

RC2:

Review of the article titled “Machine Learning model for inverting convective boundary layer height with implicit physical constraints and its multi-site applicability” by Chu and coauthors for publication in Atmos. Chem. Phys.

1. The authors have used boundary layer (BL) height from the doppler lidar (DL), surface fluxes from eddy correlation (ECOR) and energy balance Bowen Ratio (EBBR), and thermodynamic stability from Atmospheric Emitted Radiance Interferometer (AERI) to construct a machine learning (ML) model for predicting BL height. The data from ARM SGP site, and other ancillary sites around SGP have been used. The main premise of the paper is using the off-the-shelf interfaces like TPOT and AutoKeras for training and validation, thereby leaving AI to pick the ML model. After model identification, the authors have applied the model to predict BL height over different seasons and different sites. The article is relatively well-written and fits the journals scope. However, I believe that the article lacks physical depth, and could be improved. Much of the discussion is on simply adapting the data for TPOT and AutoKeras, which is not novel. The paper is also too long at this point, some of the discussion is more suitable for a dissertation rather than a paper. So, mentioning few things below that can improve the article further.

Reply: Thank you for your thorough review and constructive feedback. We revised the paper accordingly, which have significantly improved the manuscript's quality and alignment with ACP's submission standards. We acknowledge the concern regarding the lack of physical depth and the manuscript's length. To address this, we have streamlined the discussion by focusing on the novel aspects of the ML model application and moved less critical content, such as the introduction to machine learning and EBBR-related details, to the supplementary materials. Discussions on machine learning-related content are intended to facilitate better understanding for readers unfamiliar with machine learning. This revision enhances conciseness and clarity while maintaining the paper's focus on the use of the machine learning model for CBLH prediction across seasons and sites.

2. I like that you are trying to use some physical constraints as input parameters to improve the ML model. LTS, time, and sun parameters are a good start (Line 345). However, through previous research it has been shown that presence of elevated humidity above the BL can also affect the BL development through radiative effects, and same is true for high-level clouds. These effects in some part will be reflected in the ECOR fluxes, but with a time delay. In addition, wind speed, wind direction, wind shear and wind veer have also been shown to be very important. So maybe you

can include the following parameters in your input models, as they are also available at the ARM sites: wind speed, wind direction, wind shear, wind veer, surface upwelling and downwelling longwave and shortwave radiation. Surface meteorological variables will also be good to include. I understand that including cloud properties might be hard, but given the strong expertise of authors Deng, Xue and Wang, they can include ceilometer cloud fraction and base height in it. This might significantly improve the model, and the shapely analysis will tell which parameters are important. Thank you.

Reply: Thank you for your positive feedback and valuable suggestions. We agree that incorporating additional parameters such as wind speed, wind direction, wind shear, wind veer, surface longwave and shortwave radiation, and ceilometer-derived cloud fraction and base height could enhance the model's performance. This paper focused on measurements readily available from both observations and model simulations. We have revised the manuscript's conclusion to reflect these future research goals and will use SHAP analysis to assess their importance. Thank you for your insightful recommendations.

3. The authors have used AutoKeras and TPOT for selecting the best model, which is great. Given the small amount of data used in this work, the tree-based models could also be employed in TPOT. It will be good if you can tell us what model and the associated hyperparameters was picked by these two frameworks. I cannot tell if they training was done online, and if so the batch sizes etc. The hyperparameters then should be scrutinized to understand whether any more improvement can be made. On the same topic, it will be good if the authors can mention if they explicitly or the two frameworks implicitly regularized (normalized, bias correct etc.) the input parameters. I assume the (Line 345) sunrise and sunset times and time variable could be normalized by 24.

Reply: Thank you for your insightful and valuable feedback. In the revised manuscript, we have detailed the models selected by AutoKeras and TPOT, along with their associated hyperparameters, in Section 3.4. All training was conducted locally, and we have included a detailed description of the hardware environment in Section 2.3. Regarding input parameter regularization, both frameworks implicitly normalize inputs, ensuring that explicit normalization of sunrise, sunset, and time variables by 24 does not affect the results. We have clarified this in the revised manuscript in in Section 2.5 (section 2.6 in the first manuscript).

