This paper uses roughly 1-year of surface-based doppler lidar wind observations collected at two locations off the west coast of California (Morro Bay and Humboldt Bay) to evaluate the 3-hourly wind speed forecasts from the High-Resolution Rapid Refresh (HRRR) numerical weather prediction model at several different heights from 40 to 240 m above the sea surface. They have demonstrated that overall, the HRRR's performance matches the observations pretty well, with better agreement at Morro Bay vs Humboldt Bay, but that there are conditions where the model has significant wind speed errors.

Generally speaking, this paper reads very well. The figures are clean, the writing is clear, the results are largely well supported. I have listed a number of comments / suggestions below that I'd like the authors to consider and address before the paper could be accepted for publication.

We would like to sincerely thank you as a reviewer for your thoughtful and constructive comments. Your insights have greatly helped us improve the clarity and overall quality of the manuscript. We have carefully addressed all suggestions in blue and believe the revised version benefits substantially from your valuable feedback.

Major comments:

Line 102: what is the accuracy and uncertainty of the GOES derived cloud top heights? Is there any seasonal dependency to these values?

GOES measures cloud-top temperature (CTT) with an accuracy of approximately 1 K, and CTH is then inferred from lapse-rate calculations following the method of Sun-Mack et al. (2014). The bias in CTT retrievals was assessed by Yu et al. (2012), who showed that while CTTs are generally accurate, uncertainties in the lapse rate often translate into larger errors in derived CTHs. Several approaches have attempted to mitigate this issue. Zuidema et al. (2009) used sea-surface temperature (SST)–based constraints to improve CTH estimation, but this approach also showed mixed performance. A comparison study by Ghate et al. (2019) demonstrated that all major CTH retrieval techniques perform well under certain conditions and fail under others, with no clear gold standard. The most recent SATCORPS Edition 4 GOES algorithm represents the best-performing version to date; Yost et al. (2021) found mean GOES–CloudSat/CALIOP CTH differences of about 100 m.

In our study, the mean HRRR–GOES CTH differences (~150–200 m) are somewhat larger than the expected GOES uncertainty, suggesting that the discrepancy likely reflects HRRR deficiencies rather than retrieval error. Because stratocumulus clouds off the California coast are typically thin (< 200 m) and the boundary-layer lapse rate closely

follows the dry adiabat, a 1 K uncertainty in retrieved CTT corresponds to roughly 100 m in CTH uncertainty—consistent with Yost et al. (2021).

Finally, as noted in Mitra et al. (2025), a seven-year comparison between GOES and ERA5 cloudy boundary-layer heights indicated ERA5–GOES CTH differences of −0.17 ± 0.62 km (Humboldt) and −0.22 ± 0.51 km (Morro Bay). Together, these findings suggest that while GOES retrievals have inherent uncertainties, their magnitude is small compared with the systematic HRRR biases discussed in this study. As shown in Mitra et al. (2021), a seasonal dependence in GOES CTH uncertainty is possible, but its effect is minor relative to these overall differences.

We added the following paragraph in Section 2.2 to provide readers information about uncertainty and seasonality related to the GOES cloud retrievals:

GOES determines cloud-top height (CTH) from infrared-retrieved cloud-top temperature (CTT) and lapse-rate calculations following Sun-Mack et al. (2014). The CTT measurements have an accuracy of about 1 K (Yu et al., 2012), which corresponds to an uncertainty of roughly 100 m in CTH for a dry-adiabatic lapse rate. Previous studies have shown that inaccuracies in lapse-rate estimates can lead to CTH errors of this magnitude (Zuidema et al., 2009; Ghate et al., 2019). Comparisons of the most recent SATCORPS Edition 4 GOES retrievals with CloudSat and CALIOP indicate mean differences near 100 m (Yost et al., 2021). A recent seven-year analysis comparing ERA5 and GOES data reported ERA5–GOES differences of −0.17 ± 0.62 km (Humboldt) and −0.22 ± 0.51 km (Morro Bay) (Mitra et al., 2025). These results imply that GOES retrieval uncertainties are on the order of 100–200 m—smaller than the HRRR–GOES differences reported here—and that the HRRR bias in CTH likely reflects model rather than retrieval limitations.

