We thank the reviewer for their thoughtful suggestions and constructive criticism that have helped us improve our manuscript. Below, we provide responses to reviewer concerns and suggestions in blue font.

Review of egusphere-2023-1943: "HSRL-2 Retrievals of Ocean Surface Wind Speeds" by Dmitrovic et al.

In this work the authors describe an algorithm for deriving ocean surface wind speed estimates from measurements of ocean surface backscatter acquired by the NASA-Langley high spectral resolution lidar (HSRL). The NASA HSRL can accurately characterize the signal attenuation above the ocean surface and reliably partition the surface signal pulse into pure surface and ocean subsurface components, and hence can deliver high quality measurements of surface integrated attenuated backscatter (β_{surf}). The accuracy of the wind speed retrieval thus depends on the equation relating β_{surf} to wave slope variances and the fidelity of the model used to convert wave slope variances to wind speeds. The authors derive wind speeds using two different models – the classic Cox-Munk (1954) and a lidar-specific model developed by Hu et al. (2008) – and compare these results to near-simultaneous dropsonde measurements of wind speeds. The paper is well organized and well written and its subject matter is entirely appropriate for Atmospheric Measurements Techniques. There are, however, several issues that should be addressed prior to publication.

My primary concern is that the authors' equation (2) fails to acknowledge the possible presence of whitecaps and/or sea foam. While the authors cite Josset et al., 2010b as their source for equation (2), that work explicitly includes reflection from the whitecap fraction within any footprint; see section 2.2 and equations (2) and (21) therein. Furthermore, comparing the upper and lower panels of figure (2) in Hu et al., 2008 suggests that omitting whitecap contributions could have a significant impact on the results derived in this paper. On the other hand, Lancaster et al. (2005) suggest that the "contribution of whitecaps to the nadir lidar measurements is seen in Figure 2 to be negligible".

You bring up an excellent point and as noted in Josset et al. (2010b) there needs to be more research on the effects of whitecaps. Josset notes, "There is also a need to better assess the large uncertainties associated with the lidar return of foam patches and their effect on subsurface lidar returns." Also, Hu et al. (2008) provided new insight to this contribution by using the integrated surface depolarization ratio to add an empirical correction to the surface scattering of whitecaps. The details of how this correction was determined were not provided and there are significant differences between our lidar and CALIPSO. These include the spot size of the beam on the surface (8 m compared to ~90 m for CALIPSO and HSRL-2 has data collected and stored over 100 laser shots (averaged due to x10 repetition rate of the laser), while CALIPSO records every shot. We note that the CALIPSO data presented by Hu et al. (2008) was averaged globally and thus were statistically dominated by clear water cases where the change in depolarization would better correlate with the whitecaps rather than the subsurface particulate scattering. Moreover, the data collected from ACTIVATE includes significant sampling along the coastal waters where the subsurface contribution from ocean particulates can significantly impact surface depolarization data. Lastly, the residual depolarization due to the different set of optics will likely change the empirical relationship compared to CALIPSO.

While recognizing that this correction is important at high wind speeds (> 10 m s^{-1}), we limited our discussion to the other two main factors in deriving wind speed using HSRL-2/1) the subsurface contribution due to the high vertical resolution to derive the ocean backscatter and 2) determining an accurate value of the atmospheric attenuation. It is of high interest to look at the surface depolarization, which we have calculated from HSRL-2 during these flights. There is a clear relationship between the surface depolarization with wind speed as expected, but the ocean subsurface contribution is also evident and correlates with increased scattering in the ocean. In addition, the minimum surface backscattered reflection could provide additional information on the average reflection of the whitecaps. For instance, looking at the 97th percentile the surface backscatter is 0.02. A method to account for the ocean subsurface contribution to the integrated depolarization would need to be accounted for in the retrievals. Critical to addressing all of these issues, data collected at higher wind speeds with correlative data is required. We note that Hu et al. (2008) had AMSR-E data with a large number of matchups in clear air and likely oligotrophic (low particulate scattering) ocean conditions. Therefore, we believe that the whitecap correction for HSRL-2 is not ready for publication without further analysis and evaluation, which has started but is still in the early phase.

We have added a general discussion about whitecap correction at the end of Sect. 2.4.

Added: “In addition to the specular reflection from the surface, whitecaps or sea foam can increase the lidar backscatter signal. As noted in Josset et al. (2010b), the contribution of scattering by the whitecaps on the ocean surface has been treated as Lambertian scattering. There is a wavelength dependence of the scattering at longer wavelengths due to the water absorption, based on measurements presented by Dierssen (2019) covering wavelengths from 0.4 – 2.5 µm. Measurements presented here are at 532 nm, a region of the visible spectrum where scattering from foam is relatively constant with wavelength. The contribution of whitecaps is typically modeled with a constant average reflectance and an effective area weighted fraction that varies with surface wind speed (Whitlock et al., 1982; Koepke, 1984; Gordon and Wang, 1994; Moore et al., 2000). Following Moore et al. (2000), we have estimated the average reflectance due to the whitecaps as a function of surface wind speed and the difference becomes > 1 m s^{-1} for surface wind speeds > 15 m s^{-1} based on this relationship. As presented below, there are limited data (49 data points) above 13.3 m s^{-1} that can be compared to the dropsonde surface wind speeds to evaluate this relationship. Moreover, since the correction depends on surface wind speed, an iterative calculation is required to use this relationship as the backscatter is dependent on wind speed.

