Supporting Information for "Obukhov length estimation from spaceborne radars"

Owen O'Driscoll[1], Alexis Mouche[1], Bertrand Chapron[1], Marcel Kleinherenbrink[2], Paco López-Dekker[2]

[1]Univ Brest, CNRS, Ifremer, IRD, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, F29280, Plouzané, France

[2]Geoscience and Remote Sensing, Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

Contents of this file

1. Text S1 to S3
2. Figures S1 to S4
3. Tables S1 to S5

Introduction The supporting information contains an expanded description of data preparation in Text S1. This section details the retrieval of a wind field from SAR imagery, the preparations of both ERA5 and buoy data for use with the COARE.5 algorithm, the decorrelation of ERA5 validation as a result of spatial resolution and a comparison between buoy and ERA5 validation data. Text S2 expands upon the methodology. It details the statistics derived from the two- and one-dimensional spectra and the analysis frameworks used to obtain results. Text S3 provides a supplement to the results. It contains

an analysis of the results in a geographical representation, the effects of turbulence scale and viewing geometry on estimation and a variance budget bringing together estimation results and the expected maximum attainable estimation as a result of noise and errors in validation data.

1. Text S1. Data preparation

1.1. Wind-field retrieval

SAR backscatter intensity is generally related to a current-relative, neutral wind field at 10 m above the ocean surface (throughout this study simply referred to as wind field) through a geophysical model function (GMF). GMFs require an a-priori wind direction, which is provided by the ERA5 global reanalysis product along with information on the SAR viewing geometry. This study employs CMOD5.N, a GMF developed to estimate wind speeds for $\mathcal{O}$(10km) C-band scatterometer observations (Hersbach, 2008). SAR resolution is still orders of magnitude greater than that of scatterometers, for which CMOD5.N is designed. Therefore we cannot discount the possibility that CMOD5.N is incapable of quantifying high-resolution wind-field properties. The GMF assumes the observed backscatter to be solely related to the local wind conditions; any imprints caused by rain bursts, swell, surface slicks or currents can lead to erroneous estimates of the wind field. With the exception of currents, most of these errors can be separated in the frequency domain or avoided by restricting analyses to homogeneous scenes.

1.2. ERA5 data

Included in the retrieved ERA5 data are: 10 m horizontal wind vectors, 2 m air and dew point temperature, sea-surface temperature, surface pressure, surface thermal radi-

ation downwards (long-wave radiation), surface solar radiation downwards (short-wave radiation) and boundary layer height. Wind vectors are converted to a magnitude and orientation w.r.t. North and relative humidity is computed from the air and dew-point temperatures using the Arden-Buck equation. Parameters boundary layer height, long-wave and short-wave radiation have a minimal impact on the computation of L_{ERA5}, but are nonetheless required inputs to run the COARE3.5 algorithm. Alternatively these parameters could be attributed a constant value. No surface-current vectors are available from ERA5 such that the wind speed can not be converted to a current relative equivalent, as advised by the COARE3.5 algorithm.

1.3. Buoy data

Buoy data from WHOI (2022); NDBC (2022); NOAA (2022a, 2022b) is used as a secondary validation source. It serves as a check on ERA5's validation capacity and as an independent validation for a small subset of co-located SAR observations. SAR observations are considered co-located with one of 269 buoys if they occur within 4 hours and a 200 km radius of a measurement. To increase the number of co-locations the analysis was performed over January 2014 - April 2022. During this time window, a total of 7565 co-locations occur between scenes classified as rolls and 103 unique buoys. All buoys measured air/sea temperature, relative humidity (or dew point temperature from which relative humidity can be calculated) and wind speeds/directions. COARE3.5 provides an ideal tool to calculate L_{buoy} from buoy measurements as it can take into account the variable measurements heights of each buoy. None of the buoys measured all the required parameters to run the full COARE3.5 algorithm, necessitating supplementary

ERA5 data. This includes the spatiotemporal co-located ERA5 data for boundary layer height, long-wave radiation and short-wave radiation. Additionally, only a few buoys measured surface pressure. For consistency the ERA5 surface pressure values were used for all buoys instead. Similar to ERA5, no (consistent) observations of the currents were made such that the wind speed could not be made current relative.

1.4. ERA5 decorrelation

The coarse $1° \times 1°$ resolution of used ERA5 data leads to decorrelation between observations and validation. By quantifying this effect we can determine what fraction of the error budget is controlled by ERA5's resolution and whether or not a finer resolution is required. The average decorrelation is predicted by analyzing the mean difference between an ERA5 retrieved L_{ERA5}-field with a shifted version of itself. To increase the representability of results for co-location between SAR and ERA5, the analysis is performed between -20° and 20° latitude (corresponding the highest concentration of observed rolls), all shifted data points involving positive L_{ERA5}'s are removed and the distribution of L_{ERA5} is further filtered such that it is similar to that of ERA5 co-located with SAR (not shown). Assuming that the decorrelation gradient between an offset of 0° and 1° holds on a sub-grid level, that all offsets are purely longitudinal and that the average offset between observations and the points hypothetically represented by ERA5 within the sub grid is 0.25°, the found maximum achievable R^2 and minimum median absolute error (MAE) on a logarithmic scale are approximately 0.94 and 0.02 respectively.

1.5. Buoy-ERA5 comparison

Comparing a subset of SAR observations co-located with both ERA5 and buoys enables us to assess the validation capabilities of ERA5 with respect to the ground truth measurements from the buoys. Table S1 contains comparison results. A direct comparison between these two validation sources gives a MAE of 0.202 which is equal to a median relative error of 59%. An R^2 of 0.258 indicates that just over a quarter of the variance from ERA5 is directly explained by validation retrieved from buoys, highlighting the volatility of Obukhov-length estimation. The difference between validation sources may be caused by consistent offsets. To remove the effect thereof, a ML algorithm is trained to regress buoy data (raw measurements and those computed from COARE3.5 but excluding any unique buoy/network identifier) towards L_{ERA5}. Indeed the R^2 value for the predictions of this algorithm, $\widehat{L}_{\mathrm{buoy_{ML}}}$, increases nearly three-fold to 0.759: when accounting for consistent offsets, buoys are able to explain approximately 76% of ERA5's variance. This ballpark figure provides a lower bound on the fraction of ground-truth-related atmospheric information contained in ERA5's data set. A separate analysis on a greater buoy-ERA5 co-locations data set yielded slightly better results ($R^2 \approx 0.8$, not shown) even after excluding parameters potentially contaminated by ERA5, e.g. parameters calculated through COARE3.5. Therefore, a maximum of approximately 24% of L_{ERA5}'s variance remains unexplained of which a minimum of approximately 6%-points are caused by spatial decorrelation (see Text S1.4), leaving a maximum of 18% of L_{ERA5}'s variance to be miscellaneous noise. When comparing estimates from the analytical approach, $\widehat{L}$, with the different validation sources, L_{buoy} and L_{ERA5}, results are closer to