Sun-Mack, S., Minnis, P., Chen, Y., Kato, S., Yi, Y., Gibson, S.C., Heck, P.W. and Winker, D.M. (2014) Regional apparent boundary layer lapse rates determined from CALIPSO and MODIS data for cloud-height determination. Journal of Applied Meteorology and Climatology, 54, 990–1011. https://doi.org/10.1175/JAMC-D-13.081.1.

Yu, F., Wu, X., Raja, M.K.R.V., Li, Y., Wang, L. and Goldberg, M.D. (2013) Diurnal and scan angle variations in the calibration of GOES imager infrared channels. IEEE Transactions on Geoscience and Remote Sensing, 51, 671–683

Zuidema, P., Painemal, D., DeZoeke, S.P. and Fairall, C.W. (2009) Stratocumulus cloud-top height estimates and their climatic implications. Journal of Climate, 22, 4652–4666.

C. R. Yost, P. Minnis, S. Sun-Mack, Y. Chen and W. L. Smith, "CERES MODIS Cloud Product Retrievals for Edition 4—Part II: Comparisons to CloudSat and CALIPSO," in

IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 5, pp. 3695-3724, May 2021, doi: 10.1109/TGRS.2020.3015155.

Mitra, A., Ghate, V., Krishnamurthy, R.: Wind and Climate Variability within the Californian Offshore Wind Energy Areas, Argonne Technical Report , ANL/NSE-24/26, Pending Release at U.S. Department of Energy (DOE) OSTI, https://doi.org/10.2172/2568064, 2025

Mitra, A., Di Girolamo, L., Hong, Y., Zhan, Y., & Mueller, K. J. (2021). Assessment and error analysis of Terra-MODIS and MISR cloud-top heights through comparison with ISS-CATS lidar. Journal of Geophysical Research: Atmospheres, 126(9), e2020JD034281. https://doi.org/10.1029/2020JD034281

Ghate, V. P., Mechem, D. B., Cadeddu, M. P., Eloranta, E. W., Jensen, M. P., Nordeen, M. L., and Smith, W. L.: Estimates of entrainment in closed cellular marine stratocumulus clouds from the MAGIC field campaign, Q. J. Roy. Meteor. Soc., 145, 1589–1602, https://doi.org/10.1002/qj.3514, 2019

Line 102: how are multi-layer cloud systems handled? Certainly there are cases where there is an upper level cirrus cloud that obscures the lower atmosphere?

In cases of multilayer cloud systems (e.g., thin cirrus overlying thicker low clouds), the GOES infrared retrieval typically reports a cloud-top height (CTH) corresponding to the upper cirrus layer. Such retrievals yield CTHs higher than our low-cloud thresholds and are therefore automatically excluded from our analysis as false “mid-level” or “high-cloud” detections. We have confirmed that these filtering criteria effectively remove most multilayer cases from the dataset.

We acknowledge that in some situations, cirrus clouds may be optically thick enough to completely obscure the lower cloud layer, preventing detection by GOES infrared channels. However, these high-cloud retrieval periods are not used to draw conclusions about the marine boundary layer or to evaluate HRRR wind-speed biases, as our analysis focuses solely on periods classified as low-cloud or clear-sky conditions.

To incorporate this information, in the revised manuscript section 2.2, we added the following paragraph:

Instances of multilayer cloud systems (e.g., thin cirrus overlying low stratocumulus) were handled by filtering out retrievals with cloud-top heights exceeding our low-cloud thresholds. In such cases, the GOES infrared retrieval reports the upper-level cirrus as the primary cloud top, leading to spuriously high CTH values that are automatically excluded from the analysis. Situations with optically thick cirrus that fully obscure the

low-level cloud deck are likewise classified as high-cloud conditions and are not used when evaluating boundary-layer properties or HRRR wind-speed biases.

Line 111 regarding your "minor discrepancies" comment: could this be a definition and/or sensitivity issue? For example, the GOES can only identify a cloud layer if the optical depth is above some threshold (which might depend on other atmospheric conditions), whereas the model might define a cloud if there are any hydrometeors in a volume. How was this handled? Was an instrument simulator applied to the HRRR's output to try to mimic the GOES observations? (I suspect not, but you should at least be clear that this is an issue that could impact these analyses).