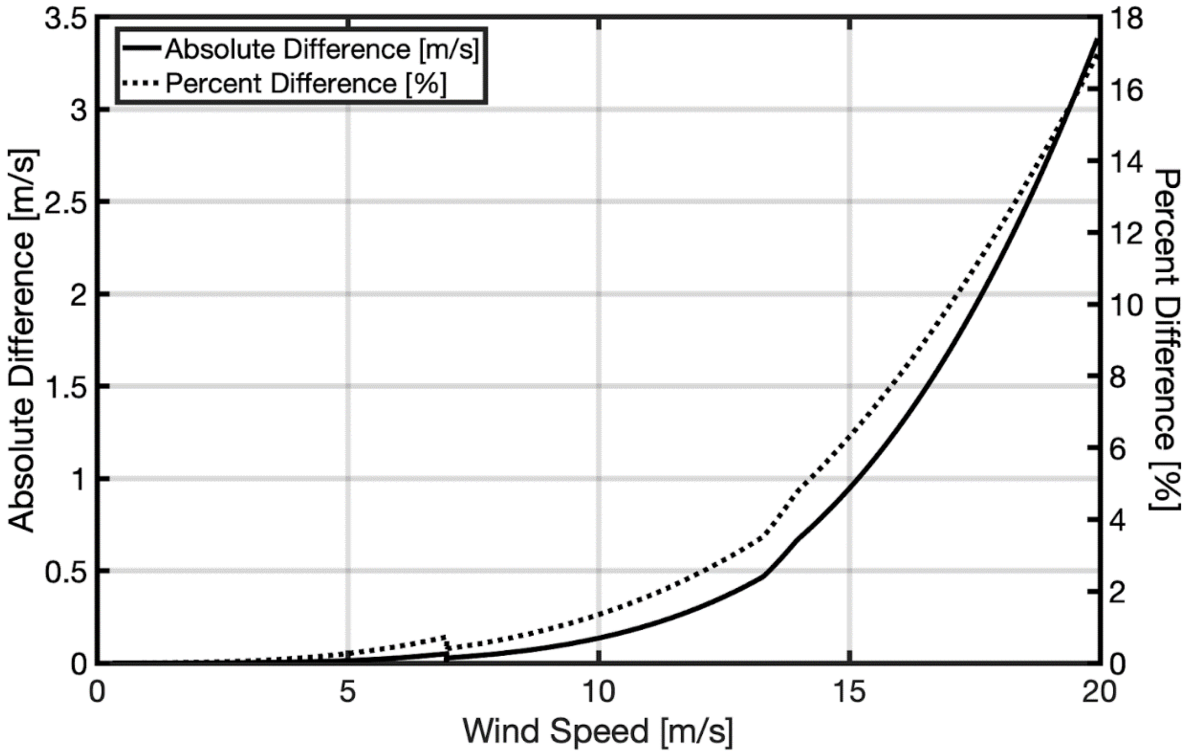


Figure 2. Estimated absolute difference in calculated surface wind speed if reflectance from whitecaps is not included. The lidar surface backscatter is higher than the specular reflectance if whitecaps are present, which results in a lower estimated surface wind speed if not accounted for in the retrieval.

Alternatively, Hu et al. (2008) used a full month of CALIPSO integrated surface depolarization ratio (ratio of the integrated cross polarized channel to the integrated co-polarized channel across the surface) and applied an empirical correction to the reflectance that was determined using AMSR-E data as the ground-truth data set to increase the correlation of the data sets. The correlation was based on much more data than the ACTIVATE matchups between HSRL-2 and dropsondes, limiting the utility of a similar analysis with the HSRL-2. In addition, there are significant differences in the configurations of CALIPSO and HSRL-2 that limit implementation of the same empirical relationship. First, CALIPSO's integrated surface depolarization includes the subsurface contributions due to its 30 m vertical resolution, whereas the HSRL-2 surface depolarization is integrated over only a few meters as shown in Fig. 1. Second, the CALIPSO data is based on global data, which is dominated by oligotrophic (clear) waters, whereas a significant fraction of the HSRL-2 - dropsonde comparisons are from eutrophic and mesotrophic waters near the coast and along the shelf. Third, there is a significant difference in footprint size between HSRL-2 and CALIPSO (8 m versus 90 m), with HSRL-2's instantaneous footprint area being greater than 2 orders of magnitude smaller and, considering HSRL-2's along-track averaging (100 laser shots) compared to CALIPSO's single shot data, greater than one order of magnitude smaller in terms of area over which surface depolarization is integrated."

References

Gordon, H. R. and Wang, M.: Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors, Appl. Opt., 33, 7754-7763, 10.1364/AO.33.007754, 1994.

Dierssen, H. M.: Hyperspectral Measurements, Parameterizations, and Atmospheric Correction of Whitecaps and Foam From Visible to Shortwave Infrared for Ocean Color Remote Sensing, Frontiers in Earth Science, 7, 10.3389/feart.2019.00014, 2019.

Josset, D., Zhai, P.-W., Hu, Y., Pelon, J., and Lucker, P. L.: Lidar equation for ocean surface and subsurface, Opt. Express, 18, 20862-20875, 10.1364/OE.18.020862, 2010b.

Koepke, P.: Effective reflectance of oceanic whitecaps, Appl. Opt., 23, 1816-1824, 10.1364/AO.23.001816, 1984.

Moore, K. D., Voss, K. J., and Gordon, H. R.: Spectral reflectance of whitecaps: Their contribution to water-leaving radiance, Journal of Geophysical Research: Oceans, 105, 6493-6499, https://doi.org/10.1029/1999JC900334, 2000.

Whitlock, C. H., Bartlett, D. S., and Gurganus, E. A.: Sea foam reflectance and influence on optimum wavelength for remote sensing of ocean aerosols, Geophysical Research Letters, 9, 719-722, 1982.

Hu, Y., Stamnes, K., Vaughan, M., Pelon, J., Weimer, C., Wu, D., Cisewski, M., Sun, W., Yang, P., Lin, B., Omar, A., Flittner, D., Hostetler, C., Trepte, C., Winker, D., Gibson, G., and Santa-Maria, M.: Sea surface wind speed estimation from space-based lidar measurements, Atmos. Chem. Phys., 8, 3593-3601, 10.5194/acp-8-3593-2008, 2008.

I note that the authors' equation (2) also omits the atmospheric two-way transmittance term give in equation (21) in Josset et al., 2010b. Given the discussion in the introduction about calibration transfer and the assertion on line 116, this omission seems a bit surprising.

Thanks for the comment since it highlights the need to be clearer on this point. Although correct in Josset et al. (2010b), the formulation is given as the attenuated backscattered signal, which includes the atmospheric attenuation as pointed out in your comment. Here we provide the backscattered (180°) reflected radiance (units sr^{-1}) of the incident light level just above the surface as this is the quantity of interest and is directly related to the wave-slope variance and therefore wind speed. We have changed this from surface backscatter to surface backscatter (180°) reflected radiance. Note that the attenuation in our formulation is introduced when the signals from the lidar are included as shown in Eq. 9.