the buoy than ERA5 as indicated by the greater R^2 and lesser MAE. Two new ML algorithms are trained, $\widehat{L}_{\mathrm{ML}_{\mathrm{ERA5}}}$ and $\widehat{L}_{\mathrm{ML}_{\mathrm{buoy}}}$, using all the SAR-derived parameters. The former is trained towards validation from ERA5 and the latter towards validation from the buoys. The performance difference between either $\widehat{L}_{\mathrm{ML}_{\mathrm{ERA5}}}$ or $\widehat{L}_{\mathrm{ML}_{\mathrm{buoy}}}$ and their respective validation is statistically insignificant. This suggests that both validation sources are approximately equally representative of the atmosphere as extracted from SAR imagery. Results appear slightly better when regressing towards validation from ERA5, perhaps because the more precise point-wise buoy measurements experience greater spatiotemporal decorrelation than the large-scale averaged ERA5 values. On a small sample of co-locations, results from the validation comparison substantiate the use of ERA5 for validation: even though ERA5 differs from the buoys, it differs consistently such that predicting ERA5 is akin to indirectly predicting buoy values.

2. Text S2. Supporting Methodology

2.1. Spectral computations

2.1.1. Two-dimensional spectrum

First five rows and columns are clipped from all sides of the wind field, thereby removing pixels prone to discontinuities and errors. Next each wind field is subdivided into four equal-sized tiles of approximately ten by ten kilometers. A two-dimensional Hann window is applied on each wind-field tile followed by the subtraction of the mean (removing the 0^{th} frequency component). Next the Power Spectral Density (PSD) of the windowed and mean-removed wind-field tile is computed in the spatial domain yielding $S_i(\xi_x, \xi_y)$, where ξ is the spatial frequency with units of m^{-1} in the x-axis and y-axis of the i^{th} wind-field tile.

The four PSD's per wind field are averaged to create a single smoothed PSD representative of the whole wind field, $S(\xi_x, \xi_y)$. The hypothetical power loss due to windowing is compensated. For non-homogeneous scenes this correction may be insufficient, such that

$$\sigma^2_{\text{windfield}} \neq \int_{\xi_x} \int_{\xi_y} S\left(\xi_y, \xi_x\right)\, d\xi_y\, d\xi_x. \tag{1}$$

The fraction of variance gained/lost after applying a window correction, as compared to the variance of the original wind field, is kept and will be referred to as the window effect

$$\text{window effect} = \frac{\int_{\xi_x} \int_{\xi_y} S\left(\xi_x, \xi_y\right)\, d\xi_y\, d\xi_x}{\sigma^2_{\text{windfield}}}. \tag{2}$$

This parameter partially captures the inhomogeneity of the wind fields in question. The Cartesian $S(\xi_x, \xi_y)$ is linearly interpolated to a polar coordinate system giving $S(\theta, \xi)$ where θ and ξ are the angle with respect to image north (satellite azimuth direction) in degrees and spatial frequency in m^{-1} respectively. SAR imagery is loaded at 100 m resolution in both axes ($\Delta x \approx \Delta y$), such that

$$\xi_{\text{Nyquist}} \geq \frac{1}{2 \cdot \Delta x} = \frac{1}{200}. \tag{3}$$

Variance in $S(\theta, \xi)$ for $\xi > \xi_{\text{Nyquist}}$ contains high-frequency energy sampled in fortuitous angles only. The total energy beyond ξ_{Nyquist} is stored as the variance beyond Nyquist.

A wind-streak containing wind field is shown in upper left of Figure S1 with its corresponding two-dimensional PSD, $S(\xi_x, \xi_y)$, displayed in the upper right. The majority of

energy is located at high frequencies ($\xi > 1/600$ m^{-1}) with a narrow circular concentration of energy at the maximum vertical extents indicating a swell system. The wind-streak related energy is observed within the bandpass of interest ($1/600$ m$^{-1} \geq \xi > 1/3000$ m^{-1}) where the energy concentration is perpendicular to the streak orientation. Energy within the lowpass frequencies ($\xi < 1/3000$ m^{-1}) contains information about mesoscale phenomena.

The lower left of Figure S1 contains annotations of further parameters derived from $S(\xi_x, \xi_y)$. The orientation of maximum integrated energy over frequencies within a Gaussian smoothed bandpass gives the energy direction which, being the orientation of the rolls, could be used as a proxy for wind direction e.g. Figure 3c in Wang et al. (2019). The 180° spectral ambiguity is removed by selecting the orientation closest to that from the ERA5 wind direction. The energy direction with respect to geographical North is converted to a direction with respect to the radar range direction δ (energy direction in range). A spectral area within 20° to the left and right of the energy direction is referred to as the beam. Variance within this beam and the intersection of the beam with the bandpass are stored as variance beam and variance bandpass beam respectively. Contours are shown for which frequencies greater than $1/3000$ m^{-1} a cumulative integration of $S(\xi_x, \xi_y)$ per angle achieves 25%, 50% and 75% of the average energy within the bandpass. For each contour the mean, median, standard deviation and median absolute deviation in pixels from the center are stored. These zero-dimensional parameters give insight into the spread and directionality of the spectral energy.

2.1.2. One-dimensional spectrum

The one-dimensional spectrum $S(\xi)$ in the lower left of Figure S1 is computed from the two-dimensional spectrum in the upper right by integrating over θ such that

$$S(\xi) = \frac{\int_{\theta=0}^{360} S\left(\theta, \xi\right)\, d\theta}{\Delta\theta} \tag{4}$$

where $\Delta\theta$ is the angular spacing of polar spectrum in radians. In the two-dimensional example spectrum of Figure S1 one can see high-frequency energy from swell located perpendicular to the orientation of wind-streak energy within the bandpass. A polar integration over θ prevents this energy from projecting itself on the otherwise smooth $-5/3$ inertial subrange. The one-dimensional spectrum within the bandpass is scaled with $\xi^{2/3}$ such that the higher frequencies increase in amplitude with respect to the lower frequencies. The peak of this scaled spectrum is selected as the spectral peak. Without scaling one would always select the mesoscale as containing a spectral peak when both a mesoscale and microscale peak are present. Spectral peaks located at $\xi < 1/1500$ m^{-1} are assumed to fall within the mesoscale whereas peaks at $\xi > 1/1000$ m^{-1} are assumed to fall within the microscale. Spectral troughs are selected as the lowest point between the spectral peak and $\xi = 1/300$ m^{-1}.

