A Method for Estimating and Partitioning the Covariance of Wind Speed Observation Error from GNSS Reflectometry

C. E. POWELL[a] AND C. S. RUF[a]

[a] *University of Michigan, Ann Arbor, Michigan*

(Manuscript received 5 March 2024, in final form 16 February 2025, accepted 17 April 2025)

ABSTRACT: This work introduces a mechanism for identifying and partitioning the sources of correlated error in wind speed observations made by spaceborne Global Navigation Satellite System-Reflectometry (GNSS-R). Observations by the Cyclone GNSS (CYGNSS) constellation of GNSS-R satellites are matched with reanalysis wind speeds to determine the retrieval error. A correction for some of the errors is constructed using ancillary environmental parameters. As a result of the corrections, the wind speed retrieval is improved by 11%. Further, we examine the correlated structure of both the wind speed error and the correction to provide physically plausible explanations for the correlated errors that are present in GNSS-R observations.

SIGNIFICANCE STATEMENT: This work is designed to enhance the utility of GNSS-R data for numerical weather forecasting. Weather models often discard some useful data from observatories because the models are designed to ingest data with completely independent and uncorrelated errors. It is often simpler to thin or average a data stream of observations—to satisfy the assumption of independent error—than it is to modify the data assimilation scheme so it can accommodate correlated errors. This is the case because the true error covariance behavior of the observations is often not well characterized. For GNSS-R, this can significantly impair the utility of the observatory as thinned datasets degrade the representativity of the measurements. This work introduces several innovations to explore how the errors in nearby observations are correlated with one another, with the goal of constructing an observational error covariance model that can be appropriately ingested into numerical weather models.

KEYWORDS: Instrumentation/sensors; Measurements; Remote sensing; Satellite observations; Error analysis; Numerical weather prediction/forecasting

1. Introduction

NASA's Cyclone Global Navigation Satellite System (CYGNSS) is a constellation of eight small satellites that measure ocean surface wind speed using a technique known as Global Navigation Satellite System-Reflectometry (GNSS-R) (Ruf et al. 2018). GNSS-R works by having spaceborne (or airborne) receivers tracking GNSS signals and measuring the power scattered from the surface. In the case of CYGNSS, each satellite contains a specially designed global positioning system (GPS) receiver known as the delay Doppler mapping instrument that measures both the direct and reflected signals from up to four GPS satellites at any given time (Ruf et al. 2022). CYGNSS then estimates local wind speed by using the received power as a proxy for surface roughness at the point of specular reflection, which is related to surface wind speed through an empirical geophysical modeling function (GMF) (Ruf and Balasubramaniam 2019).

Surface wind speed is a critical measurement for initializing numerical weather prediction (NWP) models. One of the primary use cases for CYGNSS is for measuring surface winds within tropical cyclones, as GPS signals are nearly transparent through clouds and precipitation. Traditional measurements of surface winds within tropical cyclones are challenging, as in situ observations such as buoys are sparse, aircraft observations can be dangerous and expensive, and other satellite measurements can degrade in these conditions (Quartly et al. 1996; Portabella et al. 2012).

As an opportunistic dataset, CYGNSS cannot control where it measures. The sample location is a function of the geometry between a CYGNSS receiver in low-Earth orbit, the topography of Earth's surface, and a GPS satellite in medium-Earth orbit. This architecture results in "tracks" of observations as the specular point streaks across Earth's surface. At an orbital inclination of ±35°, CYGNSS was architected to maximize the likelihood of capturing wind speed measurements near tropical cyclones.

However, the utility of this dataset is degraded by a common process in NWP data assimilation (DA). In modern DA, most satellite observations are thinned or "superobbed" to ensure that samples are independent and contain uncorrelated error (Bauer et al. 2011; Gao et al. 2019; Hoffman 2018). While thinned and superobbed CYGNSS measurements have been shown to improve tropical cyclone forecast skill (Zhang et al. 2017), we assert that this still significantly reduces the potential impact the observatory can provide to the forecasting community.

A common DA framework poses the forecast problem variationally (Lorenc et al. 2000):

Denotes content that is immediately available upon publication as open access.

Corresponding author: C. E. Powell, cepowell@umich.edu

DOI: 10.1175/JTECH-D-24-0031.1

© 2025 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

$$\arg\min_{\mathbf{x}} \mathbf{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]), \quad (1)$$

where $\mathbf{J}(\mathbf{x})$ represents the cost function, $\mathbf{x}$ is the variational argument (i.e., the state vector of geophysical model variables of dimension n), $\mathbf{x}_b$ is the background geophysical variables, $\mathbf{B}$ is the $n \times n$ background error covariance matrix, $\mathbf{y}$ is the observation vector of dimension p, H is the forward model operator which transforms from geophysical space to observation space, and $\mathbf{R}$ is the $p \times p$ observation error covariance matrix.

For observation errors to be represented as independent and uncorrelated, $\mathbf{R}$ would only have nonzero elements along the main diagonal. In contrast, physically representative $\mathbf{R}$ matrices would specify nonzero off-diagonal elements. This work suggests a mechanism to estimate these off-diagonal elements as a function of space and time for GNSS-R. Several authors have shown that accurately estimating observation error correlation can improve forecast skill for other types of measurements (Stewart et al. 2013; Rainwater et al. 2015; Stewart et al. 2008; Simonin et al. 2019; Daley 1992). We desire to extend this improvement to GNSS-R observations of wind speed.

The investigation of correlated observation error has been explored for a number of different observatories including interchannel correlations in infrared (Stewart et al. 2014; Bormann et al. 2016) and microwave sounders (Bormann and Bauer 2010; Campbell et al. 2017). In other studies, spatial error correlations have been explored for Doppler radars (Waller et al. 2016a), geostationary imagers (Waller et al. 2016b), and atmospheric motion vectors (Cordoba et al. 2017). There are several ways to estimate the observation error correlation statistics. One of the most common techniques is by estimating $\mathbf{R}$ which is from observation-minus-background and observation-minus-analysis residuals from assimilated forecasts (Desroziers et al. 2005), which is relatively simple to compute for data that are routinely assimilated into forecasts. Other techniques include applying a "three-cornered hat" approach (Sjoberg et al. 2021; Semane et al. 2022), which requires collocation of three independent datasets, but this technique generally assumes the three datasets have uncorrelated errors. Recently, "bottom-up" approaches of constructing error inventories have been introduced to provide estimates of correlated error structure from first principles (Smith et al. 2021; Yang et al. 2024; Powell et al. 2024b).

This work expands upon the Powell et al. (2024b) instrument correlation model to include the correlated error in CYGNSS wind speed retrieval and applies a "top-down" estimation to assist in partitioning sources of correlated error between the instrument and the combined retrieval and representation errors.

2. Error model construction

a. The CYGNSS observation

The CYGNSS observatory exploits a well-studied relationship that connects a statistical parameter of ocean surface roughness—the mean square slope (MSS)—and local wind stress (Cox and Munk 1954; Brown 1979; Kitaigorodskii 1983; Elfouhaily et al. 1997; Liu et al. 2000; Hwang and Fan 2018). As local wind speed increases, so does the MSS. The MSS is directly related to the CYGNSS measurement parameter, normalized bistatic radar cross section (NBRCS) σ^o by Walsh et al. (1998) and references therein [Clarizia and Ruf (2016)]:

$$\sigma^o(\theta) = \frac{|\Re(\theta)|^2}{\mathrm{MSS}}, \quad (2)$$

where θ is the angle of incidence of the scattering geometry and $\Re(\theta)$ is the Fresnel reflection coefficient for the smooth ocean surface at the specular point. CYGNSS measures the NBRCS through a modified radar-range equation, where the power received by the CYGNSS receiver relates to NBRCS (Gleason 2020):

$$\sigma^o = \frac{P^g 4\pi L^{\mathrm{atm}} L^Z G^Z}{P^Z \zeta^E G^R R^{\mathrm{tot}} \overline{A}}, \quad (3)$$

where σ^o is the NBRCS in normalized units ($\mathrm{m}^2\ \mathrm{m}^{-2}$); P^g is the calibrated power received by the nadir science antenna; L^{atm} is the loss due to attenuation in the atmosphere; G^Z is the CYGNSS zenith antenna gain in the direction of the GPS satellite; P^Z is the power received from the CYGNSS zenith receiver; ζ^E is the specular–zenith ratio for the GPS effective isotropic radiated power as described in Wang et al. (2022); P^Z is the power received from CYGNSS's zenith antenna, representing the direct GPS-CYGNSS signal; G^R is the gain of the CYGNSS nadir science antenna in the direction of the Earth specular point; R^{tot} is the total range loss of the transmission via the Earth specular point; L^Z is the transmission range loss from the GPS source to the CYGNSS zenith antenna (written as a loss in the numerator to distinguish from the R^{tot} range losses); and $\overline{A}$ is the effective scattering area of the specular point. In (3), all parameters on the right-hand side are known for a given observation geometry except p^g and P^Z, which form the measurement input for estimating σ°.

