ORIGINAL ARTICLE

Field of the paper

Assimilation of GNSS Reflectometry Delay-Doppler Maps with a Two-dimensional Variational Analysis of Global Ocean Surface Winds

Feixiong Huang[1] | James L. Garrison[1] | S. Mark Leidner[2] | Giuseppe Grieco[3] | Ad Stoffelen[3] | Bachir Annane[4] | Ross N. Hoffman[5]

[1]School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN, USA

[2]Atmospheric and Environmental Research, Lexington, MA, USA

[3]Royal Netherlands Meteorological Institute, De Bilt, The Netherlands

[4]Cooperative Institute for Marine and Atmospheric Studies, Miami, FL, USA

[5]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

Correspondence
Feixiong Huang, School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN, USA
Email: huang712@purdue.edu

Funding information
NASA Grant NNX15AU18G; EUMETSAT OSI SAF Visiting Scientist Program

Direct remote sensing observations (e.g., radar backscatter, radiometer brightness temperature, or radio occultation bending angle) are often more effective for use in data assimilation (DA) than the corresponding geophysical retrievals (e.g., ocean surface winds, soil moisture, or atmospheric water vapor). In the particular case of Global Navigation Satellite System Reflectometry (GNSS-R), the lower-level delay-Doppler map (DDM) observable shows a complicated relationship to the ocean surface wind field. Prior studies have demonstrated DA using GNSS-R wind retrievals inferred from DDMs. The complexity of the DDM dependence on winds, however, suggests that the alternative approach of directly ingesting DDM observables into DA systems, without performing a wind retrieval, may be beneficial. We demonstrate assimilation of DDM observables from the NASA Cyclone Global Navigation Satellite System (CYGNSS) mission into global ocean surface wind analyses using a two-dimensional variational analysis method. Bias correction and quality control methods are described. Several models for the required observation error covariance matrix are developed and evaluated, concluding that a diagonal matrix performs as well as a fully populated matrix empirically tuned to a large ensemble of CYGNSS observation data. 10-meter surface winds from the European Centre for Medium-Range Weather Forecasts (ECMWF) operational forecast are used as the background (i.e., prior in the variational analysis). Results are compared to independent scatterometer (ASCAT, OSCAT) winds. For one month (June 2017) of data the root-mean-square difference (RMSD) was reduced from 1.17 to 1.07 m/s and bias from -0.14 to -0.08 m/s for the wind speed at the specular point. Within a 150-km-wide swath along the specular point track, the RMSD was reduced from 1.20 to 1.13 m/s. These RMSD and bias statistics are smaller than other CYGNSS wind products available at this time.

Keywords — GNSS-R, data assimilation, winds

1 | INTRODUCTION

Global Navigation Satellite System Reflectometry (GNSS-R) is a remote sensing technique that uses satellite navigation (GNSS) transmitters as non-cooperative sources of opportunity in a bistatic radar configuration (Zavorotny et al., 2014). GNSS-R observations have been collected using receivers on stationary (Soulat et al., 2004), airborne (Garrison et al., 2002) and orbiting (Gleason et al., 2005; Foti et al., 2015; Ruf et al., 2018) platforms. Ocean surface wind speed is one variable that can be estimated from GNSS-R observations. The Rayleigh criterion indicates that the ocean surface, under most conditions, will appear rough in the L-band wavelength ($\approx$ 20 cm) used by satellite navigation signals. GNSS signals are therefore scattered from a region on the rough ocean surface that is much larger than the first Fresnel zone. That region, surrounding the specular reflection point, is called the glistening zone.

Early spaceborne GNSS-R missions, UK-DMC (Gleason et al., 2005; Clarizia et al., 2009) and TDS-1 (Foti et al., 2015), have successfully demonstrated the feasibility of measuring ocean surface winds from space. The NASA Cyclone Global Navigation Satellite System (CYGNSS) mission, launched in 2016, is a constellation of eight (8) micro-satellites using GNSS-R for sensing ocean surface winds (Ruf et al., 2013). All CYGNSS micro-satellites are in low Earth orbit (LEO) at an inclination of 35°, each capable of measuring 4 simultaneous reflections, providing up to 32 measurements per second between -38° to 38° in latitude. Since precipitation is transparent at L-band frequencies, CYGNSS can give observations in regions experiencing heavy precipitation including mesoscale convective systems and the inner core of tropical cyclones (TCs). Such regions are rarely observed by conventional higher frequency satellite scatterometers, which experience significant rain attenuation and can only observe the surface in between areas of heavy precipitation. Those observations have the potential to improve the understanding of tropical oscillations and the prediction of TCs. Furthermore, the CYGNSS constellation of eight micro-satellites in low-inclination (35°) orbits provides wind observations across the global tropics with a 7 hour mean revisit time, filling the temporal and spatial gaps from conventional microwave instruments, which are mostly in polar orbits (Ruf et al., 2016).