Estimated spectral peaks and spectral troughs can be located at the same frequency as the latter is not scaled prior to analysis. For these undesirable cases the spectrum is successively scaled with decreasing powers during trough detection such that the amplitude of high frequencies decreases with respect to that of the low frequencies. Though this increases the probability of finding a trough at high frequencies, the quality of the

slope between peak and trough for these cases may be of poor quality. Several filtering operations are applied to remove such cases.

The range between peak and trough is assumed to be the inertial subrange with a $-5/3$ spectral slope. As a quality check this slope is scaled with $\xi^{5/3}$ which, for a perfect inertial subrange, yields a normalised spectral amplitude identical for each frequency. The frequency-weighted standard deviation of $S(\xi) \cdot \xi^{5/3}$ is divided by the median value giving a frequency-weighted normalised slope deviation of the inertial subrange, $z_{S(\xi)}$. Poorly identified inertial subranges are detected through high values of $z_{S(\xi)}$ and subsequently removed.

The used similarity-theory equations require a PSD as a function of temporal frequency $S(n)$, where n is the temporal frequency with units Hz. Invoking Taylor's hypothesis, temporal frequency n is calculated from spatial frequency ξ as

$$n = \xi \overline{U}, \tag{5}$$

where $\overline{U}$ indicates the averaged wind field over the entire scene. In order to decrease sensitivity to systematic errors and outliers in the wind field—caused by erroneously attributing the magnitude of observed backscatter to the wind field alone, neglecting any current-induced effects or due to the inhomogeneity of the estimated wind field—this study uses the median wind speed $\widetilde{U}$ in all places where literature prescribes $\overline{U}$ or U. Following Equation 1 from J. C. Kaimal, Wyngaard, Izumi, and Coté (1972), this leads to the calculation of $S(n)$ as

$$S(n) = \frac{1}{\overline{U}} S(\xi). \tag{6}$$

2.2. Analytical framework

2.2.1. Shear component

Wind-induced surface stress τ is related to friction velocity u_* and wind velocity following

$$u_*{}^2 = C_d \overline{U}^2, \tag{7}$$

$$\tau = \rho_{\text{air}} u_*{}^2, \tag{8}$$

where C_d is the drag coefficient, $\overline{U}$ the mean horizontal wind speed and ρ_{air} the air density (Fairall et al., 1996). The drag coefficient C_d is calculated following

$$C_d = \left[\frac{\kappa}{\ln(z) - \ln(z_0) - \psi_m} \right]^2, \tag{9}$$

where z is the measurement height (10 m for CMOD5.N), z_0 the aerodynamic roughness length, κ the von Kármán constant (According to Table 2 in Foken (2006) $\kappa \approx 0.40$) and ψ_m the diabatic wind profile function (Stull, 1988). For neutral conditions ($zL^{-1} = 0$), the ψ_m term becomes zero and C_d will be referred to as C_{dn}. Since neutral wind speed is retrieved from the GMF, C_{dn} will be used for all shear-component calculations. In turn the aerodynamic roughness length z_0 can be calculated according to the Charnock relation as (Fairall et al., 1996)

$$z_0 = a\frac{{u_*}^2}{g} + 0.11\frac{\nu}{u_*}. \tag{10}$$

The Charnock constant a is an empirical measure used to quantify the sea state (Smith et al., 1996). Young, Sikora, and Winstead (2000) use a constant value of 0.011. According to Hersbach (2011), the a value for seas with swell hovers around 0.01 whereas for young steep seas it is closer to 0.04. Typical values fall within this range. For consistency, this study will use a value of 0.011. A constant value of 9.8 ms^{-2} is chosen for gravitational acceleration g and, according to Equation 78 presented in Andreas et al. (1989), the kinematic viscosity of air ν is approximately $1.5 \cdot 10^{-5}\mathrm{m}^2\mathrm{s}^{-1}$ for an air temperature of 293 K. This series of equations is solved iteratively taking an initial estimate of C_{dn}. Convergence is achieved within a few iterations. The calculated values for z_0, C_{dn} and u_* are used in the successive algorithms. Parameters ρ_{air}, a and ν are approximated rather than retrieved from measurements. A summary of the convective algorithm is outlined in Table S2.

2.2.2. Convective component

The convective atmospheric component is estimated from the variance present within the inertial subrange. Therefore a portion of the microscale spectrum must be identified as following the -5/3 power law in the one-dimensional spectrum $S(n)$. As the analysis is frequency normalised it does not matter at which frequency in the inertial subrange the analysis is performed, as long as it falls on the same -5/3 slope. Rewriting Equation 4 from J. Kaimal et al. (1976) yields

$$w_* = \sqrt{\frac{(2\pi)^{2/3} {f_i}^{2/3} n S(n)}{\alpha \psi^{2/3}}}, \tag{11}$$

where α is the Kolmogorov constant of approximately 0.5 and ψ the dimensionless energy dissipation rate. Results obtained in Young et al. (2000) used a mixed-layer value of $\psi = 0.6$. This study uses $\psi = 1.0$ which, based on Figure 4 in J. Kaimal et al. (1976), is deemed more representative for the surface layer. Lastly, f_i represents the dimensionless frequency calculated following

$$f_i = \frac{n Z_i}{\overline{U}} \tag{12}$$

with Z_i being the boundary layer height. In further calculation towards Obukhov length, the value of Z_i will cancel itself out, making its estimation unnecessary. A value for the convective velocity scale, $w_{*,i}$, is obtained for each analysed frequency n_i within the inertial subrange. These values should be near identical meaning they can be combined using a frequency-weighted average to form a single convective velocity scale, w_*, representative of the entire inertial subrange. The final value of w_* is combined with a rewritten version of Equation 4.2a from Stull (1988) yielding

$$\overline{w'T_v'} = \frac{{w_*}^3 \overline{T_v}}{g Z_i}. \tag{13}$$

Plugging $\overline{w'T_v'}$ into Equation 5.7c from Stull (1988) yields Obukhov length L following

$$L = -\frac{\overline{T}_v u_*^3}{\kappa g \overline{w'T_v'}}. \tag{14}$$

The $\overline{T_v}$'s cancel themselves out in Equations 14 and 13. Having calculated L, a stability correction is applied to the original wind field to account for unstable atmospheric stratification. Young et al. (2000) calculates a stability correction factor χ as

$$\chi = 1 - \left(\frac{\psi_m \sqrt{C_{dn}}}{\kappa} \right), \tag{15}$$

where C_{dn} is retrieved from the shear algorithm and ψ_m is computed by plugging the obtained value of L into

$$x = \left(1 + 16 \left| \frac{z}{L} \right| \right)^{1/4}, \tag{16}$$

followed by plugging x into (corrected for typo)

$$\psi_m = \ln \left[\left(\frac{1 + x^2}{2} \right)^2 \right] - 2\tan^{-1}(x) + \frac{\pi}{2}. \tag{17}$$

