soil moisture itself, given its role in controlling the Bowen ratio and surface heating (which are strongly infliential as seen here). Also, it is interesting that LHF doesn't emerge along with H. Using EF or Bowen ratio might have been a strong as well if included.

XGB Training: Did the authors test the sensitivity of the training result to the dataset used (ERA5 vs. M2, for example)? As presented, the results are strongly dependent on the relationship of AMDAR vs. ERA5 in terms of PBLH.

L319: 'Complement' CALIPSO(?), or is the HSRL superior in terms of identifying a robust PBLH? Does this put the quality of CALIPSO PBLH in question? There may be something to say here about both estimates being based on aerosol backscatter gradients, but yielding such different results vs. XGB.

Section 4.3: This is also interesting, in that the HSRL is based on aersol backscatter, but the spirals are manually based on other variables, yet yield similar results that are quite different from CALIPSO. Again, I would suggesting considering more rigorously the uncertainty in the CALIPSO estimates.

Conclusions: Rather short conclusions that could offer more applicability to the community, and what this dataset could be used for (and how confidently), as well as what new PBL profile observations would be most valuable in the future (e.g. spaceborne or more routine suborbital measurements) to reduce the uncertainty in the estimates applied here.