Thank you for raising this question. The minor discrepancies between the GOES- and HRRR-derived cloud-top heights and cloud fractions may indeed partly reflect differences in how clouds are defined and detected in each system. No satellite observation simulator (e.g., an instrument forward operator) was applied to the HRRR output in this study. Therefore, some portion of the observed differences between HRRR and GOES results likely arises from these definitional and sensitivity differences. We have revised the manuscript to explicitly acknowledge this as a potential source of uncertainty in the cloud comparisons. The added sentences are shown below:

These differences may partly reflect the distinct definitions of cloud used in the two datasets. The GOES retrieval identifies a cloud layer only when the optical depth exceeds a threshold that varies with viewing geometry and atmospheric conditions, while HRRR defines clouds based on the simulated presence of hydrometeors within model grid cells. Because no satellite instrument simulator was applied to the HRRR output, such definition and sensitivity differences might contribute to the observed discrepancies in cloud-top height and cloud fraction.

Line 134: One of the limitations of this study is that the liquid water path (LWP) cannot be compared between the GOES and model easily. My sense is that the model has too little LWP in its cloud, and thus in the daytime there is too little absorption of shortwave radiation, and thus too little diabatic heating in the cloud, which is why the cloud did not deepen. Your suggestion that it could be an issue with longwave radiative cooling is also possible, but again, this connects to errors in getting the diurnal evolution of the LWP in the cloud correct. So please expand this discussion a bit more.

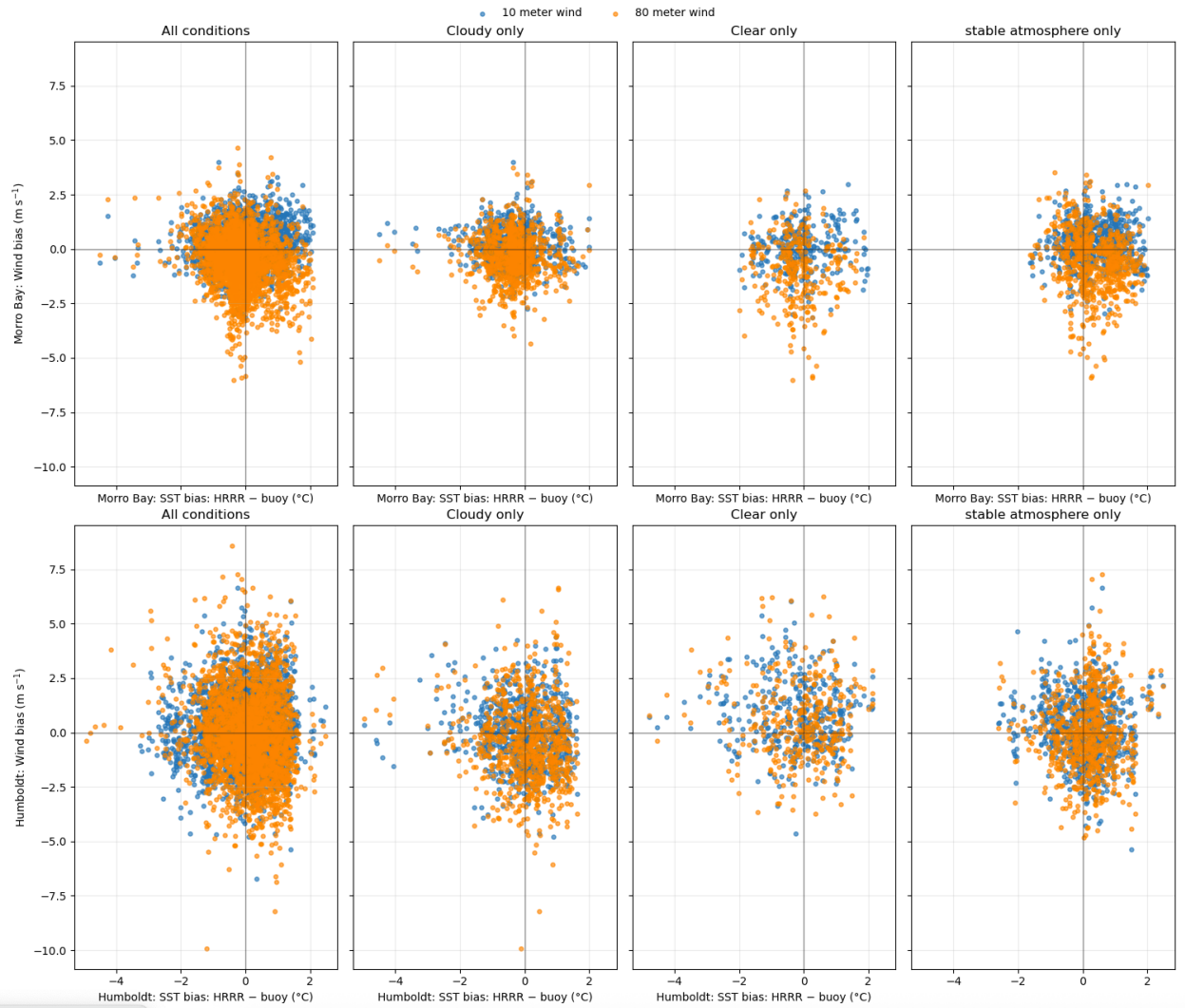
Thank you for raising this question. We agree that one limitation of our study is the lack of direct comparison of liquid water path (LWP) between GOES and HRRR, which constrains our ability to diagnose the model's representation of cloud radiative processes. The absence of a diurnal cycle in HRRR cloud-top height may indeed result from insufficient simulated LWP. A low LWP would reduce daytime shortwave