Next χ is multiplied with all the mean wind-speed terms in the convective algorithm. Since $S(\xi)$ captures the variance of U, its correction becomes χ^2. However, following Equation 6, $S(\xi)$ is divided by U (and thus χ) yielding $S(n)$, such that the correction for $S(n)$ becomes $\chi^2 \cdot \chi^{-1}$. The values for f_i, $S(n)$ and w_* are updated iteratively until convergence. A summary of the convective algorithm is outlined in Table S3.

2.3. Machine Learning

2.3.1. Data filtering

ML algorithms are trained towards matrix y containing the transformed validation values of Obukhov length $\mathrm{Log}_{10}(|L_{\mathrm{val}}|)$, where L_{val} is either L_{ERA5} or L_{buoy}. All observations in y with a positive L or a positive heat flux orientated into the ocean are removed; the

analytical model is incapable of predicting these phenomena. The remaining y has size $m \times 1$ with m observations.

The matrix of independent variables X of size $m \times n$ contains m observations for n parameters. The n parameters in X are devoid of any validation data (either from ERA5 or buoys) with the exception of the ERA5-retrieved wind direction which is implicitly contained in all wind-field related parameters. Included in n are zero-dimensional parameters derived from the SAR scene along with information about the acquisition, e.g. $\widetilde{U}$, u_*, L, incidence angle ϕ, etc. Spatiotemporal information (latitude, longitude and time) is withheld from the ML-algorithm to prevent it from fitting to geographical locations or seasons rather than observation.

Observations are filtered prior to training. Primary filtering is performed by thresholding on observed parameters. This includes filtering of observations for which any of the following conditions are not met: $1.5 \geq$ window effect ≥ 0.5, normalized spectral slope deviation $z_{S(\xi)} \leq 0.25$, energy direction from Figure S1 differs by less than 30° from θ_{ERA5}, 200 m $\leq$ inertial subrange length $\geq$ 20% of the peak production wavelength and 2950 m $\geq$ peak production wavelength $\geq$ 610 m. The first threshold removes inhomogeneous scenes, the second removes scenes with a poor fit to the spectral -5/3 slope, the third removes scenes with likely classification errors and the fourth and fifth thresholds remove scenes with unlikely inertial subranges. The ensemble of primary filtering operations insures that only the SAR scene are considered for which one could reasonably assume the analytical method to be applicable. In fact, primary filtering slightly penalizes ML results

in return for systematically reducing prediction noise (a minor R^2 reduction in return for a significant MAE reduction, not shown).

A secondary and more refined filtering operation aims at removing outliers. First all observation columns n in X are scaled by removing the mean and dividing by the their respective standard deviations. Next the dimension of X_{scaled} is reduced using Principal Component Analysis (PCA). The first principal components explaining a combined 80% of the total variance, $X_{\text{PCA 80\%}}$, are used. Outliers in $X_{\text{PCA 80\%}}$ are found using the unsupervised Local Outlier Factor outlier-detection algorithm with 100 nearest neighbors (implemented in Sci-kit learn, Pedregosa et al., 2011). Observations that do not meet the set thresholds or whose values can be considered outliers are subsequently removed from both y and X to maintain identical m-size. Lastly, both y and X are split into 80% training and 20% testing fractions. Performances of the analytical and ML-assisted methods are assessed on the test fraction of the filtered scenes.

For the rolls data set, approximately two-thirds of the data is removed in this manner, leaving a total of 35,907 observations (out of 124,682). The ensemble of filtering operations decreases the proportion of observations for which L_{ERA5} indicates near-neutral unstable conditions, conditions under which the Obukhov length tends to explode.

For the rolls & cells data set, the energy-direction filtering operation is dropped as the spectral energy orientation of cells cannot consistently be interpreted as a streak orientation. Other filtering operations remain the same. Of the resulting composite data set containing 100,538 observations, approximately two-thirds is classified as cells and one-third as rolls. A greater number of the former pass the filtering operation as the spectral

signature of cells tend to be more clearly defined. Additionally, fewer cell observations are removed due to co-located ERA5 indicating non-convective conditions. With rolls being generally less unstable, this plagues them particularly.

2.3.2. Regression

Regression is performed using a so-called stacked regressor where multiple regressors are trained on the data independently followed by training a final regressor on the output of the original regressors. This final regressor learns from the combination of original results, allowing it to make a new and improved prediction. The following independent regressors are used: LightGBM, XGBoost, CatBoost, Bayesian Ridge and Lasso LARS. The former three use state of the art tree-based boosting (Ke et al., 2017; Chen & Guestrin, 2016; Prokhorenkova et al., 2018, respectively) which, when trained on sufficient data, consistently outperforms the latter two (both implemented by Pedregosa et al., 2011), which are added to increase diversity for the stacking regressor. The outputs of these five regressors are combined, or stacked, using a final Ridge regressor (Pedregosa et al., 2011).

Hyperparameters of the independent regressors are tuned using a Tree-structured Parzen Estimator (Bergstra et al., 2011) and hyperband-based pruner (Li et al., 2017) implemented in the open-source optimization software Optuna (Akiba et al., 2019). Optimization targets the minimization of the mean MAE between y and estimate $\hat{y}$ after a 5-fold cross validation on the training data. Alternatively one could opt for the maximization of R^2 for which results are nigh identical (not shown). The hyperband pruner is called after successively training on 10%, 20%, 30%, 40% and 60% of the training data. Hyperparameter optimization per regressor is constrained to 200 iterations or 40 minutes

run time, depending on which condition is met first. Hardware and software specifications are: Python 3.9.12, Ubuntu 20.04.4 LTS, 64 GB of DDR4 at 3200 MHz and an Intel Xeon(R) W-11955M CPU @ 2.60 GHz × 16.