CYGNSS retrieves local wind speed from an empirical GMF, which matches CYGNSS measurements to ERA5 reanalysis ocean surface winds (Clarizia and Ruf 2016; Ruf and Balasubramaniam 2019). This GMF is used operationally as a forward-model lookup table (LUT), mapping wind speed and measurement incidence angle to NBRCS. The relationship is monotonic for a given incidence angle, so an inversion can be performed by interpolating between the tabulated values for a given incidence angle and NBRCS.

b. Definitions

In this section, we introduce a simple model for how wind speed observation errors occur. We follow the language used by Janjić et al. (2018) with minor modifications for purposes of brevity and clarity. We assert that wind speed observation error is simply the difference between the observed and true states of an observation vector $\mathbf{y}$:

$$\mathbf{e}_{\mathrm{obs}} = \mathbf{y}_{\mathrm{obs}} - \mathbf{y}_{\mathrm{true}}, \quad (4)$$

where $\mathbf{e}_{\mathrm{obs}}$ is the observation error, $\mathbf{y}_{\mathrm{obs}}$ is the vector of retrieved wind speed observations for a given measurement of

NBRCS, and $\mathbf{y}_{\text{true}}$ is the vector of true values of wind speed. With this formulation, an $\mathbf{R}$ matrix can be constructed for samples i and j by taking the covariance of $\mathbf{e}_{\text{obs}}$:

$$R(i, j) = \text{cov}[\mathbf{e}_{\text{obs}}(i), \mathbf{e}_{\text{obs}}(j)]. \tag{5}$$

We further assert that the sources of wind speed observation error consist of *instrument error*, *retrieval error*, and *representation error*, which are assumed to be independent and can be summed to produce the total error:

$$\mathbf{e}_{\text{obs}} = \mathbf{e}_{\text{inst}} + \mathbf{e}_{\text{ret}} + \mathbf{e}_{\text{rep}}. \tag{6}$$

Instrument and retrieval errors concern deviations from truth due to the design and operation of the observing system. Instrument errors are inherent in all physical systems because the measurement is never perfectly accurate. This occurs from imperfections in manufacturing, limited knowledge of the operating environment, and deviations from operational assumptions.

Observatories rarely measure the target parameter directly; often these systems take a series of measurements in *measurement space* which represents the physical quantity an instrument measure. These physical quantities—power received, time of flight, or NBRCS in our case— re representative of the target parameter but require a transformation to *observation space* to be understood in terms of geophysical parameters such as temperature, pressure, and wind speed. Retrieval errors can occur in situations where a measurement proxy is used for a target observation and the mapping between measurement space and observation space is not one-to-one or unique.

Representation error occurs when the instrument and its reference measurement do not correspond to identical versions of the geophysical parameter. For example, they may represent different spatial or temporal averages of the parameter as discussed in Janjić, or they may be made at different times or locations as explored in Powell et al. (2024a). Representation error can vary depending on the choice of "ground truth" reference.

We can define these error types with some mathematical formalism to illustrate the relationships between them. Instrument error occurs because the observing system is not measuring the true measurement space parameter:

$$\mathbf{e}_{\text{inst}} = H_{\text{ret}}(\mathbf{y}_{\text{meas,obs}}) - H_{\text{ret}}(\mathbf{y}_{\text{meas,true}}), \tag{7}$$

where $\mathbf{y}_{\text{meas,obs}}$ is the actual measurement vector (in our case, the measurement of NBRCS), $\mathbf{y}_{\text{meas,true}}$ is the true value of the measurement vector (i.e., the measurement of NBRCS if instrument calibration had been perfect and the measurements were noise-free), and H_{ret} is the retrieval operator that transforms measurement vectors to wind speed observational vectors. Using similar logic, we define retrieval error as

$$\begin{aligned}\mathbf{e}_{\text{ret}} &= H_{\text{ret}}(\mathbf{y}_{\text{meas,true}}) - H_{\text{ret,perf}}(\mathbf{y}_{\text{meas,true}}) \\ &\cong . H_{\text{ret}}(\mathbf{y}_{\text{meas,obs}}) - H_{\text{ret,perf}}(\mathbf{y}_{\text{meas,obs}}),\end{aligned} \tag{8}$$

where H_{ret} is the actual retrieval operator and $H_{\text{ret,perf}}$ is a perfect retrieval operator (i.e., it maps an error-free measurement vector to the true geophysical variable vector with no ambiguity or retrieval error). For simplicity, we assume that the retrieval error is approximately the same when actual measurements are used in place of error-free measurements. As this formulation suggests, the actual retrieval operator, H_{ret}, is imperfect and includes retrieval errors. Additionally, (8) implies that instrument and retrieval errors are independent. In practice, this assumption may not always hold and imposes conditions on the sensitivity of the retrieval to instrumental errors. For our illustrative exercise, we consider these errors to be independent and partitionable.

Because the representation error is fundamentally comparative, it is challenging to generalize. For instance, a satellite observation compared against a buoy will have different representation errors than when compared to another type of satellite observation or a model output. Further, in our case, representation errors and retrieval errors may be explicitly related, as the training data for the empirical CYGNSS retrieval are ERA5 reanalysis matchups (Ruf and Balasubramaniam 2019). Because of this, we will treat the retrieval and representation errors as a combined entity $\mathbf{e}_{\text{rr}}$:

$$\mathbf{e}_{\text{rr}} = \mathbf{e}_{\text{ret}} + \mathbf{e}_{\text{rep}}. \tag{9}$$

Wind speed observation error now becomes

$$\mathbf{e}_{\text{obs}} = \mathbf{e}_{\text{inst}} + \mathbf{e}_{\text{rr}}. \tag{10}$$

In this construction, we assert that $\mathbf{e}_{\text{inst}}$ and $\mathbf{e}_{\text{rr}}$ are the separable and independent components of wind speed observation error.

c. *Wind speed observation error covariance model*

This work considers how the wind speed observation error varies in space and time, in particular how its correlation varies with separation. We make two general assumptions: first, that the representation–retrieval error $\mathbf{e}_{\text{rr}}$ is isotropic in space and can be parameterized by spatial distance between two samples i and j; and second, that the instrument error $\mathbf{e}_{\text{inst}}$ is isotropic in time and can be parameterized by temporal distance between two samples i and j.

These assumptions warrant discussion. The retrieval component of $\mathbf{e}_{\text{rr}}$ signifies all of the reasons CYGNSS may be sensitive to variations in NBRCS *other* than locally driven winds. For instance, the CYGNSS wind speed retrieval may be incorrect when the satellite samples areas of ocean surfaces covered by oil slicks, biomass, or some other contaminant or when a significant longwave swell modulates the overall roughness condition or any host of other possible reasons why the mapping between surface roughness and wind speed is not as expected. The representation component of $\mathbf{e}_{\text{rr}}$ signifies when CYGNSS captures true wind speed dynamics and variability that is not properly resolved by ERA5. In each of these cases, anisotropies may apply locally for various reasons. However, we posit that with a sufficiently large sample size, each constituent source of the combined representation–retrieval error has a characteristic length scale that does not vary significantly in the aggregate. The characteristic length scale of $\mathbf{e}_{\text{rr}}$ is later shown to be of order 100s of km, which

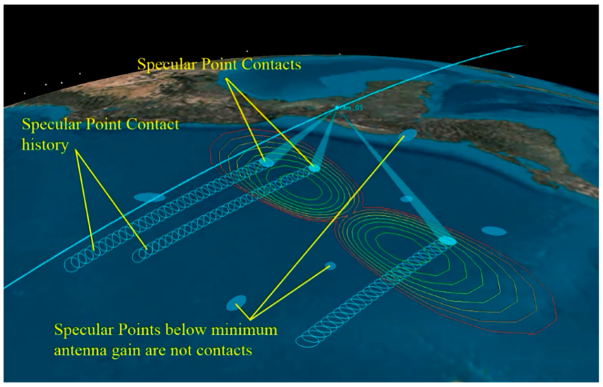


FIG. 1. Illustration of the specular reflection measurement geometry. Measurements are made at the point of contact, when a specular point enters the sensor's antenna pattern (shown as concentric yellow ellipses with border in red). The time history of contacts creates the track of measurements corresponding to a particular GPS/CYGNSS pair of satellites.

corresponds to time separations between measurements of 10s of seconds to minutes.

The instrument error $\mathbf{e}_{\text{inst}}$ was modeled by Powell et al. (Powell et al. 2024b) and is primarily a function of time, as the errors relate to variability in the environment of the CYGNSS receiver, evolution of the measurement geometry, and the condition of the GPS transmitter. The nature of these instrument errors is closely related to the sampling properties of the sensor, which is reviewed here. For more detail, refer to Ruf et al. (2022). The signal propagation associated with a single CYGNSS measurement consists of (i) generation of a navigation signal by one of ~30 GPS satellites at an altitude of ~20 200 km traveling at ~3.9 km s^{-1}, (ii) propagation of the signal from the GPS orbit to Earth's surface, (iii) specular reflection by the surface, and (iv) propagation from the surface to one of 8 CYGNSS satellites at an altitude of ~500 km traveling at ~7.6 km s^{-1}. Measurements of NBRCS are made in the immediate vicinity of the point of specular reflection. As the transmitting and receiving satellites move and their relative positions change, the location of the point of specular reflection moves on Earth's surface along a line referred to as the "track" of the measurement. This measurement geometry is depicted graphically in Fig. 1.