Added: "To derive surface wind speeds, the surface backscattered (180°) reflected radiance (β_{surf}, units sr^{-1}) is estimated from the surface return signal and related to the wave-slope variance (σ^2), as detailed in Josset et al. (2010b), through..."

When considering the authors' wind speed difference statistics, I kept wondering about wind speed variations over time at a fixed point. HSRL biases relative to the dropsondes are given as either 0.15 ± 1.80 m/s or 0.62 ± 1.70 m/s, depending on the model used. But the temporal offset between matched HSRL and dropsonde wind speed estimates can be as large as 15 minutes. How do these bias magnitudes compare to the natural variations in wind speed that would be measured at a fixed point over a 15-minute time interval? Perhaps wind speed variability information is readily available from the NOAA's National Data Buoy Center? A box and whisker plot showing wind speed differences as a function of temporal offset between the two data sets might also shed some light on this issue.

Thank you for this observation. We initially attempted to address this in the SI file, where Fig. S2 shows that there is a weak correlation between surface wind speed deltas and time. We have looked at the variations from the surface return and the optical depth calculation that will drive variations in surface wind speed and shown those statistics/plots in your other comment about providing an overview of the primary sources of uncertainty associated with calculating β_{surf} using equation (15). This, along with the uncertainty in the dropsonde data, do not match the variability from the comparisons observed. Therefore, there is a significant potential variation of ~1 m s^{-1} from the spatiotemporal differences. With this type of comparison, which involves single points from the dropsondes, we cannot perform a comparison over time to look at the variability unfortunately. Therefore, no change is made to the paper for this comment.

On lines 51–52, immediately after describing the rudiments of Hu et al., 2008 derivation, the authors introduce equation (1) by say, "The wind speed (U) was then approximated from the waveslope variance (σ_2) through this linear relationship". What I was expecting to see were the Hu equations subsequently given as equations (3.1) through (3.3). Instead, equation (1) is Cox-Munk. This section of the text (lines 44–52) should be rewritten to clearly distinguish between the original Cox-Munk equation and the subsequent CALIPSO derivation by Hu.

We apologize for this confusion. We have rewritten this section by introducing lidar retrievals in general and how Cox and Munk first related surface wind speed and surface reflectance. Then, we introduce CALIPSO afterwards and transition into the attenuation discussion.

These lines now read:

"...Therefore, instruments such as lidar are used to provide accurate surface wind speed measurements in various geographical locations to improve estimations of the MABL state globally. For instance, satellite lidar systems that measure aerosol and cloud vertical distributions, such as the lidar on board the NASA Cloud-Aerosol Lidar and Infrared Pathfinder Observation (CALIPSO) satellite, also have the capability to provide horizontally-resolved surface wind speed data. The underlying principle of lidar surface wind speed retrievals was first derived by Cox and Munk (1954), where bidirectional reflectance measurements of sea-surface glint are used to establish a Gaussian relationship between surface wind speeds and the distribution of wind-driven wave slopes. To probe these surface wave slopes, lidar instruments emit laser pulses into the atmosphere and measure the reflectance (or backscatter) of those laser pulses from particles, molecules, and the ocean surface. The magnitude of the measured signal is then used to estimate the variance of the wave-slope distribution (i.e., wave-slope variance) and therefore surface wind speed. Note that reflectance and backscatter are used interchangeably throughout this paper.

Although many studies have expanded upon the original Cox-Munk relationship (e.g., Hu et al., 2008; Josset et al., 2008; Josset et al., 2010a; Kiliyanpilakkil and Meskhidze, 2011; Nair and Rajeev, 2014; Murphy and Hu, 2021; Sun et al., 2023), these parameterizations do not account for atmospheric attenuation by aerosols and therefore have difficulty in calibrating the measured ocean surface reflectance accurately...”

From a quick glance at the studies cited on line 54, I do not find any support for the assertion that “CALIPSO retrievals of surface wind speeds have been used in many studies”. Josset et al., 2010a used AMSR winds to investigate “the normalized scattering cross section” of the CALIPSO lidar and the CloudSat radar. Kiliyanpilakkil and Meskhidze used AMSR winds and aerosol optical properties derived from CALIPSO. Nair & Rajeev used QuickScat winds and CALIPSO cloud heights. Sun et al. uses “numerical weather prediction wind vector assimilated with observed wind component” obtained from ALADIN.

Thank you for pointing this out. The goal of this sentence was to show that many studies have looked at the relationship between surface reflectance and surface wind speed since the original Cox and Munk formulation. However, this message was not communicated properly. Ultimately, we introduce these studies to later establish that the HSRL-2 can account for atmospheric attenuation by aerosols without making assumptions and therefore get an accurate measure of surface reflectance (and therefore, surface wind speed).

Revised: “Although many studies have expanded upon the original Cox-Munk relationship (e.g., Hu et al., 2008; Josset et al., 2008; Josset et al., 2010a; Kiliyanpilakkil and Meskhidze, 2011; Nair and Rajeev, 2014; Murphy and Hu, 2021; Sun et al., 2023), these parameterizations do not account for atmospheric attenuation by aerosols and therefore have difficulty in calibrating the measured ocean surface reflectance accurately...”

Figure 1 and its supporting description are all very nicely done. I commend the authors for their clear and informative presentation of this material.

Thank you for this nice comment as we spent considerable time highlighting the vertical information from the lidar, which is currently not well represented in the literature.

Please provide an overview of the primary sources of uncertainty associated with calculating β_{surf} using equation (15). How is $\Delta\beta_{surf}$ estimated? In practice, is there some maximum $\Delta\beta_{surf} / \beta_{surf}$ above which the retrieval is deemed too unreliable for subsequent wind speed estimation?

We have provided a discussion of the uncertainties in the backscatter reflected radiances in this discussion, but the intent of the manuscript was to show the methods of the measurement approach and then perform a direct comparison with the dropsondes to assess performance. Currently, we have not done an end-to-end assessment of the errors using the SNR from the lidar signals directly but desire to do so in the future. However, we can look at the variance of the different components of the calculation for the ocean surface backscattered radiances in Eq. 15. Based solely on detector shot-noise, we expect the calculation of the two-way optical depth (Eq. 10) to dominate the noise as this signal is much smaller as shown in Fig. 1. The variance in the ocean subsurface is small relative to the integrated surface return, so we expect the random errors from this correction to be negligible. We estimated the variability of both the integrated surface return (numerator in Eq. 15) and the optical depth term from the molecular channel (denominator) over 10 seconds. The following fractional uncertainties were calculated using all ACTIVATE data points. These are the estimated fractional uncertainties, in percent, for the 0.5 s data records.