The optimal preprocessing scaler of the X matrix is sought in addition to optimizing the regression hyperparameters. Namely, whether to scale all columns of the X matrix using one (or none) of several common preprocessing scalers available in the Scikit-learn Python package (Pedregosa et al., 2011). Similarly, the optimization procedure decides whether to transform the dependent matrix y to a normal distribution (and to how many quantiles this normal distribution should contain) prior to regression. Such a transformation can boost performance when the regressor assumes y to follow a normal distribution. The X scalers and y transformations are selected per independent regressor. Information leakage is prevented by fitting the scalers and transformations on the training data only.

3. Text S3. Supplement to results

3.1. Spatiotemporal analysis

Spatiotemporal averages of estimates are shown in Figure S2. Though the magnitude may be off, spatial features of the analytical results in the upper left overlap with those from validation in the upper right: there is increased instability near the coasts (in particular for the equatorial west coasts), a thin equatorial region with greater instability separates the northern and southern tropical regions with lesser instability, followed by a transition into oceans north and south of 30°N & 30°S respectively. Results from the ML-assisted approach in the lower right are qualitatively different from those obtained with the analytical approach and very similar to those from validation.

Spatial differences between $|\widehat{L}|$ and $|L_{\mathrm{ERA5}}|$, shown in the upper left of Figure S3, indicate under estimation of $|L|$ north of 30°N and south of 30°S whereas within these bands there is over estimation. Notable exceptions are the coasts bordering Angola, Ecuador, Senegal and the sea just north of Madagascar. Spatial differences between $|\widehat{L}_{\mathrm{ML}}|$ and $|L_{\mathrm{ERA5}}|$ can be discerned from the upper right of Figure S3. The ML algorithm is capable of explaining away most of the original under estimation of $|L|$ for the seas bordering Angola, Ecuador, Senegal and Madagascar, i.e. observations for these regions contain useful signatures. The fact that error signs of ML results appear regionally correlated indicates that regional phenomena are sources of errors. For instance, the coherent structures of flipping error signs near the equator are located where one would expect equatorial currents in close proximity to the equatorial counter current. These apparent errors could originate from the validation side, as a result of, for instance, not supplying current-relative wind speed to COARE3.5 or poor parameterisation of currents in ERA5's ocean-atmosphere interactions. The error sign of the maps in the top panel of Figures S3 may fortuitously cancel out, falsely giving the impression of low median errors. This is mitigated by analyzing the absolute median errors, provided in the lower panel, which confirm that the magnitude of errors is not significantly affected by sign cancellation.

3.2. Turbulence-scale effect

Theoretically the analytical method is only valid for the microscale inertial subrange. In Table S4 the ML prediction performance is assessed across the mesoscale and microscale in order to determine whether its applicability is also constrained by scale. We separated scales according to the wavelength of the found spectral peak. Scenes with spectral peaks

greater than 1500 m are assumed to contain mesoscale convection whereas spectral peaks located at less than 1000 m are assumed to fall within the microscale. Performances are obtained from an ML algorithm retrained on an equal number of scenes for both scales. The difference in R^2 score is statistically insignificant while a statistically significant lower MAE value for the mesoscale suggests that a ML-assisted approach is systematically better at inferring atmospheric stability from larger-scale turbulence. A sampling difference between scales is the likely cause of this result: the distribution of L_{ERA5} within the microscale is broader than that of the mesoscale (not shown), potentially penalizing the MAE metric. Nevertheless, it is evident that the ML-assisted approach is able to extract significant atmospheric information from scenes containing either micro- or mesoscale turbulence. The indifference to scale confirms that little predictive power is extracted from the scale-sensitive parameters derived in the analytical methodology.

3.3. Viewing-geometry effect

Radar observations of the ocean surface are generally sensitive to the orientation of surface waves with respect to the sensor. ML-estimation performance is assessed for different bins of ERA5 wind directions with respect to the radar range direction, δ_{ERA5}, in order to determine the effect of this inherent sensitivity to the extraction of atmospheric information. Results are presented in Table S5. A δ_{ERA5} of 0° indicates wind propagating exactly down (or up) the radar range direction with a maximum value of 90° indicating wind propagating perpendicular to the range direction. The ML algorithm is retrained per δ_{ERA5}-angle bin using an identical number of observations following the same distribution of L_{ERA5} (according to a two-sample Kolmogorov-Smirnov test the distributions across

bins are insignificantly different with all p-values greater than 0.95). The small sample size of overlapping L_{ERA5}'s across δ_{ERA5}-bins prevents statistically significant results; almost all mean values fall within two standard deviations of all other means. Nonetheless, from the 18-36 bin onwards the explainable variance R^2 decreases while the median error increases. This suggests that the quantity of atmosphere-related information extracted from SAR scenes decreases for winds propagating at greater angular differences with respect to the radar range direction. A similar observation is made in Wang et al. (2020) where the detection of rolls decreases with wind directions propagating perpendicular to the radar orientation, again suggesting a gradual loss of atmospheric information contained in SAR imagery as a function of viewing geometry. The relatively poor performance for 00-18 bin might be caused by increased sensitivity exactly down range, systematically picking up noise and/or sampling a broader range of possible parameter values associated with L. This would also explain the increased uncertainty ranges.

3.4. Variance budget

Analyses of the validation data in Text S1.5 suggest a maximum of 24.1% of variance observed in the validation data may be ascribed to noise, of which a minimum of 6%-points as a result of spatial decorrelation (Text S1.4). These ballpark figures lead to an $R^2_{\mathrm{max}} = 0.759$. Therefore, with an obtained $R^2 = 0.673$, approximately 88% of the explainable variance is resolved. The remaining 8.6%-points of unexplained variance provide an indicative limit for attainable improvement. Resolving all individual uncertainties—such as those related to the high-resolution applicability of CMOD5.N, the inability to capture/extract atmospheric information through the used zero-dimensional parameters

or the disregard of surface currents—is unlikely to improve results by more than 8.6%-points, and certainly no more than 26.6%-points (unexplained % + misc. noise %). An infographic detailing the variance budget for the rolls data set is provided in Figure S4.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A next-generation hyperparameter optimization framework* [Software].

Andreas, E. L., et al. (1989). Thermal and size evolution of sea spray droplets.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). *Algorithms for hyper-parameter optimization* (Vol. 24) [Software].

Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system* [Software].

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B., & Young, G. S. (1996). Bulk parameterization of air-sea fluxes for tropical ocean-global atmosphere coupled-ocean atmosphere response experiment. *Journal of Geophysical Research: Oceans*, *101*(C2), 3747–3764.

Foken, T. (2006). 50 years of the monin–obukhov similarity theory. *Boundary-Layer Meteorology*, *119*(3), 431–447.

Hersbach, H. (2008). *Cmod5. n: A c-band geophysical model function for equivalent neutral wind*. European Centre for Medium-Range Weather Forecasts.