A single track consists of the consecutive measurements made by a common GPS/CYGNSS pair. Depending on the locations of the two satellites and on the relative orientations of their velocity vectors, tracks typically vary in length between ~500 and 1500 km, the duration of time from the beginning to the end of the samples taken along a track varies between ~100 and 400 s, and the separation between samples taken at 1 Hz varies between ~2 and 7 km. However, the distance between samples is nearly constant within a track. Therefore, a trackwise mapping between time and space can be calculated:

$$\tau(i, j) = t_1 - t_2 \rightarrow \Delta(i, j) = c_{\text{track}}\tau, \tag{11}$$

where τ is the distance in time between samples i and j at measurement times t_1 and t_2, respectively; Δ is the distance in space on a great circle projection; and c_{track} is the nearly constant speed at which the specular point moves along a track. With this mapping, $\mathbf{e}_{\text{inst}}$ and $\mathbf{e}_{\text{rr}}$ can share a common spatial separation index, and the covarying errors can be expressed as a function of distance between samples.

The combined error covariance takes the form derived in appendix A of Powell et al. (2024b):

$$R_{\text{mod}}(i, j) = R_{\text{inst}}(\text{i}, j) + R_{\text{rr}}(i, j), \tag{12}$$

where R_{mod} is the modeled wind speed observation error covariance between samples i and j, R_{inst} is the calculated instrument error covariance between samples i and j, and R_{rr} is the calculated representation–retrieval error covariance between samples i and j. Further,

$$R_{\text{inst}}(i, j) = E_{\text{inst}}^2 K_{\text{inst}}(i, j), \tag{13a}$$

$$R_{\text{rr}}(i, j) = E_{\text{rr}}^2 K_{\text{rr}}(i, j), \tag{13b}$$

where E_{inst} is the error magnitude of the instrumentation error in wind speed observation space, K_{inst} is the instrument error correlation between samples i and j, E_{rr} is the error magnitude of the combined representation–retrieval error, and K_{rr} is the representation–retrieval error correlation between samples i and j. The correlation functions K_n have the range of $[-1, 1]$, and $K_n = 1$ if $i = j$. This construction necessarily implies that the error magnitudes E_{inst} and E_{rr} are stationary with respect to location and that the instrument and representation–retrieval errors are independent.

d. *Correlated instrument errors in wind speed observation space*

Powell et al. (2024b) provide a model for correlated instrument error, but the errors are calculated in measurement space (i.e., in units of NBRCS). To propagate these errors to wind speed observation space (in units of m s^{-1}), the instrument error covariance is passed through the GMF. The instrument error correlation function K_{inst} is the tuned $\tilde{\mathcal{R}}_{\text{mod}}$ from Powell et al. (2024b), which is nearly stationary across tracks and observing conditions. The instrument error magnitude takes the following form:

$$E_{\text{inst}} = \left|\frac{\partial(\text{GMF})}{\partial u}\right|^{-1} \text{std}(\sigma^o), \tag{14}$$

where GMF represents the mapping from wind speed to NBRCS. The error magnitude is therefore sensitive to wind speed, as the GMF itself is strongly dependent on wind speed. This behavior is illustrated for the fully developed seas (FDS) GMF in Fig. 2.

The instrument error model $R_{\text{inst}}(i, j)$ can be calculated as a function of wind speed and incidence angle, as shown in Fig. 3. The error covariance is computed in terms of temporal distance τ between two observations i and j within the same track and then mapped to great circle distance Δ using (11), such that R_{inst} takes the form $R_{\underline{\text{inst}}}(i, \Delta)$, where the magnitude of the error depends on the wind speed and incidence angle at

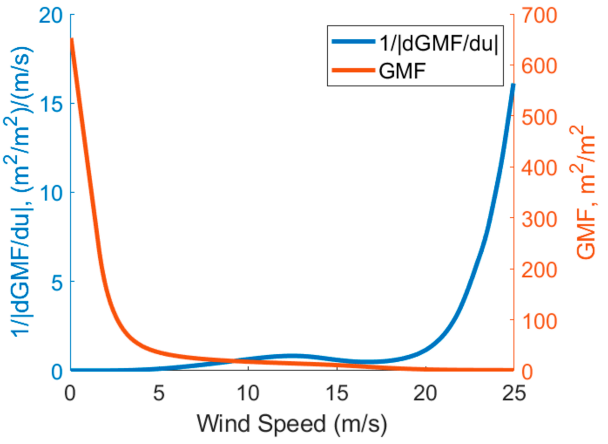


FIG. 2. The GMF (red) and the inverse of the partial derivative of the GMF with respect to wind speed (blue) at a representative incidence angle $\theta = 30°$. Note that as wind speed approaches 0, the NBRCS increases and the sensitivity to errors decreases. At higher wind speeds, the NBRCS approaches 0 and the sensitivity to errors increases significantly. The FDS GMF is valid for wind speeds < 25 m s^{-1}.

sample i, and the decorrelation rolloff is mapped to great circle distance Δ.

e. *Evaluation of representation–retrieval errors*

Direct evaluation of $R_{rr}(i, j)$ is challenging, as both the magnitude and decorrelation rolloff of the representation–retrieval error are not well known. However, we can estimate $R_{rr}(i, j)$ by direct calculation of $R(i, j)$ as in (5) according to

$$R_{rr}(i, j) = R(i, j) - R_{inst}(i, j), \tag{15}$$

where R_{rr} is the representation–retrieval covariance. For calculation of $\mathbf{e}_{obs}$, $\mathbf{y}_{obs}$ is the CYGNSS retrieved fully developed sea wind speed observation at a given point in time and space and $\mathbf{y}_{true}$ is the ERA5 reanalysis 10-m neutral winds matched up via a nearest-neighbor algorithm in space and time. It is well-documented that using reanalysis winds as ground truth is suboptimal, as the reanalysis tends to underestimate high wind speeds and smear out natural variability (Wu et al. 2024), but since the CYGNSS GMF is trained on ERA5, these deficiencies are in part incorporated into the behavior of the CYGNSS wind speed retrievals. For this reason, representation errors associated with ERA5 deficiencies may be underestimated by the procedure used here to quantify it.

The $R(i, j)$ is calculated by taking the autocorrelation along individual tracks such that

$$R(i, j) = R(\Delta) = \sigma^2_{\epsilon,obs}\rho_{\epsilon,obs}(\Delta), \tag{16}$$

where $\rho_{\epsilon,obs}(\Delta)$ is the single-track wind speed observation error autocorrelation at spatial lag Δ and $\sigma_{\epsilon,obs}$ is the standard deviation of the observation error. This construction necessarily assumes that the error is stationary and isotropic. Because we assume stationarity, we incorporate a large number of tracks and calculate a *bulk* observation correlation for M tracks:

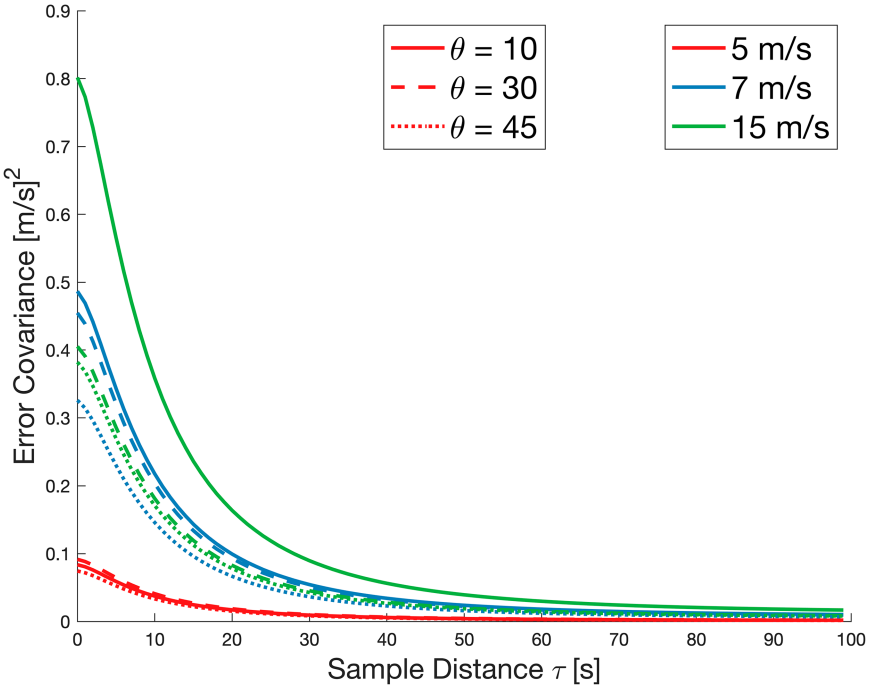


FIG. 3. Contribution to error covariance by propagating instrument error sources into wind speed observation space using the GMF. The colors correspond to different wind speeds: 5 m s^{-1} (red), 7 m s^{-1} (blue), and 15 m s^{-1} (green). The line styles correspond to measurement incidence angles: 10° (solid), 30° (dashed), and 45° (dotted).