4. Figure 4 onwards: these are nice figures, but maybe you can add another panel showing the time evolution of the difference between DL CBLH and predicted CBLH. This will truly tell if the model accurately captures the daytime evolution.

Reply: Thank you for the excellent suggestion. Accordingly, we have incorporated figures illustrating the temporal evolution of the discrepancies between the DL-derived CBLH and the predicted CBLH into Figure 4c and Supplementary Figure S5.

5. Figure 10: The connection between convective boundary layer and surface fluxes is clear during summer months, but I cannot tell how it works in winter months. Can you please elaborate on the number of samples going into this figure, especially for the colder seasons. It is difficult to assess as to how accurate the surface fluxes might be when the environment is cold and the surface is frozen. Otherwise, you can control for it by using temperature and advection in your input parameters. Thank you.

Reply: Thank you for your insightful and valuable feedback. In Figure 10 (now Figure 8 in the revised manuscript), we have specified the winter (DJF) sample size as 240 points, one-third of the summer sample size (720 points). Winter performance is indeed the weakest, with an R^2 of 0.736. We agree that assessing surface flux accuracy in cold, frozen conditions is challenging, but for clear and scattering cloud cases used in this study, the monthly mean sensible and latent heat fluxes are still up to ~ 200 and ~ 100 W/m^2 in the winter (DJF). The consistent temporal evolutions of convective boundary layer and fluxes are still clear in winter at the SGP site. We have incorporated this discussion in Section 3.5.2 of the revised manuscript. In future work, we plan to include temperature and advection as input parameters to improve model performance.

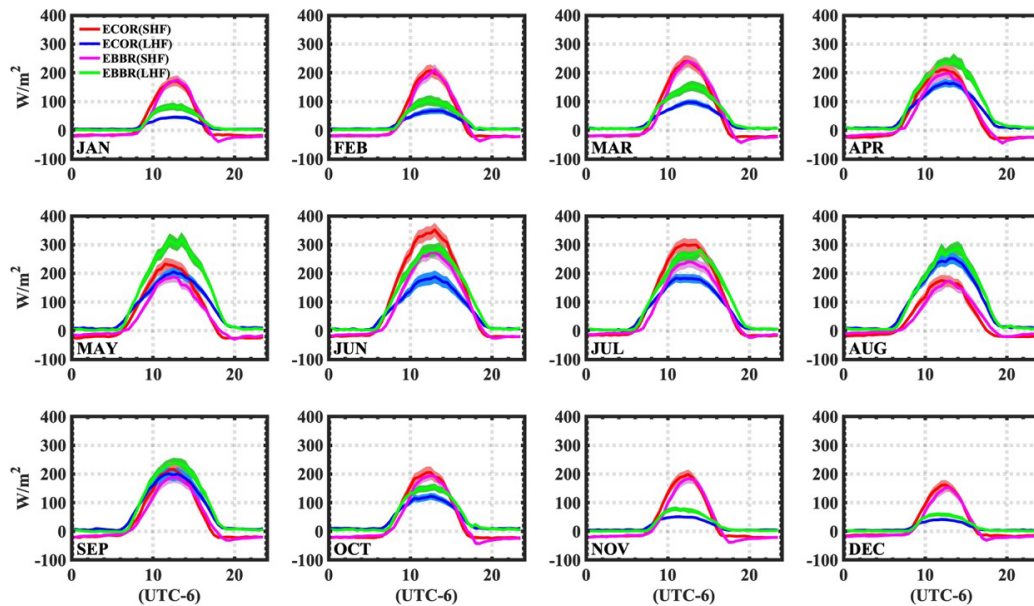


Figure: Monthly mean temporal variations of sensible heat flux (SHF) and latent heat flux (LHF) at site C1. E13 (EBBR: Energy Balance Bowen Ratio system, for measuring sensible and latent heat fluxes via the Bowen ratio method) and E14 (ECOR: Eddy Correlation system, for measuring turbulent fluxes via the eddy covariance method) are both located within 1 km of the C1 site and can thus be regarded as co-located with C1.