First, estimated fractional uncertainties are shown below for 0.5 s averaged data based on variances calculated over a 10 s window (all in %):

Surface Area

- Mean: 12
- Median: 11
- 5% quartile: 6.3
- 95% quartile: 22

Optical Depth (normalization)

- Mean: 5.6
- Median: 4.5
- 5% quartile: 3.0
- 95% quartile: 11

Backscatter Reflected Radiance Uncertainty (added in quadrature)

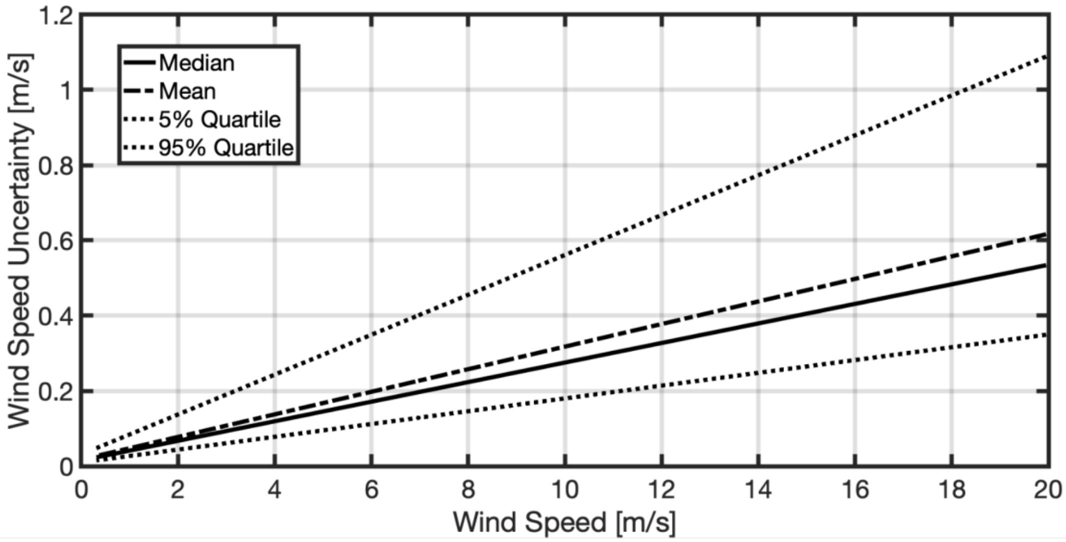
- Mean: 13
- Median: 12
- 5% quartile: 7.8
- 95% quartile: 24

Then, estimated fractional uncertainty for 10 second average based on uncertainties above are shown. This is the averaging interval used for the comparisons of the lidar to the dropsondes winds.

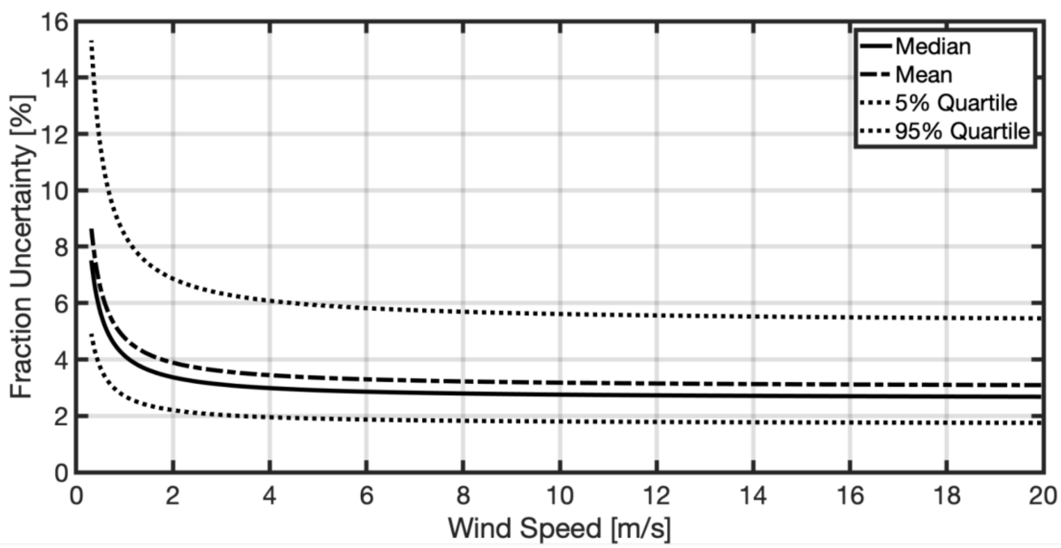
Backscatter Reflected Radiance

- Mean: 3.0
- Median: 2.6
- 5% quartile: 1.7
- 95% quartile: 5.3

Next, the estimated error in wind speeds from the estimated 10 s average backscatter reflected radiance uncertainty based on the linear Cox-Munk empirical model related wave slope variance to wind speed is shown in the plots below.



Surface wind speed uncertainty estimated from the backscatter reflected radiance uncertainty.



Same as above but showing the fractional surface wind speed uncertainty.

Discussion: The mean estimated fractional error for wind speeds is less than 4-6% based on this assessment. Over the range of measurements collected for ACTIVATE, the average estimated wind speed uncertainty is less than 0.6 m s^{-1} and is less than 1 m s^{-1} for the 95th quartile. As pointed out by the reviewer, there are also potential bias errors due to the effects of whitecaps or foam on the water surface that are expected to affect the upper end of the measured wind speed distribution. In addition, there can be bias errors due to the assumed isotropic wave-slope variances and the assumed Gaussian distribution.