$$\mathcal{R}_{\text{obs}}(\Delta) = \frac{1}{M(N-\tau)} \frac{1}{\sigma_{z_i}\sigma_{z_{i+\Delta}}} \sum_{i=1}^{N-\tau} (Z_i - \overline{Z})(Z_{i+\Delta} - \overline{Z}), \quad (17)$$

where $Z_i = \sum_{k=1}^{M} e_{\text{obs}_{i,k}}$, σ_{z_i} is the standard deviation at index i, and k is the index for track number. By substituting $\mathcal{R}(\Delta)$ for $\rho(\Delta)$, R_{rr} can be estimated as

$$R_{\text{rr}}(\Delta) \cong \sigma^2_{\epsilon,\text{obs}} \mathcal{R}_{\text{obs}}(\Delta) - R_{\text{inst}}(\Delta), \quad (18)$$

where all values are a function of spatial great circle distance. This bulk correlation in the wind speed observations is considered to be stationary in the sense that it is applicable generally for measurement tracks in different locations, and it is considered isotropic in space in the sense that it is applicable for measurement tracks in different orientations.

f. Data selection

A month of CYGNSS data was collected (1–31 January 2019) using version 3.1 of the level 2 science data product (CYGNSS 2021). These wind speed observations were filtered to only include tracks with sufficiently long track lengths ($N > 250$ samples) to enable investigation of the spatial and temporal observation error relationships using trackwise autocorrelation. Additionally, CYGNSS observations were filtered to include only wind speeds between 5 and 25 m s^{-1}, which is the domain where the CYGNSS fully developed sea retrieval is most representative. These data were matched with ERA5 hourly reanalysis winds and wave model outputs using a nearest-neighbor algorithm (Hersbach et al. 2020). The ERA5 atmospheric product has a grid spacing of 31 km, while the wave model parameters operate on a 0.36° grid. Both the atmospheric and wave parameters were preprocessed and interpolated for regular latitude–longitude gridding by the Copernicus Climate Data Store upon data access. With the filters applied, a total of 62 497 valid tracks were matched, with global observations ±40° latitude, across all 8 CYGNSS satellites.

3. Partition of representation–retrieval error

In this section, we introduce a methodology for partitioning the representation–retrieval error in a manner that suggests a plausible physical explanation. The processing consists of four main steps:

(i) Identify model variables that correlate with wind speeds and wind speed errors.
(ii) Parameterize wind speed error as a function of a selected model variable.
(iii) Correct the retrieved wind speed with the parameterized functions.
(iv) Repeat this process using corrected wind speed.

By applying successive corrections using this strategy, we can approximate individual components of error. To mitigate the influence of cross correlation between model variables, we use the corrected wind speed as the starting point for the successive correction. The explainable error covariance is then

$$R_{\text{explained},k}(\Delta) \cong R[\boldsymbol{\epsilon}_k](\Delta), \quad (19)$$

where $R_{\text{explained},k}$ is the explainable error covariance for a given correction k as a function of spatial lag Δ, $\boldsymbol{\epsilon}_k$ is the vector of wind speed correction in observation space for a given correction cycle k, and each $R[\mathbf{x}] = \sigma_x^2 \mathcal{R}_x$, where σ_x^2 is the variance of term x and $\mathcal{R}_x$ is the bulk autocorrelation of x as in (17).

a. Selection of correctable variables

Because ERA5 contains a fully coupled wind-wave model, there are a number of model variables that may modulate MSS, local wind speed output, or some combination of the two. In general, it is counterproductive to use variables that are closely related to the target observation—local wind speed in this case—to perform the corrections. This is because we end up "correcting" with another representation of the ground truth, which can decrease the independence of the observed wind speed. We have compiled a list of variables and correlated the matchups with CYGNSS FDS-observed wind speed, ERA5 reanalysis wind speed, and the wind speed error $\mathbf{e}_{\text{obs}}$. Table 1 displays an abbreviated list of 10 ERA5 model variables sorted by the absolute value of wind speed error correlation. This list was curated from an initial analysis of 20 variables to illustrate the considerations for selecting variables for this process.

The variables listed in Table 1 are illustrative of a primary heuristic for determining optimal model variables using this methodology: we are looking primarily for variables that correlate *strongly* with wind speed error $\mathbf{e}_{\text{obs}}$, but *not strongly* with ERA5 wind speed output. For the purposes of this analysis, we pass over any variable that has absolute value correlation coefficient with ERA5 wind speed greater than 0.6. This threshold is primarily designed to avoid using model variables that have nearly one-to-one mappings with CYGNSS ground truth. This would have the effect of correcting errors with ground truth and provide limited insight into the plausible physical basis of error sources. However, this method considers variables with a low level of partial correlation with ground truth, provided it is sufficiently decorrelated that there may exist a plausible physical rationale for the error.

The first variable in Table 1, coefficient of drag with wind waves, is nearly completely correlated with ERA5 wind speed. Upon inspection, the ERA5 wave model calculates this value as a linear function of the neutral wind speed over the ocean, so this high correlation is to be expected. The third item, Charnock parameter, accounts for the aerodynamic roughness of the ocean surface and depends directly on the sea state via wind-induced stress. Because these variables are primarily proxies for ground truth, we do not consider them in our analysis.

The top two remaining variables that correlate most strongly with wind speed error are the mean wave period (MWP) and wave spectral skewness (WSS). Both describe general statistical properties of the ocean surface and are calculated from the two-dimensional wave spectrum produced by the wave model. Note that another parameter is significant wave height (SWH), which is a known error source in the CYGNSS retrieval (Pascual et al. 2021). The version of the CYGNSS wind speed product used for

TABLE 1. Abbreviated list of ERA5 model variables matched up with the CYGNSS wind speed observations. These variables are sorted by the absolute magnitude of correlation with wind speed error $\mathbf{e}_{obs}$. The correlation of the model matchup with CYGNSS-observed and ERA5-modeled wind speed is also displayed. Variables in italics were passed over in analysis because they are too strongly correlated with ERA5 wind speed. Variables in bold are used in this analysis because they are strongly correlated with wind speed error but are sufficiently decorrelated from ERA5 wind speed.

Model variable (ERA5 variable parameter ID)	Correlation with wind speed error $\mathbf{e}_{obs}$	Correlation with CYGNSS wind speed	Correlation with ERA5 wind speed
Coefficient of drag with waves (140233)	*−0.5310*	*0.5725*	*0.9331*
Mean wave period (140232)	**0.4609**	**0.1841**	**−0.1942**
Charnock (148)	*−0.4463*	*0.3324*	*0.6498*
Wave spectral skewness (140207)	**−0.3033**	**0.3897**	**0.5896**
Sig. wave height (140229)	0.1583	0.7767	0.5782
Wave spectral peakedness (140254)	0.1521	0.1569	0.0228
Air seat temp diff [(167) – (34)]	−0.1115	−0.2295	−0.1198
2-m temp (167)	−0.1020	−0.3768	−0.2602
Sea surface temp (34)	−0.0707	−0.3359	−0.2478
Surface pressure (134)	0.0364	−0.1478	−0.1619

this analysis includes a correction for SWH, but there remains some residual correlation with wind speed error.

b. *Generating wind speed correction*

We will start with MWP, which is the average time between two wave crests, inclusive of both wind waves and swell. This variable is strongly correlated with wind speed error but only moderately correlated with ERA5 wind speed. However, because there is some residual dependence on ERA5 wind speed, we develop a correction using a two-step process:

1) Correct for wind speed dependence to generate model variable wind speed anomalies.
2) Parameterize wind speed errors for variable wind speed anomalies with respect to the model.

Both steps can be visualized by depicting the two variables in a two-dimensional histogram, as shown in Fig. 4. Because the MWP appears to be roughly linearly dependent on ERA5 wind speed, a linear regression is computed through the region of highest density ($N > 1 \times 10^4$ counts) to minimize the impact of outliers, as shown in Fig. 4a. This regression provides an expected MWP as a function of wind speed $\lambda_{exp}(\mathbf{y})$. The MWP anomaly is now

$$\lambda_{anom}(y) = \lambda_{exp}(y) - \lambda_{obs}, \tag{20}$$

where λ_{anom} is the MWP anomaly, λ_{exp} is the parameterized expected MWP from the empirical linear regression, and λ_{obs} is the observed MWP for the matchup data. The MWP anomaly is then matched with the wind speed observation error $\mathbf{e}_{obs}$

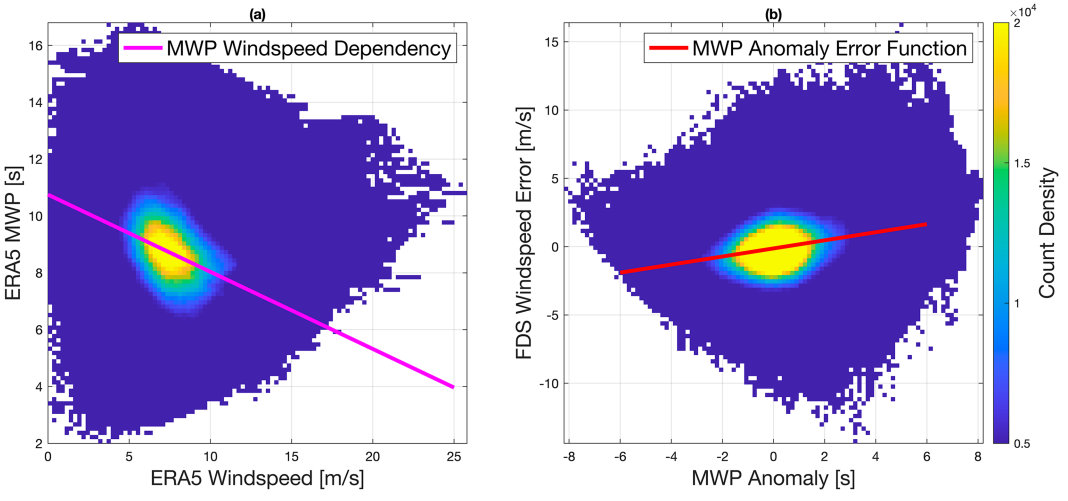


FIG. 4. (a) A two-dimensional histogram of MWP vs ERA5 wind speed. The color shading of the bins represents the frequency of occurrence to illustrate the shape of the overall distribution of wind speed dependence on MWP. A linear regression is computed (magenta) through the region of highest count density ($N > 1 \times 10^4$) to minimize the effects of outliers. This regression represents the wind speed dependence of MWP. The MWP anomaly is calculated as distance from the regression for a given wind speed. (b) A two-dimensional histogram with wind speed error vs MWP anomaly (note the change in axes). A linear regression is computed (red) to estimate the impact of MWP anomaly on WS error.