Hersbach, H. (2011). Sea surface roughness and drag coefficient as functions of neutral wind speed. *Journal of Physical Oceanography*, *41*(1), 247–251.

Kaimal, J., Wyngaard, J., Haugen, D., Coté, O., Izumi, Y., Caughey, S., & Readings, C. (1976). Turbulence structure in the convective boundary layer. *Journal of Atmo-*

spheric Sciences, *33*(11), 2152–2169.

Kaimal, J. C., Wyngaard, J., Izumi, Y., & Coté, O. (1972). Spectral characteristics of surface-layer turbulence. *Quarterly Journal of the Royal Meteorological Society*, *98*(417), 563–589.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). *Lightgbm: A highly efficient gradient boosting decision tree* (Vol. 30) [Software].

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2017). *Hyperband: A novel bandit-based approach to hyperparameter optimization* (Vol. 18) (Software No. 1). JMLR. org.

NDBC. (2022). *Data source NDBC and TAO buoy network* [dataset]. (`https://www.ndbc.noaa.gov`, `https://tao.ndbc.noaa.gov/tao/data_download/search_map.shtml` : Accessed: 2022-02-01)

NOAA. (2022a). *Data source for TRITON, RAMA and PIRATA buoy networks. made available by IFREMER* [dataset]. (`http://tds0.ifremer.fr/thredds/catalog/CORIOLIS-OCEANSITES-GDAC-OBS/DATA/catalog.html`: Accessed: 2022-02-01)

NOAA. (2022b). *Data source PAPA buoy network* [dataset]. (`https://www.pmel.noaa.gov/ocs/data-links`: Accessed: 2022-02-01)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). *Scikit-learn: Machine learning in Python* (Vol. 12) [software].

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *Catboost: unbiased boosting with categorical features* (Vol. 31) [Software].

Smith, S. D., Fairall, C. W., Geernaert, G. L., & Hasse, L. (1996). Air-sea fluxes: 25

years of progress. *Boundary-Layer Meteorology*, *78*(3), 247–290.

Stull, R. B. (1988). *An introduction to boundary layer meteorology* (Vol. 13). Springer Science & Business Media.

Wang, C., Mouche, A., Foster, R. C., Vandemark, D., Stopa, J. E., Tandeo, P., ... Chapron, B. (2019). Characteristics of marine atmospheric boundary layer roll vortices from sentinel-1 sar wave mode. In *Igarss 2019-2019 ieee international geoscience and remote sensing symposium* (pp. 7908–7911).

Wang, C., Vandemark, D., Mouche, A., Chapron, B., Li, H., & Foster, R. C. (2020). An assessment of marine atmospheric boundary layer roll detection using sentinel-1 sar data. *Remote Sensing of Environment*, *250*, 112031.

WHOI. (2022). *Data source Stratus and WHOTS buoy networks of the Woods Hole Oceanographic Institution* [dataset]. (`http://uop.whoi.edu/projects/`: Accessed: 2022-02-01)

Young, G. S., Sikora, T. D., & Winstead, N. S. (2000). Inferring marine atmospheric boundary layer properties from spectral characteristics of satellite-borne sar imagery. *Monthly weather review*, *128*(5), 1506–1520.