absorption within the cloud, leading to too little diabatic heating and limiting cloud deepening, while also affecting longwave cooling at night. Therefore, both underestimation of daytime heating and misrepresentation of nocturnal longwave radiative cooling could contribute to the weak diurnal variability of simulated cloud-top height. We have expanded the discussion in Section 3.1 to reflect this broader interpretation and to acknowledge that errors in HRRR's representation of the diurnal LWP cycle likely underlie its muted diurnal cloud-top height variability. The added sentences in the revised manuscript are shown below:

One likely contributing factor is the model's underrepresentation of cloud liquid water path (LWP), which cannot be directly compared with GOES in this study. A low simulated LWP would reduce the amount of shortwave absorption within the cloud during the daytime, resulting in insufficient diabatic heating and limited cloud deepening. Conversely, too little condensate would also weaken longwave cooling at night, reducing turbulent mixing at cloud top. These combined effects could explain the muted diurnal variation in HRRR cloud-top height and indicate that biases in the model's treatment of cloud microphysics and radiative processes jointly contribute to the discrepancy.

Line 230: You are hypothesizing, here and later in the paper, that the model is not capturing stability properly and that could be the source of the bias. However, error in the sea surface temperature (SST), which is a boundary condition for the model, is another possibility. And it is possible that there are different errors in SST in Morro vs Humboldt Bays, especially in clear skies? That would affect the stability profile, which would then feed back into the turbulence and wind profiles. I believe the buoys have SST observations on them, and it would be a pretty straight-forward analysis to determine if SST biases are correlated with the wind biases. This might result in your need to update the statement at line 350.

Thank you for the suggestion. We conducted additional analyses to examine whether SST biases are correlated with wind-speed biases. We evaluated scatter plots for all time periods, for cloudy conditions, for clear-sky conditions, and for cases with stable stratification (defined when the buoy-measured air–sea temperature difference, Tair – SST, is positive). In all cases, we found no clear correlation between SST bias and wind-speed bias, as shown below:



In the revised manuscript, we added this information as given below:

However, we did not find any meaningful correlation between SST bias and wind-speed bias when examining the relationship over the entire period, as well as separately for cloudy conditions, clear-sky conditions, and stable atmospheric cases.

Line 253: you state “similarity-based wind speed profile model”. The way this is worded suggests that the HRRR is using this approach. Similarly, in line 348, you give the same impression. The HRRR uses Monin-Obukhov theory in the surface layer; however, above the surface layer the HRRR uses an eddy diffusivity mass flux approach. See papers by Olson et al (BAMS 2019) and NOAA Tech Memo (https://doi.org/10.25923/n9wm-be49).

Thank you for your comment. We revised the sentences to avoid the confusion and is given as below in the revised manuscript:

As discussed by Optis et al. (2016) and references therein, similarity-based wind-speed profile formulations derived from Monin–Obukhov similarity theory (MOST) can exhibit substantial bias under strongly stratified boundary layers. Because HRRR employs surface-layer and turbulence parameterizations that draw from this theoretical framework, such limitations may contribute to the under- or overestimation of wind speeds observed under stable, clear-sky conditions in our analysis. It should be noted, however, that HRRR applies MOST primarily within the surface layer, while turbulent mixing above this layer is represented using an eddy-diffusivity mass-flux (EDMF) approach that accounts for nonlocal transport and convective plumes (Olson et al., 2019a,2019b).