TABLE 2. List of ERA5 model variables matched up with the MWP-corrected wind speed. These variables are sorted by the absolute magnitude of correlation with MWP-corrected wind speed error $\mathbf{e}_{\text{obs}}$. The correlation of the model matchup with MWP-corrected CYGNSS wind speed and MWP-corrected ERA5 wind speed is also displayed. Variables in italics were passed over in analysis in the first step for being too strongly correlated with uncorrected ERA5 wind speed. Variables in bold are used in this analysis. Note that the MWP correction increases the WSS correlation with ERA5 wind speed because MWP and WSS are related and the correction itself introduces additional wind speed dependence. For the purposes of this analysis, we assert that this does not violate the original rationale for excluding other parameters because we know that WSS is not a robust proxy for ERA5 winds.

Model variable (ERA5 variable name)	Correlation with MWP-corrected wind speed error	Correlation with MWP-corrected CYGNSS wind speed	Correlation with MWP-corrected ERA5 wind speed
Coefficient of drag with waves (140233)	*−0.5943*	*0.6322*	*0.9173*
Charnock (148)	*−0.4877*	*0.3793*	*0.6449*
Wave spectral skewness (140207)	**−0.2433**	**0.5134**	**0.6527**
Mean wave period (140232)	**0.2091**	**−0.0592**	**−0.3886**
Air sea temp diff [(167) − (34)]	−0.1015	−0.2332	−0.112
Surface pressure (134)	0.0605	−0.1463	−0.1499
Significant wave height (140229)	−0.0583	0.656	0.4314
Wave spectral peakedness (140254)	0.0556	0.0748	−0.049
2-m temp (167)	−0.0315	−0.3453	−0.2134
Sea surface temp (34)	0.0058	−0.2985	−0.1977

as shown in Fig. 4b. A similar regression analysis can be performed to create an estimator for wind speed error as a function of MWP anomaly $\mathbf{e}_{\text{obs},\lambda}(\lambda_{\text{anom}})$, which is shown in red.

The MWP-corrected wind speed can be now calculated as

$$\mathbf{y}_{\lambda} = y - \mathbf{e}_{\lambda}[\lambda_{\text{anom}}(\mathbf{y})], \tag{21}$$

where $\mathbf{y}_{\lambda}$ is the MWP-corrected wind speed, $\mathbf{y}$ is a wind speed vector, and $\mathbf{e}_{\lambda}[\lambda_{\text{anom}}(\mathbf{y})]$ is the MWP wind speed error, which is a function of MWP anomaly, which itself is a function of wind speed. This corrected wind speed is now used as the new basis for subsequent corrections. The values in Table 1 can be updated to illustrate how the MWP correction impacts the correlations of other parameters. These new values are shown in Table 2.

A similar process occurs for the successive correction for WSS. WSS is a statistical property of the ocean surface and is a measure of deviation from Gaussianity. Operationally, this variable is used to forecast "freak" waves and is calculated from third-order cumulants of the sea surface elevation in the wave model (ECMWF 2018). While one can expect WSS to have some physical coupling with surface wind speed, it can also be driven by other wave model parameters that impact the ocean surface elevation statistics. We calculate the WSS correction from the MSS-corrected ERA5 wind speed matchup as illustrated in Fig. 5. The only differences from the prior iteration are that we started with the MWP-corrected wind speed and applied a quadratic fit (as opposed to linear) for the WSS anomaly error function based on the shape of the density plot. The parameters of both corrections are displayed in Table 3.

The WSS-corrected wind speed can be calculated in a similar fashion as (21):

$$\mathbf{y}_{\lambda,w} = \mathbf{y}_{\lambda} - \mathbf{e}_{w}[w_{\text{anom}}(\mathbf{y})], \tag{22}$$

where $\mathbf{y}_{\lambda,w}$ is the MWP- and WSS-corrected wind speed, $\mathbf{y}_{\lambda}$ is the MWP-corrected wind speed, $\mathbf{y}$ is a wind speed vector, and $\mathbf{e}_{w}[w_{\text{anom}}(\mathbf{y})]$ is the WSS wind speed error, which is a function of WSS anomaly, which itself is a function of wind speed.

4. Results and discussion

If the parameterized models for wind speed error as functions of MWP and WSS are applied as corrections to improve the wind speed retrieval (for the quality-controlled tracks at 5–25 m s^{-1}), the root-mean-square error decreases from 1.75 m s^{-1} with no corrections to 1.58 m s^{-1} with the MWP correction and further to 1.54 m s^{-1} with both the MWP and WSS corrections. This represents an overall improvement of 11% in the wind speed retrieval.

In addition, we can estimate the covariance structure for each error component individually and compare with the overall wind speed observation error covariance. The individual components of error covariance can be calculated via (17) and (19):

$$R_{\lambda}(\Delta) = \sigma^2_{\epsilon_{\lambda}} R[\boldsymbol{\epsilon}_{\lambda}](\Delta), \tag{23a}$$

$$R_{w}(\Delta) = \sigma^2_{\epsilon_{w}} R[\boldsymbol{\epsilon}_{w}](\Delta), \tag{23b}$$

where R_{λ} is the explainable error covariance due to MWP, R_{w} is the explainable error covariance due to WSS, $\sigma_{\epsilon_{\lambda}}$ is the standard deviation of the MWP correction in wind speed observation space, $\sigma_{\epsilon_{w}}$ is the standard deviation of the WSS correction in observation space, $\mathcal{R}[\boldsymbol{\epsilon}_{\lambda}]$ is the bulk autocorrelation of the MWP correction in observation space as a function of spatial lag Δ, and $\mathcal{R}[\boldsymbol{\epsilon}_{w}]$ is the bulk autocorrelation of the WSS correction in observation space as a function of spatial lag Δ.

The overall wind speed observation error covariance is calculated as suggested in (5) and (19):

$$R_{o}(\Delta) = R[\mathbf{y}_{o} - \mathbf{y}_{\text{ERA}}](\Delta), \tag{24a}$$

$$R_{\text{corr}}(\Delta) = R[\mathbf{y}_{\lambda,w} - \mathbf{y}_{\text{ERA}}](\Delta), \tag{24b}$$

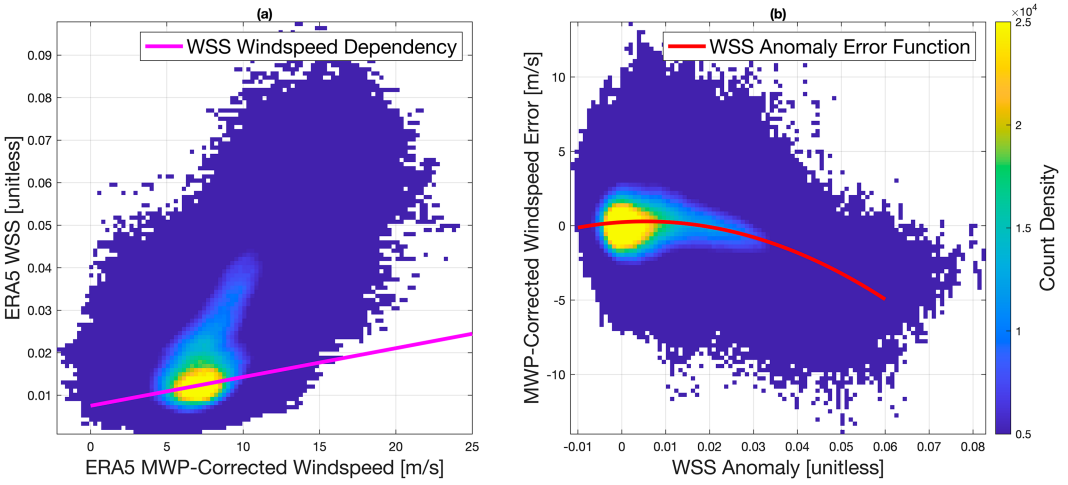


FIG. 5. (a) A two-dimensional histogram with WSS vs MWP-corrected ERA5 wind speed. The color shading of the bins represents the frequency of occurrence to illustrate the shape of the overall distribution of (MWP-corrected) wind speed dependence on WSS. A linear regression is computed (magenta) through the region of highest count density ($N > 2 \times 10^4$) to minimize the effects of outliers. This regression represents the wind speed dependence of WSS. The WSS anomaly is calculated as distance from the regression for a given wind speed. (b) A two-dimensional histogram with MWP-corrected wind speed error vs WSS anomaly (note the change in axes). A quadratic regression is computed (red) to estimate the impact of WSS anomaly on wind speed error.

where R_o is the initial observation error covariance as a function of spatial lag Δ, R_{corr} is the observation error covariance after the MWP and WSS corrections as a function of spatial lag Δ, $\mathbf{y}_o$ is the uncorrected CYGNSS-observed wind speed, $\mathbf{y}_{\lambda,w}$ is the MWP- and WSS-corrected CYGNSS-observed wind speed, and $\mathbf{y}_{\text{ERA}}$ is the ERA5 matched wind speed.