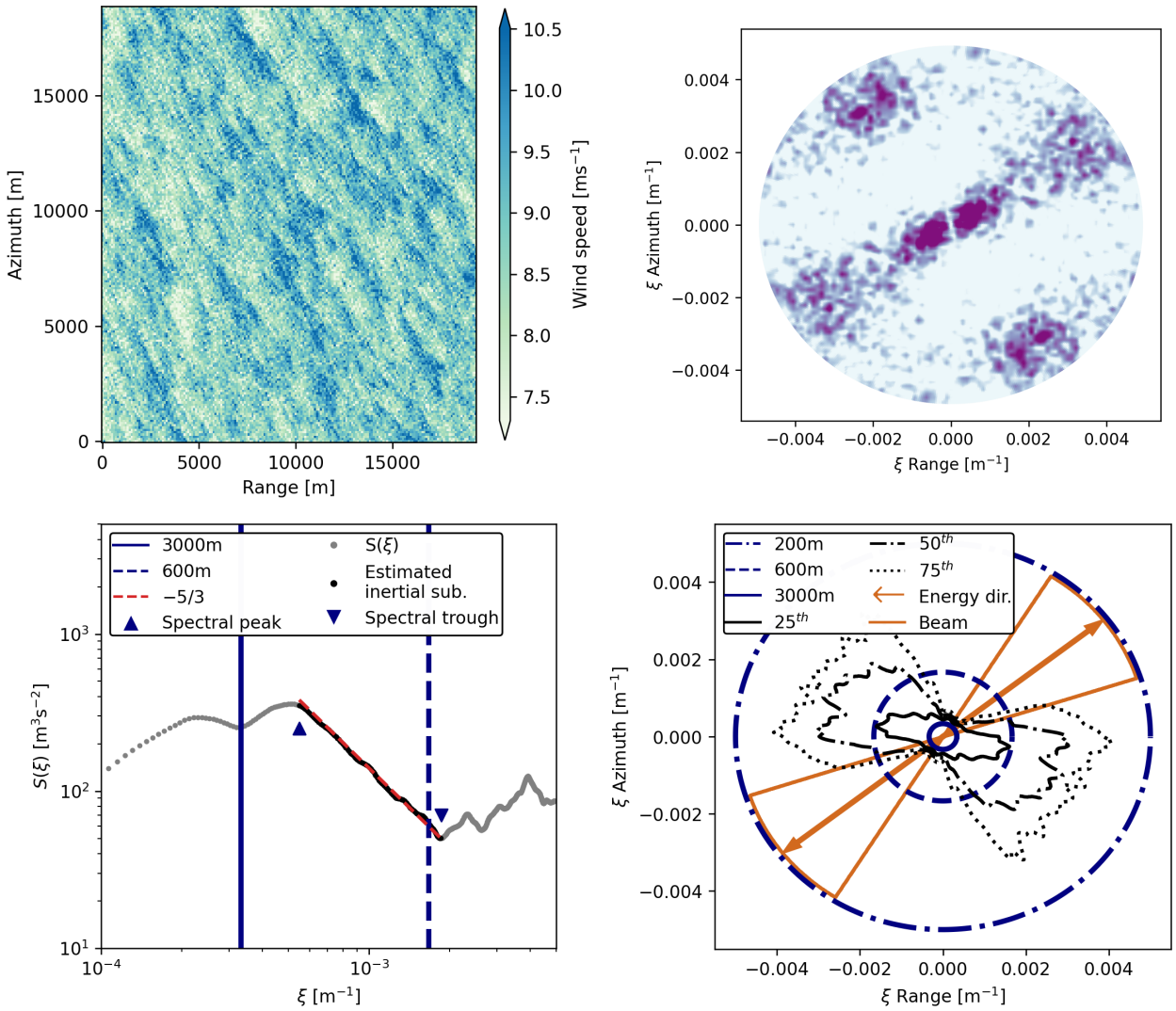

Figure S1. Spectral example of WV scenes classified as containing rolls. Sentinel-1B, date: 07-07-2021, absolute orbit: 027685, vignette : 42, product unique ID: FF3B. (Upper left) Wind field as retrieved from the GMF. (Upper right) Logarithm of wind field's 2D PSD, $S(\theta, \xi)$. Arbitrary units with more purple indicating greater amplitude. (Lower left) 1D integrated spectrum of 2D PSD , $S(\xi)$. (Lower right) Annotations for $S(\theta, \xi)$.

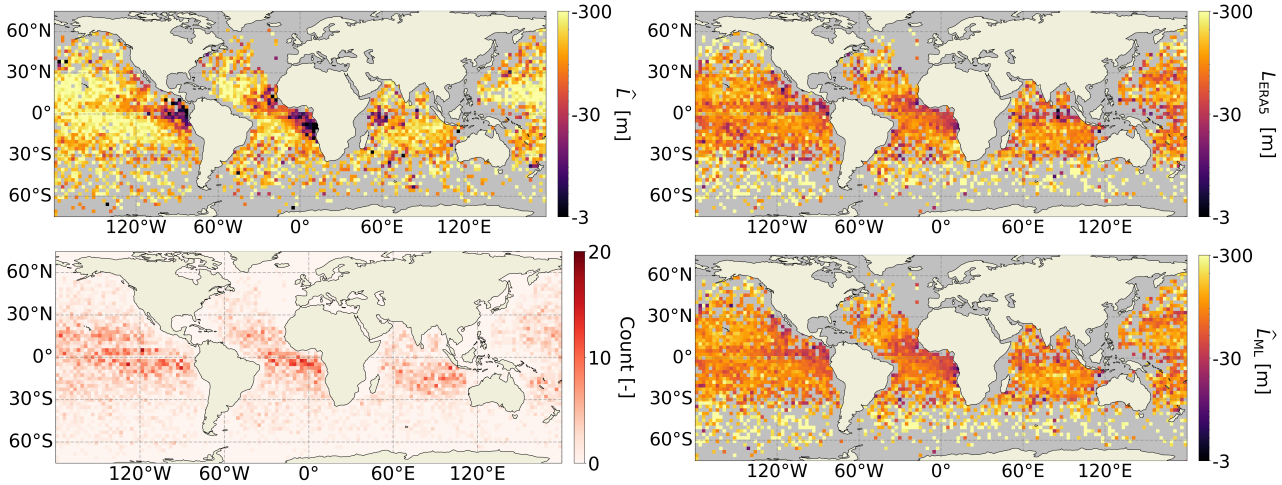

Figure S2. Global median estimation results on $2.5° \times 2.5°$ for the 20% testing data set during 2020-2021. Medians calculated on logarithmic scale. (Upper left) $\hat{L}$ (Upper right) L_{ERA5} (Lower left) Scene count (Lower right) $\hat{L}_{\mathrm{ML}}$.

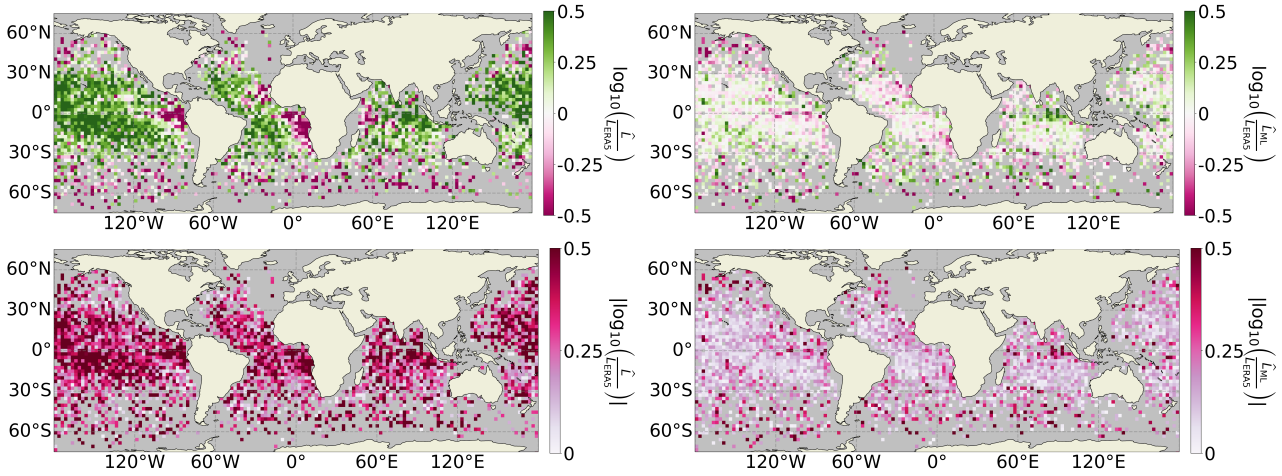

Figure S3. Global median estimation and validation differences on $2.5° \times 2.5°$ for the 20% testing data set during 2020-2021. Medians calculated on logarithmic scale. (Upper left) difference $\hat{L}$ and L_{ERA5}. (Upper right) difference $\hat{L}_{\mathrm{ML}}$ and L_{ERA5} (Lower left) Absolute difference $\hat{L}$ and L_{ERA5}. (Lower right) Absolute difference $\hat{L}_{\mathrm{ML}}$ and L_{ERA5}.