The intent of this manuscript was largely to provide an overall assessment of the retrieved wind speeds using correlative measurements from an accepted measurement (i.e., dropsondes) to quantify errors (which might be an upper limit) and to demonstrate the performance of using the high vertical resolution High Spectral Resolution lidar technique to account for both the atmospheric optical transmission and ocean subsurface scattering. The comment also asks if there is a limit of conditions that this is acceptable, and the answer is yes. We have screened data at the 0.5 s fundamental averaging interval when water clouds are detected and limited the conditions on the aircraft roll and pitch. The cloud screening is implemented for multiple reasons that include removing highly attenuated signals, which can make determining the surface location uncertain in addition to the lower signal levels. The limits on the pitch and roll were done to prevent timing delays in the incident angle of the laser beam with the surface, which can rapidly change during turns. No limits were made on the calculated optical depth for the ACTIVATE data other than screening for water clouds. We have provided the 5% and 95% quartiles to show that the signal to noise from cloud free regions is high for the ACTIVATE conditions.

Figures 4–7: ordinary least squares problems can be extremely sensitive to large outliers. Did the authors consider applying an outlier rejection scheme (e.g., Tukey fencing) before computing the regression lines shown in these figures?

Thank you for bringing this up. Initially, we only removed points that were outside the 30 km and 15 min collocation constraints. Based on your question, we've applied Tukey fencing using the standard k = 1.5 and see that most of the surface wind speeds above 13.3 m s^{-1} would be eliminated. However, we decided to leave the high wind speed values in the analysis but recognize that the number of comparisons are limited and we point out that further measurements at these higher winds speeds are needed to fully assess both the empirical relationships and also the contribution due to whitecaps in the lidar backscatter signal. Therefore, no changes are made to the paper.

Lines 470–477: I am totally bewildered by the authors' data screening criteria; i.e., "if a wind speed retrieval is taken in an area with a high cloud fraction [...] the retrieval is deemed a missing value". Please explain what is meant by "cloud fraction" in this context. Is this vertical cloud fraction within an individual HSRL profile? Perhaps naively, I would think that (a) wind speed retrievals would be possible any time the ocean surface was reliably detected and (b) a much better QA metric could be derived from the quality of the surface backscatter signal.

Thank you for this important comment. It inspired us to restructure the case study section by removing the 11 January 2022 research study and expanding on the HSRL-2's retrieval capabilities using Research Flight 29 on 28 August 2020 and Research Flight 14 on 1 March 2020 (Sect. 3.1 now). You are correct that surface wind speed retrievals would be possible any time the ocean surface was reliably detected, which is what the 1 March 2020 discussion communicates now. The HSRL-2 can detect the surface in the breaks/gaps between clouds, so the retrievals still reliably provide data on days with broken cloud scenes. Therefore, we are not constrained to cloud-free conditions like in Hu et al. (2008), which is a significant point to highlight. However, a later comment mentions that there are limitations where the retrievals are no longer applicable such as in the cases of overcast clouds or dense fog.

Revised:

"…

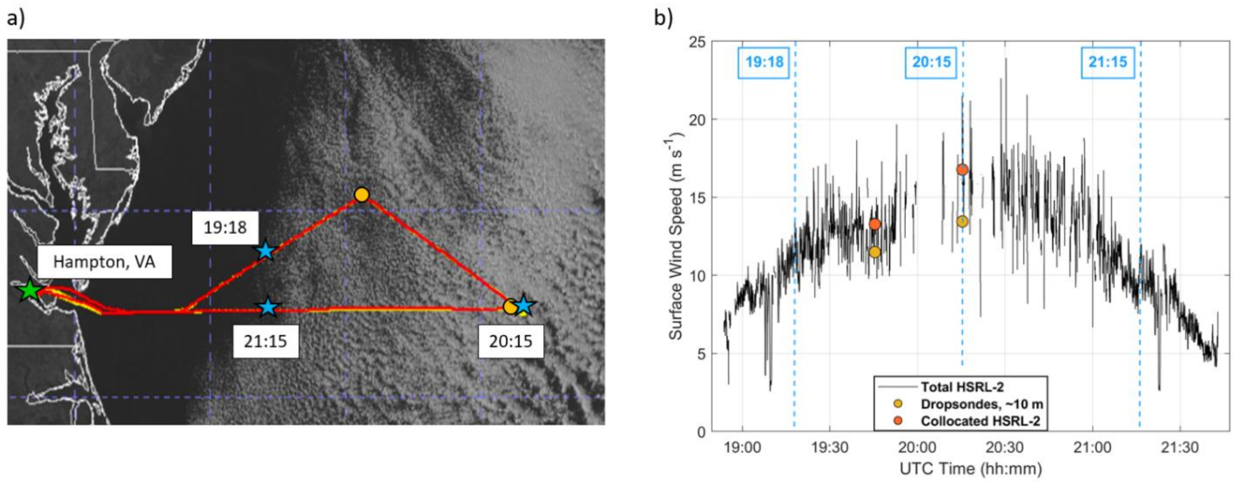


Figure 5: a) Flight map of the King Air (red line), Falcon (yellow line), and dropsondes (dark yellow circles) overlaid onto Geostationary Operational Environmental Satellite (GOES-16) cloud imagery for Research Flight 14 on 1 March 2020. Blue stars represent time stamps where the King Air crosses over from cloud-free to cloudy areas. b) Time series of surface wind speed data from HSRL-2 and dropsondes for the same flight, where lines signify total HSRL-2 surface wind speed data and circles indicate collocated surface wind speed data points. Blue dashed lines represent time stamps of interest as indicated in a).

As the aircraft approaches the cloud scene at 19:18, there is a noticeable and steady increase of HSRL-2 surface wind speeds. The reverse observation is seen when the aircraft approaches 21:15, where the HSRL-2 surface wind speeds start to decrease steadily. As highlighted in the 28 August 2020 case study, the high horizontal spatial resolution of the HSRL-2 retrievals enables these spatial gradients to be observed. Another important takeaway is the HSRL-2 is still able to sample the surface in cloud scenes, as seen by the almost complete surface wind speed profile in Fig. 5b. Although a gap in data occurs at 20:15 where cloud cover is most substantial, some retrievals are still present in that area. The reason is that the HSRL-2 can probe the surface through gaps between clouds, allowing for the surface wind speed retrievals to take place. Although the HSRL-2 retrievals would be unavailable in overcast cloud scenes, the ability of the instrument to sample the surface in broken cloud fields and not just cloud-free scenes is a significant benefit of the lidar and the HSRL technique."

Minor Remarks

The formatting of equation (1) is ambiguous. Decimal notation would be much, much better I think (e.g., 0.003 instead of 3.0E – 3).