Figure 6 depicts the spatial error covariance for instrument error covariance R_{inst} as defined in (13a) for an exemplar track at the global average wind speed of 7 m s^{-1} and at 30° incidence, partitioned error covariances R_λ and R_w, and the overall observation error covariances R_o and R_{corr}.

Several key features can be identified from Fig. 6. First, the instrument-derived error covariance is generally a small component of the wind speed observation error covariance and decays with a spatial scale size on the order of ~100 km. Second, the error covariance for MWP decays much more slowly in the first ~500 km and then decreases more rapidly toward zero near ~1000 km. This suggests that the mean wave period of the ocean exhibits a high correlation for spatial separations of less than ~500 km and decorrelates at larger separations. The WSS partition is a smaller component of the overall error covariance than MWP and it decorrelates more rapidly, becoming negligible at spatial lags greater than 500 km. This suggests that wave spectral skewness has a smaller impact on CYGNSS wind speed retrieval error and its spatial correlation scale size is smaller than that of MWP. These differences in characteristic scale sizes and rolloffs provide useful insights for future investigations of ocean surface roughness. Third, both the corrected and uncorrected topline wind speed error covariance decay to zero at ~1000 km.

a. Validation

A key test of validation is determining whether the corresponding improvement in the corrected wind speed observation error covariance is due to the specific partitions identified. To calculate the impact of the corrections on overall error covariance structure, the cross correlation between R_o and R_{corr} needs to be accounted for. We asserted earlier that the error partitions are independent, but R_o and R_{corr} are expected to be highly correlated. The difference in error covariances becomes

$$R_{o-\text{corr}}(\Delta) = R[(\mathbf{y}_o - \mathbf{y}_{\text{ERA}}) - (\mathbf{y}_{\lambda,w} - \mathbf{y}_{\text{ERA}})](\Delta), \tag{25a}$$

$$= R[\mathbf{y}_o - \mathbf{y}_{\lambda,w}](\Delta), \tag{25b}$$

$$= \sigma^2_{y_o - y_{\lambda,w}} \mathcal{R}[\mathbf{y}_o - \mathbf{y}_{\lambda,w}](\Delta), \tag{25c}$$

$$= \sigma^2_{y_o - y_{\lambda,w}} \mathcal{R}[\boldsymbol{\epsilon}_\lambda + \boldsymbol{\epsilon}_w](\Delta), \tag{25d}$$

TABLE 3. List of regression parameters for the corrections applied in this section. Note that the regression coefficients c_i are listed in descending order, where c_i take the form $c_n x^n + c_{(n-1)} x^{(n-1)} + \cdots$.

Abscissa	Ordinate	Types of regression	Regression coefficients
ERA5 wind speed (m s^{-1})	MWP (s)	Linear	(−0.2175, 10.7503)
MWP anomaly (s)	Wind speed error (m s^{-1})	Linear	(0.2957, −0.1286)
MWP-corrected ERA5 wind speed (m s^{-1})	WSS (unitless)	Linear	(6.7731×10^{-4}, 0.0075)
WSS anomaly (unitless)	Wind speed error (m s^{-1})	Quadratic	(-1.7584×10^{-3}, 19.0555, 0.2402)

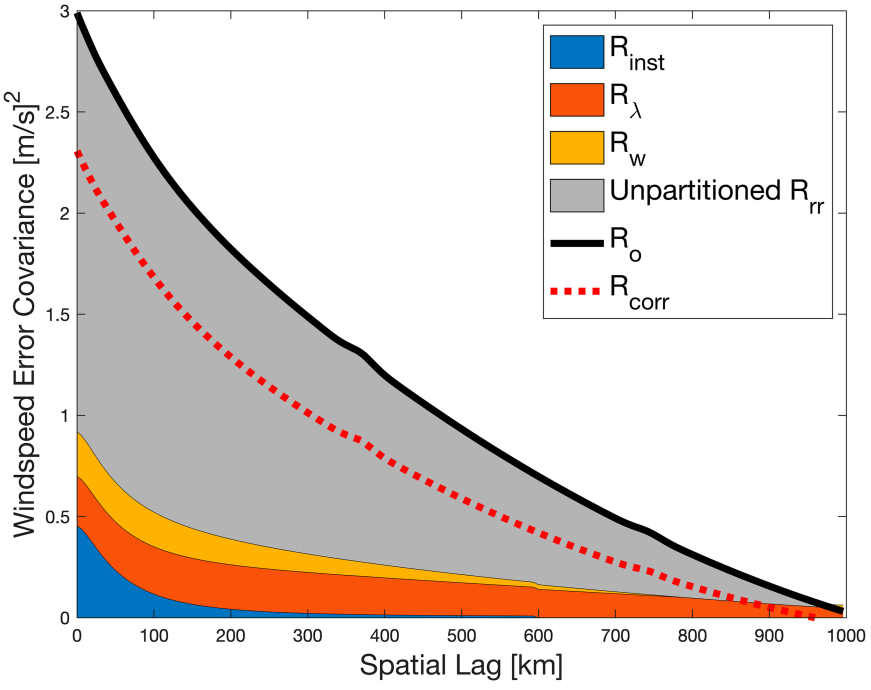


FIG. 6. A stacked area chart illustrating the error covariance partitions from this work. The blue region represents the wind speed observation error covariance from R_{inst} for a reference track at global average wind speed. The orange region represents the observation error covariance due to MWP. The yellow region represents the observation error covariance due to WSS. The three colored regions are defined in the model as independent and are plotted by stacking the contributions upon one another. The gray region is defined as undiagnosed representation–retrieval error in this model and is constructed as the difference between R_o and the sum of the three defined error covariance contributions. The R_o is traced in black and the corrected error covariance R_{corr} is dotted in red, demonstrating the across-the-board improvement in observation error covariance due to the corrections.

where $R_{o-\text{corr}}$ is the difference in covariance between the initial and corrected wind speed errors, and the variance of the difference in wind speeds $\sigma^2_{y_o - y_{\lambda,w}}$ is computed by

$$\sigma^2_{y_o - y_{\lambda,w}} = \sigma^2_{y_o} + \sigma^2_{y_{\lambda,w}} - 2\rho_{o-c}(\sigma_{y_o}\sigma_{y_{\lambda,w}}). \tag{26}$$

In (26), σ_{y_o} is the standard deviation of uncorrected wind speed, $\sigma_{y_{\lambda,w}}$ is the standard deviation of the corrected wind speed, and ρ_{o-c} is the correlation coefficient between the uncorrected and corrected wind speeds. The error covariance difference $R_{o-\text{corr}}(\Delta)$ is plotted together with R_λ and R_w in Fig. 7.

Figure 7 suggests that the improvement in wind speed error due to the corrections, and the resulting impacts to the shape of the observation error covariance, nearly perfectly corresponds with the sum of the individual error covariance partitions. Further, this plot also suggests that the process of successive corrections mitigates enough residual correlation between error terms that the correction terms can be considered independent between one another.

b. Limitations

This work is presented with several caveats. The first is the assumption that wind speed observation errors can be partitioned into several constituent parts, maintaining independence such that the magnitudes can be summed together. The second is that this assumption holds true as when these partitions are correlated and mapped to a common argument, such as great circle distance.

The authors acknowledge the pitfalls associated with using ERA5 wind speed measurements as "truth." ERA5's grid spacing averages out the dynamics of many important subscale phenomenologies, resulting in an artificially slower decorrelation rolloff. The variations in CYGNSS observations may very well be real differences in wind speed.

As seen in Fig. 6, a significant amount of representation–retrieval error remains undiagnosed. While this work can begin to partition certain components of $\mathbf{e}_{\text{rr}}$, it is unlikely that this method will be able to completely partition wind speed observation error until the reference ground truth is matched to the observation appropriately in scale and representation. Matching disparate datasets for commensurate representation is a much larger topic that is outside the scope of this work, but this limitation suggests an opportunity to begin quantifying the error structures of representation error.