Figure 8 and the discussion starting at line 299: There are both resolved clouds (i.e., where the entire model grid cell is cloudy) and subgrid-scale clouds (i.e., where there is partial cloudiness in a grid cell). Unfortunately, the HRRR does not save the liquid water content operationally to an output file. Thus, this figure can only show the resolved cloud liquid clouds. It is possible that there is significant subgrid-scale cloud liquid between the "gaps" shown in Fig 8a and discussed at line 305, but we just don't know for sure. You can get the LWP from the HRRR from the HRRR data archive on the AWS cloud server; the LWP is the vertical integral of both the resolved and subgrid-scale liquid water content profiles.

Thank you for bringing up this point. We attempted to extract the liquid water path (LWP) from the HRRR output available to us but were unable to locate this variable in our dataset. If you could provide additional guidance on how to retrieve LWP from the HRRR archive, we would be glad to explore this further for the revision.

In the meantime, we have added the following sentences to the manuscript to clarify the limitations associated with using the cloud water mixing ratio variable:

It is important to note that Figure 8 shows only the resolved cloud liquid water mixing ratio fields from HRRR. The operational HRRR output does not include subgrid-scale cloud liquid water, and therefore the figure cannot represent partially cloudy grid cells where unresolved cloud condensate may be present. As a result, some of the apparent "gaps" in the HRRR cloud field may in fact contain subgrid-scale cloud liquid that is not captured in the resolved output.

Minor comments:

Line 13: "multiple instruments"

Thank you. I corrected it

Line 14: there are many types of lidar. Please say “including Doppler lidar”

Thank you. I followed your suggestion

Line 54: …resource assessments over the United States. (the HRRR’s domain)

Thank you. I added “over the United States”

Line 80: Isn’t GOES-17 the only satellite that is relevant for this paper?

Yes, since we are only using the time frame of the DOE buoys. That line was in reference to the entire SATCORPS dataset, not the subset of the data used in question. To clarify the confusion, we added a following sentence in the revised manuscript:

For this study, GOES-17 is the only satellite providing data that overlaps with the buoy observation period.

Line 130: “in both locations” --> “in either location”

Thank you. I corrected it.

Line 131: I believe you have this reversed, as the GOES shows higher values in the daytime in Fig 2

Thank you for raising the question. The sentence was meant for the typical marine stratocumulus situation, which we do not see in our GOES analysis. To avoid the confusion, I revised the sentence as shown below:

Under typical marine stratocumulus conditions, cloud tops are higher at night and lower during the day due to stronger nocturnal cloud-top radiative cooling, our GOES analysis (Fig. 2) shows the opposite behavior at both sites; the reason for this discrepancy is presently unclear.

Figure 3: please use the same y-axis range for all panels within the figure, as it would make it much easier to compare the different panels

Thank you for the suggestion. We now have the same y-axis panels for all panels in the revised manuscript. We also changed the color scheme as it previously didn’t pass the color blindness test. Here is the remade figure 3.