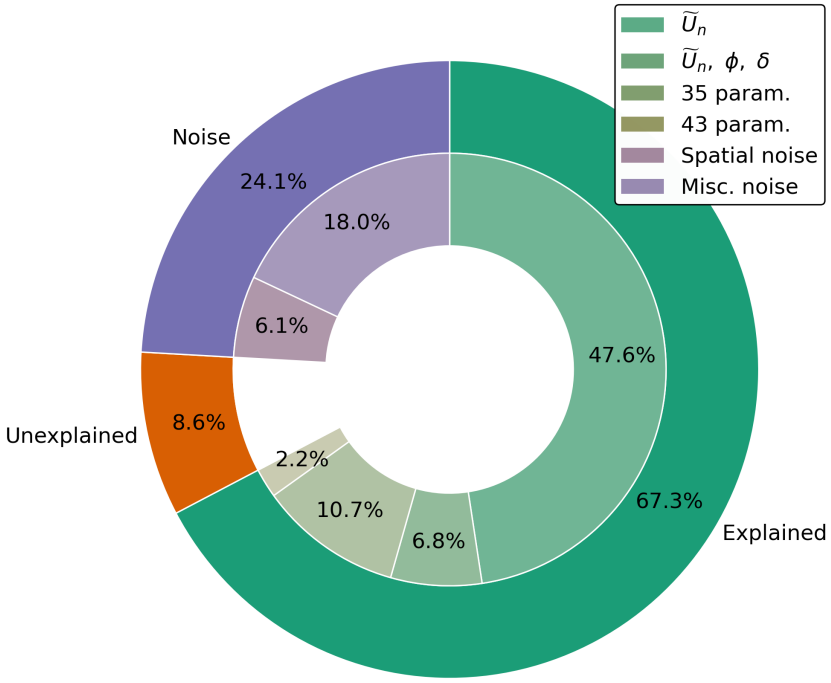

Figure S4. Approximate explainable variance budget for L_{ERA5} estimation on rolls data set. A ML algorithm trained on estimated neutral median wind speed, $\widetilde{U_n}$, is capable of explaining 47.6% of variance. Including two viewing geometry related parameters, ϕ and δ, increases the explained variance by 6.8%-points. Including a further 32 parameters (for a total of 35) derived from two-dimensional spectra adds 10.7%-points. Including eight parameters (for a total of 43) uniquely derived in the analytical method adds a mere 2.2%-points. A comparison between buoys and ERA5 indicates that a minimum of 75.9% of ERA5's L_{ERA5} is physically based. The remaining 24.1%-points are attributed to noise of which approximately a quarter is attributed to spatial decorrelation. The difference between the 75.9% of L_{ERA5} explained by the buoys, and the 67.3% explained by SAR yields 8.6% of unexplained but attainable signal. Results should be seen as ballpark figures representative for the data set of this study. Exact values are likely to vary with the maximum explainable variance being a function of geographical location, data distribution, viewing geometry and or other (unconsidered) factors.

Table S1. Estimation performance between $\log_{10}(|L_x|)$ and $\log_{10}(|L_y|)$ where $\log_{10}(|L_y|)$ is the objective for 2056 observations with both buoy and ERA5 validations. Results are shown for ERA5 validation co-located in space and time with buoy observations rather than SAR acquisitions, allowing for a comparison between buoy and ERA5 validation.

L_x-L_y	L_{buoy}-L_{ERA5}	$L_{buoy_{ML}}$-L_{ERA5}	$\widehat{L}$-L_{ERA5}	$\widehat{L}$-L_{buoy}	$\widehat{L}_{ML_{ERA5}}$-L_{ERA5}	$\widehat{L}_{ML_{buoy}}$-L_{buoy}
R^2	0.258 ± 0.083	0.759 ± 0.047	-1.374 ± 0.219	-0.572 ± 0.134	0.495 ± 0.053	0.396 ± 0.096
MAE	0.202 ± 0.010	0.087 ± 0.007	0.359 ± 0.017	0.259 ± 0.011	0.120 ± 0.014	0.136 ± 0.013

Table S2. Summary of shear algorithm in Text S2.2.1

Step	Equations	Operations	Constants
1	S7	$u_* = \sqrt{C_{dn}\overline{U_n}^2}$	initial guess C_{dn}
2	S10	$z_0 = a\frac{{u_*}^2}{g} + 0.11\frac{\nu}{u_*}$	a, g, ν
3	S9	$C_{dn} = \left[\frac{\kappa}{\ln(z) - ln(z_0)}\right]^2$	z, κ
4		Iterate till convergence	

Table S3. Summary of convective algorithm in Text S2.2.2

Step	Equations	Operations	Constants
1	S5	$n = \xi\overline{U}\cdot\chi$	$\chi = 1$ for first iteration
2	S12	$f_i = \frac{nZ_i}{\overline{U}\cdot\chi}$	Initial estimate Z_i
3	-	$U \rightarrow S(n)$	
4	S11	$w_* = \sqrt{\frac{(2\pi)^{2/3}{f_i}^{2/3}nS(n)\cdot\chi}{\alpha\psi^{2/3}}}$	α, ψ
5	S13	$\overline{w'T_v'} = \frac{{w_*}^3 T_v}{gZ_i}$	Initial estimate of $\overline{T_v}$
6	S14	$L = -\frac{\overline{T}_v u_*^3}{\kappa g\overline{w'T_v'}}$	κ, g, u_* from the shear algorithm
7	S16	$x = \left(1 + 16\left\lvert\frac{z}{L}\right\rvert\right)^{1/4}$	z
8	S17	$\psi_m = \ln\left[\left(\frac{1+x^2}{2}\right)^2\right] - 2\tan^{-1}(x) + \frac{\pi}{2}$	
9	S15	$\chi = 1 - \left(\frac{\psi_m\sqrt{C_{dn}}}{\kappa}\right)$	C_{dn} from the shear algorithm
10		Iterate till convergence	

Table S4. ML performance for scenes whose inertial subrange is assumed to lie in either the microscale (ξ_{peak} [610, 1000] m) or mesoscale (ξ_{peak}: [1500, 2950] m). A maximum of 7910 scenes occurred within the microscale range. An identical number of scenes with peaks in the mesoscale range was chosen to prevent artificially improving performance.

	microscale	mesoscale
R^2	0.667 ± 0.038	0.619 ± 0.045
MAE	0.150 ± 0.007	0.124 ± 0.007

Table S5. ML performance on a logarithmic scale for the angular difference between ERA5 wind direction and radar range direction, δ_{ERA5}. The ML algorithm is retrained per δ_{ERA5}-bin on 400 data points. These data points follow a similar validation distribution across the δ_{ERA5}-bins with 50 L_{ERA5}'s chosen for each of eight logarithmically spaced ranges between -300 m $< L_{\mathrm{ERA5}} <$ -20 m.

δ_{ERA5}	00-18	18-36	36-54	54-72	72-90
R^2	0.573 ± 0.177	0.666 ± 0.054	0.570 ± 0.092	0.554 ± 0.095	0.394 ± 0.137
MAE	0.166 ± 0.037	0.124 ± 0.029	0.156 ± 0.010	0.164 ± 0.018	0.181 ± 0.021