This analysis also makes several assumptions about the isotropy and distribution of wind speed observation error. These

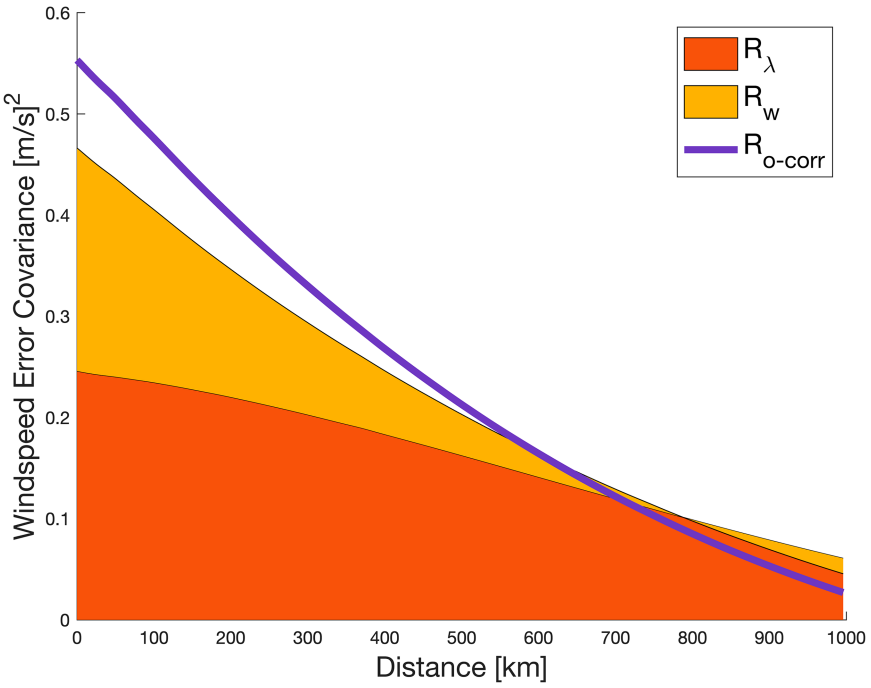


FIG. 7. A stacked area chart illustrating the validation test for the error partitions. The orange region represents the wind speed observation error covariance due to MWP. The yellow region represents the observation error covariance due to WSS. The error covariance due to the correction $R_{o-\mathrm{corr}}$ is plotted in purple. The overall shape and magnitude between the combined MWP and WSS error covariance and the calculated $R_{o-\mathrm{corr}}$ are reasonably consistent, which suggests that the retrieval improvement is the result of the correction of these specific error sources.

assumptions are enabled by sufficiently large datasets to characterize broad statistical behavior of observation errors, but these relationships may change based on location, season, or other confounding variables.

5. Concluding remarks

The results presented here demonstrate a new approach to the construction of a correlated observation error model that builds upon that of previous work. Previous "top-down" approaches have used comparisons between observations and background or analysis estimates of geophysical variables (Desroziers et al. 2005) or between triplets of collocated observations of the variables (Sjoberg et al. 2021; Semane et al. 2022) to estimate the covariance of the observations. Alternate "bottom-up" approaches have attempted to determine the observational covariance from a first principle analysis of the root causes of the correlation (Smith et al. 2021; Yang et al. 2024). Our approach combines these two options by starting with a bottom-up determination of the component of observation error correlation due to instrumental errors (Powell et al. 2024b), determining the components of correlation due to representation and retrieval error using a top-down analysis, and then constructing a composite error covariance model that combines the three components.

Some important "lessons learned" from this study include the variability of correlation scale size and the impact on the final covariance model of the choice of reference, or ground truth, observations. Regarding scale sizes, our results indicate that there are large variations in the range of times and distances over which the different sources of observational error are correlated. In our particular case, instrumental errors were found to decorrelate more rapidly than most sources of representation and retrieval error, and the composite covariance model that was constructed incorporates each of their individual scale sizes. These results are particular to the correlation behavior of the CYGNSS instrumental errors, and other types of sensors may well have very different scales for their error correlation. The important lesson learned is to allow for differences in the scales of the different error sources by constructing a composite covariance model that can incorporate them appropriately. Regarding the dependence of the final covariance model on the choice of reference observations used during the top-down portion of its construction, we chose as our reference the ERA5 model wind speed and recognize that this requires certain caveats to be placed on the accuracy of the derived covariance model. The caveats include both the known deficiencies with ERA5 (e.g., spatial smoothing of finer wind speed variations and increased uncertainty in high wind conditions) and the correlation in errors between it and the CYGNSS wind speed observations due to the use of ERA5 to derive the GMF used by its retrieval algorithm. A more robust covariance model could be constructed using either some other, independent source of reference

observations or possibly by employing a triple collocation approach in which multiple sources of reference observations are used.

The primary contribution of this study was in the development of a process by which the covariance of different error sources, derived using both bottom-up and top-down methods, can be combined together into a composite covariance model that retains and incorporates each of their individual correlation properties. Additional research is needed for improvements in each of the component error sources to improve the accuracy of the composite covariance model. In particular, some, if not most, of the limitations presented by relying on ERA5 as the reference observation could be circumvented by a more judicious choice of reference(s).

In terms of next steps to examine the utility of this new covariance model, some controlled data assimilation (DA) studies are probably in order. Possible elements of such a study would include the following: 1) a DA control run in which the CYGNSS wind speed observations are assimilated under the assumption that its errors are uncorrelated (i.e., the **R** covariance matrix is diagonal). Standard superobbing or data thinning techniques would likely be required; 2) a modification to the control run in which the extent of data thinning is related to the predominant scale size of the composite covariance model. This test would examine the hypothesis that the optimal extent of data thinning is related to the scale size of the correlation between observations. If this is shown to be the case, the composite covariance model could provide thinning guidance to traditional DA schemes that currently tend to use quite ad hoc means for determining the degree of thinning they use; and 3) an upgraded version of DA that makes full use of the composite covariance model by modifying its cost function [**J** in (1)] to support the more general, nondiagonal, observational covariance model. A test for the efficacy of the composite covariance model would compare the results of these three DA approaches.

While the construction of the composite covariance model presented here is specific to CYGNSS observations, elements of the process used are generalizable to other types of observations. Construction of the instrument error component of the model required a detailed, engineering-based analysis of the instrument design. This applies especially to whatever techniques or targets are used to calibrate the data. The engineering analysis performed for CYGNSS went well beyond what is typically done to verify the performance of an instrument, which usually consists of determining the measurement uncertainties (bias and random components) at one point in time and space. To construct similar covariance models for other types of observations, it is strongly recommended that they also perform such a covariance analysis of engineering performance. With regard to the representation and retrieval components of the covariance model, the approach followed in this study should be generally applicable, but would need to be customized appropriately to other types of observations. For example, a correlation analysis was performed to identify which ancillary geophysical parameters had variations that most closely correlated with the variations in observation error. This general approach should still work, but other parameters are likely to correlate the most with the observation error, and finally, the choice of source or sources of "ground truth" reference observations used should be made with due consideration for the retrieval algorithm used by the observation being characterized. In addition to providing accurate estimates of the true geophysical variable, prudent choices for the reference observation should have as little correlation as possible between its errors and those of the observation under test.

Acknowledgments. This work was supported in part by the National Aeronautics and Space Administration Science Mission Directorate under Contract NNL13AQ00C. This work is the authors' own and does not represent the views of any department or agency of the United States Government, including the U.S. Department of Commerce or the National Oceanic and Atmospheric Administration. The authors thank the reviewers for their time and feedback, whose contributions vastly improved the content and clarity of this work.

Data availability statement. All data used in this work are publicly available. The CYGNSS v3.1 science data record is available from the NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC) (CYGNSS 2021). ERA5 reanalysis is publicly available through the Copernicus Climate Data Store [Copernicus Climate Change Service (C3S) Climate Data Store (CDS) 2023].

REFERENCES

Bauer, P., R. Buizza, C. Cardinali, and J. N. Thépaut, 2011: Impact of singular-vector-based satellite data thinning on NWP. *Quart. J. Roy. Meteor. Soc.*, **137**, 286–302, https://doi.org/10.1002/qj.733.

Bormann, N., and P. Bauer, 2010: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Quart. J. Roy. Meteor. Soc.*, **136**, 1036–1050, https://doi.org/10.1002/qj.616.

——, M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767–1780, https://doi.org/10.1002/qj.2774.

Brown, G. S., 1979: Surface roughness slope density estimates for low sea state conditions. *J. Geophys. Res.*, **84**, 3987–3989, https://doi.org/10.1029/JB084iB08p03987.

Campbell, W. F., E. A. Satterfield, B. Ruston, and N. L. Baker, 2017: Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Mon. Wea. Rev.*, **145**, 1019–1032, https://doi.org/10.1175/MWR-D-16-0240.1.

Clarizia, M. P., and C. S. Ruf, 2016: Wind speed retrieval algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) mission. *IEEE Trans. Geosci. Remote Sens.*, **54**, 4419–4432, https://doi.org/10.1109/TGRS.2016.2541343.

Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023: ERA5 hourly data on single levels from 1940 to present. Accessed 3 March 2024, https://doi.org/10.24381/cds.adbb2d47.

Cordoba, M., S. L. Dance, G. A. Kelly, N. K. Nichols, and J. A. Waller, 2017: Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **143**, 333–341, https://doi.org/10.1002/qj.2925.

Cox, C., and W. Munk, 1954: Measurement of the roughness of the sea surface from photographs of the sun's glitter. *J. Opt. Soc. Amer.*, **44**, 838–850, https://doi.org/10.1364/JOSA.44.000838.

CYGNSS, 2021: CYGNSS Level 1 science data record version 3.1. NASA Physical Oceanography Distributed Active Archive Center, accessed 25 January 2024, https://doi.org/10.5067/CYGNS-L1X31.