Thank you for this comment. Reviewer 1 noted that it is uncommon to write equations in the introduction, so we removed Eq. 1 and left any discussion on wind speed – wave-slope to the Methods section (i.e., Sect. 2.4) and switched to decimal notation as suggested.

Line 105: what value did the authors use for the Fresnel coefficient? Note: Hu et al., 2009 use 0.0209, Josset et al., 2010a use 0.0213, and Venkata and Reagan 2015 use 0.0205.

Thank you for noting this omission and it follows a more general comment from the other reviewer. We have provided a paragraph on the relevant parameters of the instrument and geometry that is implemented before the methodology. Specifically, we use a constant value of 0.0205 for the Fresnel coefficient listed in Venkata and Reagan (2016) due to the limited range of angles analyzed and the constant wavelength from the lidar. The change in the text is provided below.

Inserted: "...C_F is the Fresnel coefficient and is set to 0.0205 as given in Venkata and Reagan (2016)..."

Line 105: what is the typical off-nadir angle for the HSRL measurements? Should readers assume nadir pointing, so that $\theta = 0$?

As noted in the previous comment, we have added an instrument description which we hope addresses the lack of information provided on the viewing geometry. Specifically, the calculations are limited to relatively small angles, but we note that it is important to account for them in the calculations. The nominal aircraft pitch would vary depending on flight conditions from 3 - 5° for the aircraft used, but the angle of incidence could be as large as 10° accounting for both the pitch and roll of the aircraft. We do not go into specific details in the text, but we did limit the roll angles to +/- 3° from the median values to prevent rapid changes in the aircraft altitude data during the data averaging interval and could have a slight lag in time (< 0.5 s). Most flights were conducted with limited turns (< 2 - 3) when over the water.

Added: "For the surface wind speed calculations, data are screened to limit the pitch and roll to less than +/- 3° from the median values, which are approximately 0° for the roll and 3° - 5° for pitch on the King Air."

Line 110: "Eq. 3.3 is similar identical to the log-linear relationship proposed by Wu (1990)."

We apologize for this confusion. We originally tried to communicate that Hu did not simply reuse the Cox-Munk and Wu relationships and these identical results came from his own derivation.

Now, the line reads: "The relationships shown in Eqs. 3.1 – 3.3 were derived by Hu using the comparisons between AMSR-E surface wind speeds and CALIPSO backscatter reflectance mentioned in Sect. 1 and agree identically with the Cox-Munk relationship for surface wind speeds between 7 m s^{-1} and 13.3 m s^{-1} and the log-linear relationship proposed by Wu (1990) for surface wind speeds above 13.3 m s^{-1}."

Line 111: change "to be" to "being"

This line is now removed because we realize it is misleading. The Venkata-Reagan model is also used in CALIPSO retrievals.

Line 192: practically speaking, is there some maximum AOD above which surface wind speeds are not considered reliable? Or are there perhaps some meteorological conditions in which the method is not applicable (e.g., exceptionally dense surface-hugging fogs)?

You are correct and we comment on this in the discussion of the uncertainty above. In general, for cloud-free conditions, the retrievals are possible as noted in the 1 March 2020 case study. We screen for clouds based on the atmospheric backscatter measurements and do not perform the retrievals as there are a lot of conditions that make this challenging including even finding the surface. We agree that dense fog would be one of these conditions and that would be excluded due to the screening process.

Added: "Although the HSRL-2 retrievals would be unavailable in overcast cloud scenes, the ability of the instrument to sample the surface in broken cloud fields and not just aerosol- and cloud-free scenes is a significant benefit of the lidar and the HSRL technique."

Figure 3: use different line colors and/or line types to plot the two different sets of HSRL wind speed retrievals.

We removed most Cox-Munk comparisons throughout the paper, so this figure (now Fig. 4b) reflects this change.

Lines 299–300: In the figure caption, the authors say, "A few collocated Hu08 and CM54 wind speed data points are on top of each other owing to similar values." They could (and should) eliminate any ambiguity by specifying UTC for these pairs of points.

As mentioned in the previous reply, most Cox-Munk comparisons were removed from most of the paper. Therefore, this line has been removed.

Lines 319–325: I would have appreciated a bit more detail here. The authors' description does not provide sufficient information to distinguish the bisector method from other 'errors in variables' techniques (e.g., Deming regression and orthogonal distance regression). Is the bisector method especially effective for problems of this sort? Or will any errors in variables method do equally well?

We appreciate you bringing up this point. This section has been expanded in Sect. 2.5, but we will provide an explanation here as well. Although OLS-bisector, Deming, and orthogonal distance regressions are all errors-in-variable techniques as you mentioned, we have reason to believe that OLS-bisector is the better choice for this study. Orthogonal distance regression assumes that the total error in Y is equal to the total error in X, which is probably not true considering the wind speed data are coming from two different instruments. Deming regression assumes that the measurement error ratio between X and Y is constant (default of 1), so one must provide their own ratio before performing the regression (which is not necessarily straightforward) (Wu and Yu, 2018).

Least squares bisector is where Y is regressed on X (standard OLS) and then X is regressed on Y (inverse OLS) (Ricker, 1973). Then, the technique minimizes the distance of the observation points by drawing a line that bisects the angle of the two regression lines. Then, the variance and covariance of this line are calculated to provide a measure of error in both X and Y. We do need to assume that the relationship between X and Y is approximately linear, but we prefer this method because it is straightforward to use and can calculate the error present in both data sets without making a priori assumptions like in orthogonal distance or Deming.

Added: "Since OLS-bisector is less common than linear regression, a brief explanation of their differences is provided. In linear regression, X is treated as the independent variable while Y is treated as the dependent variable. In other words, one observes how Y varies with changes to fixed X values. OLS-bisector is known as an errors-in-variable regression technique, where X and Y are both dependent variables and thus both subject to error. OLS-bisector regresses Y on X (standard OLS) and then regresses X on Y (inverse OLS), then bisects the angle of these two regression lines (Ricker, 1973). Although other errors-in-variable techniques exist (e.g., Deming regression, orthogonal distance regression), OLS-bisector is chosen because it calculates the error present in both data sets using the bisector rather than assuming an error a priori like the examples mentioned (Wu and Yu, 2018)."

References

Ricker, W. E.: Linear Regressions in Fishery Research, Journal of the Fisheries Research Board of Canada, 30, 409-434, 10.1139/f73-072, 1973.