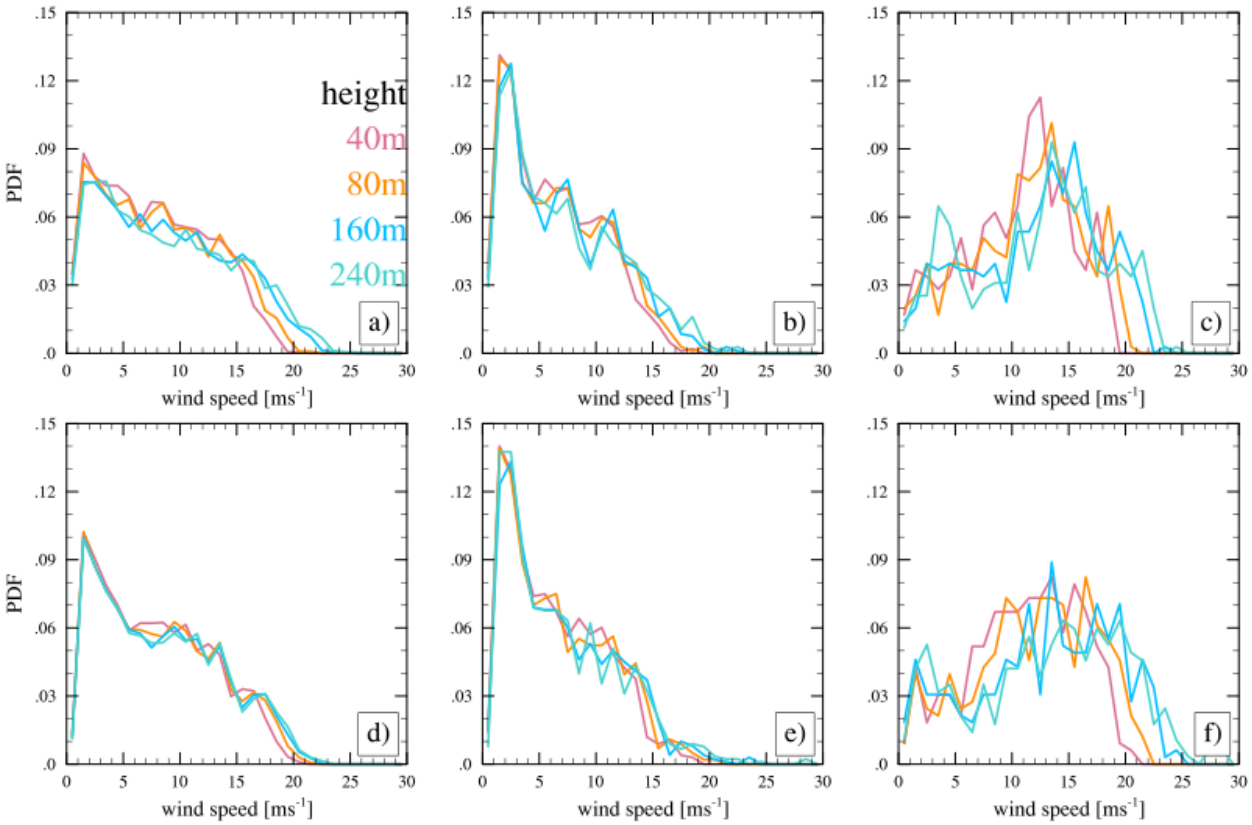
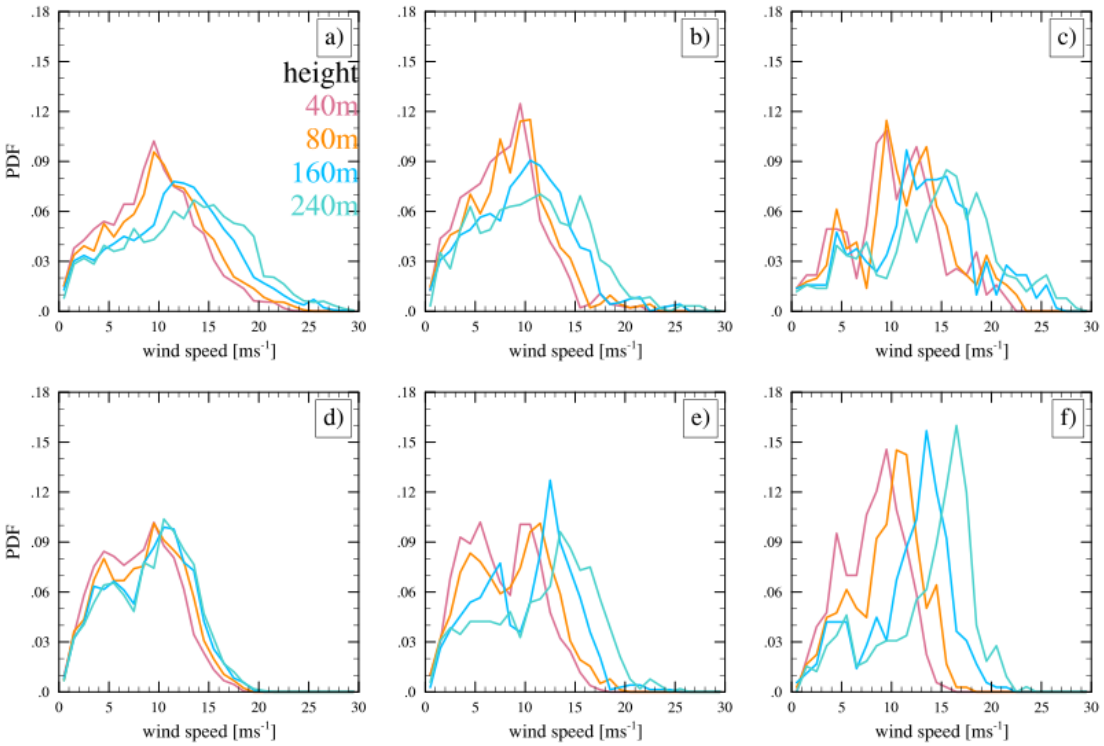