The delay-Doppler map (DDM), generated by cross-correlating the received signal with a replica of the transmitted signal over a range of delays and Doppler frequencies, is the fundamental physical GNSS-R measurement. Many algorithms have been developed to retrieve ocean surface wind speed and other observables from the DDM (Clarizia et al., 2009, 2014; Rodriguez-Alvarez and Garrison, 2016; Clarizia and Ruf, 2016; Clarizia et al., 2018; Clarizia and Ruf, 2017; Huang et al., 2019a; Reynolds et al., 2020; Clarizia and Ruf, 2020). Under nominal operations, the CYGNSS generates DDMs of 17 time delays $\times$ 11 Doppler frequencies in arbitrary units of "counts". At the CYGNSS science operation center (SOC), the DDM counts are first calibrated to units of power (W) and then converted to bistatic radar cross-section (BRCS), resulting in the Level 1 data product (Gleason et al., 2016). Two observables, the normalized bistatic radar cross-section (NBRCS) and leading edge slope (LES), are computed using only a 3×5 delay-Doppler window of the DDM centered around the bin closest to the predicted specular point delay, thereby providing 25 km spatial resolution (Clarizia and Ruf, 2016). 25-km resolution surface wind speeds at specular points (a Level 2 data product) are retrieved using empirically-developed geophysical model functions (GMFs) relating wind speed to the NBRCS and LES (Ruf and Balasubramaniam, 2018). CYGNSS Level 2 wind speed retrievals were found to have an overall RMSD with respect to ECMWF analyses of 1.96 m/s below 20 m/s and an overall RMSD with NOAA P-3 Stepped Frequency Microwave Radiometer (SFMR) wind observations of 6.45 m/s above 20 m/s (Ruf et al., 2018).

CYGNSS data, have the potential for improving NWP analyses and forecasts through data assimilation (DA). Before launch, synthetic retrieved wind speeds were produced by an end-to-end simulator for many DA studies. Assimilating simulated CYGNSS wind products, using the variational analysis method (VAM), into regional NWP analyses for hurricane cases showed the capability to correct the storm position (Leidner et al., 2018). Simulated winds were also assimilated into the Hurricane Weather Research and Forecasting (HWRF) model by a Gridpoint Statistical Interpolation (GSI) analysis system and evaluated by observing system simulation experiments (OSSEs). These results showed that CYGNSS observations could improve the forecast of TCs both in track and intensity (Zhang et al., 2017; Annane et al., 2018). Another DA experiment, based on multiscale tropical weather systems, showed that simulated CYGNSS winds could improve the low-level wind and temperature (Ying and Zhang, 2018). Recent results from assimilating actual CYGNSS winds also showed improvements in forecasts of TC track, intensity, and structure (Cui et al., 2019; Li et al., 2020). A preliminary study of assimilating CYGNSS winds into global NWP models demonstrate CYGNSS winds' capability to provide more detail in the analysis of global tropical surface winds (Leidner et al., 2020).

While the CYGNSS Level 2 retrieved wind speeds have been used in many DA studies, CYGNSS Level 1 DDM power can be assimilated directly, following similar approaches used for radiance (Andersson et al., 1994), radar backscatter, radiometer brightness temperature (Lievens et al., 2017) and radio occultation bending angle (Cucurull et al., 2013). Potential advantages of assimilating Level 1 DDMs, in contrast to Level 2 wind speed retrievals, include the following:

1) The observables, NBRCS and LES, used for the CYGNSS wind retrieval are calculated by assuming the geometries and power parameters for all DDM bins in the 3×5 box are the same. The failure of this assumption can introduce non-geophysical dependence on the observables. Direct assimilation of the DDMs can account for these non-geophysical factors.

2) Direct assimilation of DDMs using a physically-based forward operator can incorporate additional physical factors such as nonlocal components of the wave field.

3) The full DDM contains more information on the ocean reflections over a larger region of the glistening surface than a wind speed retrieval estimated from only a few bins around the specular point.

4) With a larger footprint (≈ 100 km), the assimilation of full DDMs can impact the analysis over a broader area.

5) The CYGNSS specular point moves at a speed of about 6 km/s on the earth's surface, allowing each point on the ocean surface along the track to be observed by more than 15 sequential DDMs. This feature provides a large number of "multi-look" observations and could achieve better accuracy if the observation errors are characterized properly to avoid over-fitting.

Several DA system components are required to successfully assimilate remotely sensed data such as DDMs. First, DDM assimilation requires a forward model for DDM power as a function of the surface wind speeds. A forward operator and Jacobian have been developed, in which the states are wind speeds on a 10-km grid covering the glistening zone (Huang et al., 2020a). This high resolution grid can represent wind speed variation within the large footprint of the full DDM. Second, a bias correction scheme is required since the modeled DDM power computed by the forward operator is sensitive to bias in the estimated power parameters. Third, the error covariance of the DDMs must be estimated.