Wu, C. and Yu, J. Z.: Evaluation of linear regression techniques for atmospheric applications: the importance of appropriate weighting, Atmos. Meas. Tech., 11, 1233-1250, 10.5194/amt-11-1233-2018, 2018.

One Reviewer's Opinion
A scatter plot of dropsonde wind speeds (U) versus matching values of σ_2 (derived using equation (3) and β_{surf} computed using equation (15)) would have made this paper enormously more interesting. Both Cox-Munk and Hu et al., 2008 are approximations of the true relationship between wave slope variance and surface wind speeds. The collocated measurements reported in this manuscript offer a superb opportunity to evaluate the relative merits of both models. Perhaps this tantalizing topic can be briefly explored in an appendix included in a revision to the current manuscript.

Thank you for this great advice. This inspired us to restructure the Results section. We first introduce Fig. 7 (in wind speed space) to show that the Cox-Munk model overestimates the dropsonde wind speeds more so than the Hu model for winds below 7 m s^{-1}. Then, we add the plot you suggest of HSRL-2 wave-slope variance versus dropsonde surface wind speeds (now Fig. 8) to better compare the merits of both models. When comparing the HSRL-2 measurements of wave-slope variance to the wave-slope variance computed using the dropsonde surface wind speeds, we find that the Hu et al. (2008) model provides a better representation of wave-slope variance than the more commonly used Cox-Munk model.

Added: "Now that the HSRL-2 retrievals have been broadly evaluated, Fig. 7 shows how their accuracy varies per 1 m s^{-1} interval in surface wind speed. This plot also provides the opportunity to compare the Hu et al. (2008) model with the models proposed by Cox and Munk (1954) and Wu (1990) to see if some of the error in the HSRL-2 retrievals can be attributed to model characteristics.

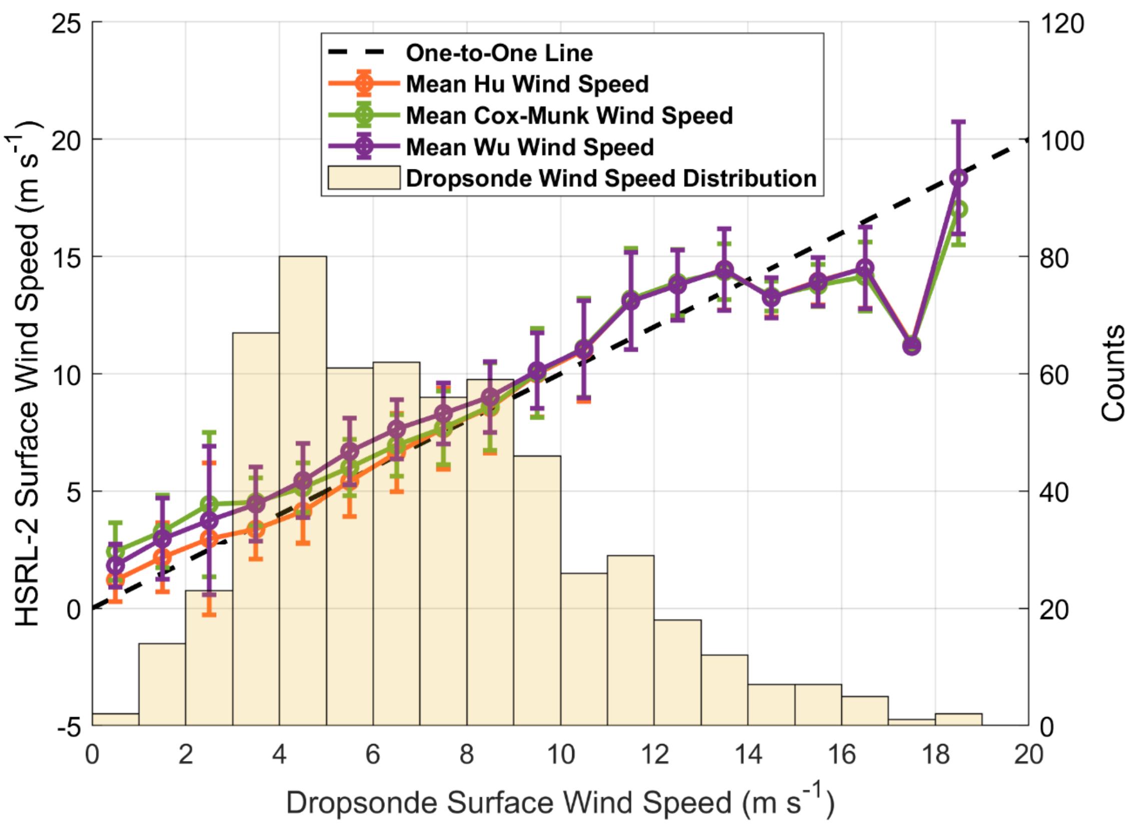


Figure 7: HSRL-2 surface wind speed using Hu, Cox-Munk, and Wu models versus mean dropsonde surface wind speed calculated per 1 m s^{-1} bin. A histogram of dropsonde surface wind speeds is also included to show their distribution.

It is seen that the mean Cox-Munk and Wu surface wind speed values are higher than the mean Hu values from 0 m s^{-1} to 7 m s^{-1}, showing that the Cox-Munk and Wu relationships overestimate dropsonde surface wind speeds more than the Hu relationship. The variability (i.e., STD) around the mean per bin is similar between the three models, which is 1.59 m s^{-1} for Hu, 1.43 m s^{-1} for Cox-Munk, and 1.55 m s^{-1} for Wu on average. Although similar, the STD of the Hu surface wind speeds found here is ~0.4 m s^{-1} lower than the one found in Fig. 6. This could be attributed to an STD not being able to be calculated for the 17 to 18 m s^{-1} bin since it only contained one point.

Although it is apparent Cox-Munk and Wu retrievals overestimate dropsonde observations for surface wind speeds below 7 m s^{-1}, it is still unclear which of the models perform better overall. Therefore, the y-axis from Fig. 7 is converted to wave-slope space and the result of this modification is shown in Fig. 8. HSRL-2 wave-slope is used because it directly reports the original measurements of surface reflectance rather than estimated values of surface wind speed. Using the original data ensures that uncertainty is coming from the actual HSRL-2 – dropsonde comparisons rather than from potential errors in the conversion from wave-slope to surface wind speed.