Daley, R., 1992: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 164–177, https://doi.org/10.1175/1520-0493(1992)120<0164:TEOSCO>2.0.CO;2.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

ECMWF, 2018: IFS documentation Cy45r1 - Part VII: ECMWF wave model. Cy45r1 Tech. Doc., 99 pp., https://www.ecmwf.int/en/elibrary/80898-ifs-documentation-cy45r1-part-vii-ecmwf-wave-model.

Elfouhaily, T., B. Chapron, K. Katsaros, and D. Vandemark, 1997: A unified directional spectrum for long and short wind-driven waves. *J. Geophys. Res.*, **102**, 15 781–15 796, https://doi.org/10.1029/97JC00467.

Gao, Y., H. Xiao, D. Jiang, Q. Wan, P. W. Chan, K. Kwong Hon, and G. Deng, 2019: Impacts of thinning aircraft observations on data assimilation and its prediction during Typhoon Nida (2016). *Atmosphere*, **10**, 754, https://doi.org/10.3390/atmos10120754.

Gleason, S., 2020: Level 1B DDM calibration algorithm theoretical basis document (Rev. 3). Algorithm Theoretical Basis Doc. 148-0137, 37 pp., https://cygnss.engin.umich.edu/wp-content/uploads/sites/534/2021/07/148-0137_ATBD-L1B-DDM-Calibration_R3_release.pdf.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoffman, R. N., 2018: The effect of thinning and superobservations in a simple one-dimensional data analysis with mischaracterized error. *Mon. Wea. Rev.*, **146**, 1181–1195, https://doi.org/10.1175/MWR-D-17-0363.1.

Hwang, P. A., and Y. Fan, 2018: Low-frequency mean square slopes and dominant wave spectral properties: Toward tropical cyclone remote sensing. *IEEE Trans. Geosci. Remote Sens.*, **56**, 7359–7368, https://doi.org/10.1109/TGRS.2018.2850969.

Janjić, T., and Coauthors, 2018: On the representation error in data assimilation. *Quart. J. Roy. Meteor. Soc.*, **144**, 1257–1278, https://doi.org/10.1002/qj.3130.

Kitaigorodskii, S. A., 1983: On the theory of the equilibrium range in the spectrum of wind-generated gravity waves. *J. Phys. Oceanogr.*, **13**, 816–827, https://doi.org/10.1175/1520-0485(1983)013<0816:OTTOTE>2.0.CO;2.

Liu, Y., M.-Y. Su, X.-H. Yan, and W. T. Liu, 2000: The mean-square slope of ocean surface waves and its effects on radar backscatter. *J. Atmos. Oceanic Technol.*, **17**, 1092–1105, https://doi.org/10.1175/1520-0426(2000)017<1092:TMSSOO>2.0.CO;2.

Lorenc, A. C., and Coauthors, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012, https://doi.org/10.1002/qj.49712657002.

Pascual, D., M. P. Clarizia, and C. S. Ruf, 2021: Improved CYGNSS wind speed retrieval using significant wave height correction. *Remote Sens.*, **13**, 4313, https://doi.org/10.3390/rs13214313.

Portabella, M., A. Stoffelen, W. Lin, A. Turiel, A. Verhoef, J. Verspeek, and J. Ballabrera-Poy, 2012: Rain effects on ASCAT-retrieved winds: Toward an improved quality control. *IEEE Trans. Geosci. Remote Sens.*, **50**, 2495–2506, https://doi.org/10.1109/TGRS.2012.2185933.

Powell, C. E., C. S. Ruf, S. Gleason, and S. C. R. Rafkin, 2024a: Sampled together: Assessing the value of simultaneous collocated measurements for optimal satellite configurations. *Bull. Amer. Meteor. Soc.*, **105**, E285–E296, https://doi.org/10.1175/BAMS-D-23-0198.1.

——, ——, D. S. McKague, T. Wang, and A. Russel, 2024b: An instrument error correlation model for global navigation satellite system reflectometry. *Remote Sens.*, **16**, 742, https://doi.org/10.3390/rs16050742.

Quartly, G. D., T. H. Guymer, and M. A. Srokosz, 1996: The effects of rain on Topex radar altimeter data. *J. Atmos. Oceanic Technol.*, **13**, 1209–1229, https://doi.org/10.1175/1520-0426(1996)013<1209:TEOROT>2.0.CO;2.

Rainwater, S., C. H. Bishop, and W. F. Campbell, 2015: The benefits of correlated observation errors for small scales. *Quart. J. Roy. Meteor. Soc.*, **141**, 3439–3445, https://doi.org/10.1002/qj.2582.

Ruf, C., and Coauthors, 2022: *CYGNSS Handbook*. 2nd ed. Michigan Publishing Services, 193 pp., https://doi.org/10.3998/mpub.12741920.

Ruf, C. S., and R. Balasubramaniam, 2019: Development of the CYGNSS geophysical model function for wind speed. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **12**, 66–77, https://doi.org/10.1109/JSTARS.2018.2833075.

——, C. Chew, T. Lang, M. G. Morris, K. Nave, A. Ridley, and R. Balasubramaniam, 2018: A new paradigm in Earth environmental monitoring with the CYGNSS small satellite constellation. *Sci. Rep.*, **8**, 8782, https://doi.org/10.1038/s41598-018-27127-4.

Semane, N., R. Anthes, J. Sjoberg, S. Healy, and B. Ruston, 2022: Comparison of Desroziers and three-cornered hat methods for estimating COSMIC-2 bending angle uncertainties. *J. Atmos. Oceanic Technol.*, **39**, 929–939, https://doi.org/10.1175/JTECH-D-21-0175.1.

Simonin, D., J. A. Waller, S. P. Ballard, S. L. Dance, and N. K. Nichols, 2019: A pragmatic strategy for implementing spatially correlated observation errors in an operational system: An application to Doppler radial winds. *Quart. J. Roy. Meteor. Soc.*, **145**, 2772–2790, https://doi.org/10.1002/qj.3592.

Sjoberg, J. P., R. A. Anthes, and T. Rieckh, 2021: The three-cornered hat method for estimating error variances of three or more atmospheric datasets. Part I: Overview and evaluation. *J. Atmos. Oceanic Technol.*, **38**, 555–572, https://doi.org/10.1175/JTECH-D-19-0217.1.

Smith, D., S. E. Hunt, M. Etxaluze, D. Peters, T. Nightingale, J. Mittaz, E. R. Woolliams, and E. Polehampton, 2021: Traceability of the Sentinel-3 SLSTR level-1 infrared radiometric processing. *Remote Sens.*, **13**, 374, https://doi.org/10.3390/rs13030374.

Stewart, L. M., S. L. Dance, and N. K. Nichols, 2008: Correlated observation errors in data assimilation. *Numer. Methods Fluids*, **56**, 1521–1527, https://doi.org/10.1002/fld.1636.

——, ——, and ——, 2013: Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. *Tellus*, **65A**, 19546, https://doi.org/10.3402/tellusa.v65i0.19546.

——, ——, ——, J. R. Eyre, and J. Cameron, 2014: Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1236–1244, https://doi.org/10.1002/qj.2211.

Waller, J. A., S. P. Ballard, S. L. Dance, G. Kelly, N. K. Nichols, and D. Simonin, 2016a: Diagnosing horizontal and interchannel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sens.*, **8**, 581, https://doi.org/10.3390/rs8070581.

——, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard, 2016b: Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Wea. Rev.*, **144**, 3533–3551, https://doi.org/10.1175/MWR-D-15-0340.1.

Walsh, E. J., D. C. Vandemark, C. A. Friehe, S. P. Burns, D. Khelif, R. N. Swift, and J. F. Scott, 1998: Measuring sea surface mean square slope with a 36-GHz scanning radar altimeter. *J. Geophys. Res.*, **103**, 12 587–12 601, https://doi.org/10.1029/97JC02443.

Wang, T., C. S. Ruf, S. Gleason, A. J. O'Brien, D. S. McKague, B. P. Block, and A. Russel, 2022: Dynamic calibration of GPS effective isotropic radiated power for GNSS-reflectometry Earth remote sensing. *IEEE Trans. Geosci. Remote Sens.*, **60**, 5800512, https://doi.org/10.1109/TGRS.2021.3070238.

Wu, L., H. Su, X. Zeng, D. J. Posselt, S. Wong, S. Chen, and A. Stoffelen, 2024: Uncertainty of atmospheric winds in three widely used global reanalysis datasets. *J. Appl. Meteor. Climatol.*, **63**, 165–180, https://doi.org/10.1175/JAMC-D-22-0198.1.

Yang, J. X., and Coauthors, 2024: SatERR: A community error inventory for satellite microwave observation error representation and uncertainty quantification. *Bull. Amer. Meteor. Soc.*, **105** (1) E1–E20, https://doi.org/10.1175/BAMS-D-22-0207.1.

Zhang, S., Z. Pu, D. J. Posselt, and R. Atlas, 2017: Impact of CYGNSS ocean surface wind speeds on numerical simulations of a hurricane in observing system simulation experiments. *J. Atmos. Oceanic Technol.*, **34**, 375–383, https://doi.org/10.1175/JTECH-D-16-0144.1.