Figure 4: same comment

Thank you for the suggestion. We now have the same y-axis panels for all panels in the revised manuscript. We also changed the color scheme as it previously didn't pass the color blindness test. Here is the remade figure 4.



Line 173: “overestimates wind shear” – this is not supported by Table 2, which shows the opposite

Thank you for raising this question. We acknowledge the discrepancy in the behavior of wind shear, which is qualitatively evident in Fig. 4 and quantitatively summarized in Table 2. We believe this difference arises from the way the data were processed to derive the PDFs and the shear exponent. For the shear exponent, we used time-averaged wind speeds, whereas the PDFs include the full range of instantaneous wind speeds within each weather regime. To avoid confusion, we have added the following clarification to the manuscript:

However, when the shear exponent is computed directly from the mean wind speeds (Table 2), HRRR slightly underestimates the average shear compared to the buoy data. This could be because that PDFs are dominated by high-wind events, which amplify apparent vertical gradients, while the mean shear exponent integrates over all wind-speed regimes and may therefore show weaker shear on average.

Line 214: “model consistently underestimates” – I disagree. Fig 5 top panel shows that the HRRR is essentially unbiased for wind speeds less than 10 m/s at Morro Bay

Thank you for raising this question. We revised the questioned sentence in the revised manuscript as given below:

At Morro Bay, HRRR is largely unbiased for wind speeds below 10 m s^{-1} but consistently underestimates the observed hub-height wind speed under cloudy conditions when wind speeds exceed this threshold, with the magnitude of underestimation increasing at higher wind speeds.

Figure 6: why didn't you use the same style of plot as in Figure 5? It would make the paper more consistent and easier to read

Figure 5 presents box-and-whisker plots of the wind-speed bias at 80 m under all, clear-sky, and cloudy conditions, while Figure 6 shows the mean wind-speed bias at multiple heights (40, 80, 160, and 240 m). We wanted to show how the mean wind bias changes with height in Figure 6.

Line 261: "decoupled during the day" – is this always true (i.e., for every cloudy day in your analysis)? I would be surprised if this was true

We agree that the statement about offshore boundary layers being "decoupled during the day" should not be interpreted as a universal condition. Our intention was to describe the general tendency in marine stratocumulus-topped boundary layers, where radiative cooling at the cloud top weakens during daytime, leading to partial decoupling between the surface and cloud layers. However, the degree of decoupling depends on local thermodynamic and dynamical conditions and may vary among cases. We have revised the text to clarify this point in the revised manuscript and the change is given below:

As a result, offshore boundary layers tend to become more decoupled during the day when cloud top cooling weakens and more coupled at night when radiative cooling is strongest. However, the degree of coupling can vary substantially depending on cloud optical thickness, synoptic forcing, and sea surface temperature gradients.

Line 275: the differences in the correlations between the two locations are important. Modify the end of this sentence to indicate the mean values of the correlation coefficient to help strengthen this point

Thank you for your suggestion. I modified the sentence in the revised manuscript and is given below:

Additionally, the correlation between the time series of lidar buoy observations and HRRR wind speed is notably weaker at Humboldt (mean correlation coefficient r = 0.70) than at Morro Bay (r = 0.93), highlighting regional differences in model performance.

Figure 7: the blue minus black lines in panes c and d do NOT equal the results shown in panels e and f. Please recompute and update the figure

Thank you for raising this question. We believe that the blue minus black lines in panels (c) and (d) differ from the results shown in panels (e) and (f) because we first composited the bias for each hour across the analysis period and then computed the composite mean for each hour. This approach may explain why the difference between the composite means is not identical to the mean of the composite differences. To avoid the confusion, we replotted (e) and (f) to have blue minus black lines in panel c and d. The new figure is in the revised manuscript and also given below.