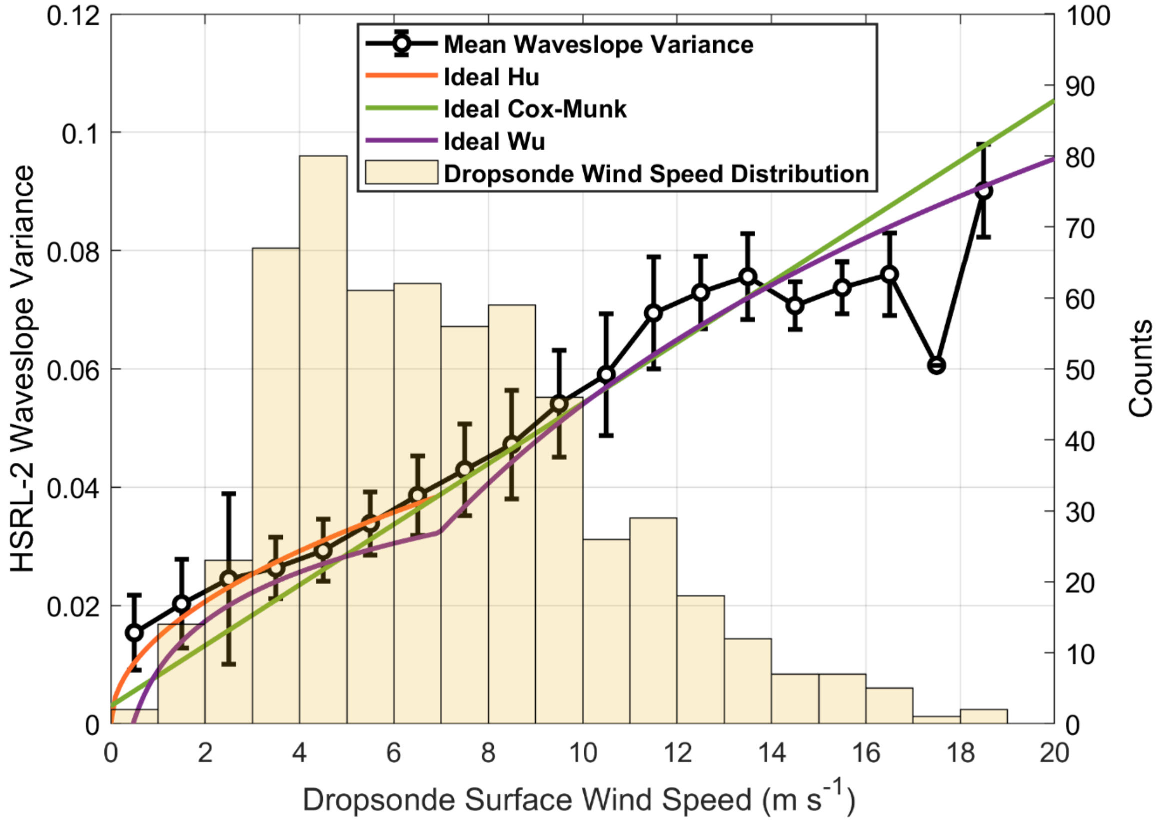


Figure 8: HSRL-2 wave-slope variance versus mean dropsonde surface wind speed calculated per 1 m s^{-1} bin. Ideal Hu, Cox-Munk, and Wu distributions are included to show how well observed dropsonde data match with each parameterization. A histogram of dropsonde surface wind speeds is also included to show their distribution.

From Fig. 8, it is more easily seen how the dropsonde surface wind speed distribution compares with Hu, Cox-Munk, and Wu parameterizations. Dropsonde surface wind speeds match quite closely to Hu and Cox-Munk parameterizations as opposed to the Wu parameterization between 7 m s^{-1} and 13.3 m s^{-1}, although some divergence is seen above ~10.5 m s^{-1}. However, a critical observation that is more apparent in Fig. 8 than Fig. 7 is how the dropsonde data most resemble the Hu distribution for surface wind speeds below 7 m s^{-1}. This improvement is substantial, especially since most of the surface wind speeds in ACTIVATE fall into this category. Surface wind speeds above 13.3 m s^{-1} substantially diverge from all models, especially above 16 m s^{-1}. As mentioned previously, there are few surface wind speed observations in this category, so more measurements are necessary to make meaningful comparisons between the two data sets. Overall, Figs. 7 and 8 demonstrate the benefits of using the Hu parameterization in this study and why surface wind speeds above 13.3 m s^{-1} are not the main focus of the comparisons in this section. Further analysis is warranted to rigorously compare the performance of various surface reflectance models and potentially apply corrections (i.e., whitecap correction for surface wind speeds above 13.3 m s^{-1}), but the aim of this paper is to evaluate LARC's HSRL-2 surface wind speed retrieval algorithm using the available ground-truth dropsonde measurements."

References

Cox and Munk, 1954: Measurement of the Roughness of the Sea Surface from Photographs of the
Sun's Glitter, *J. Opt. Soc. Am.*, **44**, 838–850, https://doi.org/10.1364/JOSA.44.000838.

Hu et al., 2008: Sea surface wind speed estimation from space-based lidar measurements, *Atmos. Chem. Phys.*, **8**, 3593–3601, https://doi.org/10.5194/acp-8-3593-2008.

Josset et al., 2010a: Multi-Instrument Calibration Method Based on a Multiwavelength Ocean Surface Model, *IEEE Geosci. Remote Sens. Lett.*, **7**, 195–199, https://doi.org/10.1109/LGRS.2009.2030906.

Josset et al., 2010b: Lidar equation for ocean surface and subsurface, *Opt. Express*, **18**, 20862–20875, https://doi.org/10.1364/OE.18.020862.

Lancaster et al., 2005: Laser pulse reflectance of the ocean surface from the GLAS satellite lidar, *Geophys. Res. Lett.*, **32**, L22S10, https://doi.org/10.1029/2005GL023732.

Venkata and Reagan, 2016: Aerosol Retrievals from CALIPSO Lidar Ocean Surface Returns, *Remote Sens.*, **8**, 1006, https://doi.org/10.3390/rs8121006.