A comprehensive summary of our method of assimilating CYGNSS DDMs into global NWP analyses is presented in this study. A two-dimensional VAM is used as the DA method. A bias correction method is also described and error characteristics (covariance matrix) of the DDM observation are discussed. One month of CYGNSS DDMs were assimilated using a 20 minute cycle. The background is the short-range forecast of 10-meter near-surface winds of the European Centre for Medium-Range Weather Forecasts (ECMWF) model. DA performance is assessed by comparison with collocated scatterometer winds. We will show two benefits of DDM assimilation: 1) A positive impact on the VAM analyses over a swath at least 150-km wide; 2) The VAM wind vector retrievals at specular points are more accurate than CYGNSS Level 2 winds and other CYGNSS wind products.

The outline of the paper is as follows. The DDM measurement is introduced in section 2. The DA method is presented in section 3. Methods for computing the DDM error covariance matrix are proposed in section 4. Section 5 assesses the DDM assimilation by validation against scatterometer winds and comparison to other CYGNSS wind products. Section 6 discusss the computational efficiency of the DDM assimilation. Conclusion remarks are given in section 7. The Appendix provides details on the development of the DDM covariance matrix model.

2 | GNSS-R DDM MEASUREMENTS

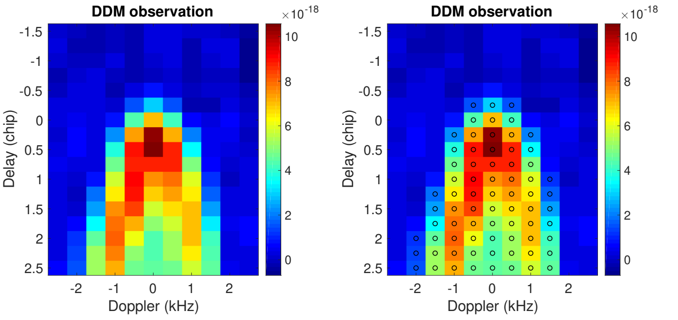
The GNSS-R DDM is calculated by first cross-correlating the reflected signal with a model of the transmitted signal over a range of delays, τ, and Doppler frequencies, f, producing a complex function, $X(t, \tau, f)$. The power of this complex voltage signal is then incoherently averaged to reduce the speckle noise. Bins at (τ, f) known not to contain signal (shorter delay than that through the specular point) are used to estimate the noise floor, Y_n, which is is subtracted from the average, giving

$$Y(t, \tau, f) = \frac{1}{N} \sum_{m=1}^{N} |X(t + (m-1)T_I, \tau, f)|^2 - Y_n \tag{1}$$

$Y(t, \tau, f)$ is calibrated to units of power in the CYGNSS Level 1 product (Gleason et al., 2016). The CYGNSS receiver uses a coherent integration time of $T_I = 1$ ms and averages $N = 1000$ samples, giving an incoherent integration time of 1 sec. CYGNSS DDMs are provided at 17 discrete delays at increments of 0.25 GPS C/A (Coarse Acquisition) code chip (244 ns) and 11 discrete Doppler frequencies at increments of 500 Hz. An example of the Level 1 DDM measurement is shown in Figure 1(a).

The “horseshoe” shape of the DDM represents power reflecting from the glistening zone, with a diameter ranging from 100 to 150 km, depending on the incidence angle and receiver altitude. Each bin of the DDM at a specific (τ, f) is sensitive to reflected power from points on the surface having a total path delay within one code chip and Doppler frequency within 1 kHz of (τ, f). Due to the geometry and delay/Doppler range selected by the receiver, some delay-Doppler bins of the DDM contain little or no information about the surface wind speed. Those observations are not useful for DA and need to be discarded. An empirical method is applied to select informative DDM bins. Only bins with power magnitude larger than 10% of the peak DDM power are selected for use in DA. The informative bins of the DDM in Figure 1(a) are shown in Figure 1(b). All K of the informative DDM bins at one time, t, are grouped into a vector

$$\mathbf{Y}(t) = \begin{bmatrix} Y(t, \tau_1, f_1) \\ Y(t, \tau_2, f_2) \\ \vdots \\ Y(t, \tau_K, f_K) \end{bmatrix} \tag{2}$$



(a) (b)

FIGURE 1 An example of the CYGNSS Level 1 17 × 11 DDM power measurement (a) and DDM informative bins used in DA shown as black circles (b). Units in watt.

which will be used as the observation in DA.

3 | DATA ASSIMILATION METHOD

3.1 | The variational analysis method

This study uses a two-dimensional VAM, based on the surface wind vector field, to assimilate DDMs. This approach was first introduced in Hoffman (1982, 1984) and Hoffman et al. (2003) to resolve scatterometer wind ambiguities and then applied to assimilate satellite wind observations from a large-scale dataset in Atlas et al. (2011). Leidner et al. (2018) used it to add wind direction information to the CYGNSS retrieved wind speed in an OSSE. It was applied to demonstrate DDM assimilation using a few examples in Huang et al. (2020a).

The VAM finds the optimal field of wind vectors, $\mathbf{x}$, that minimizes a cost function

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x}) + J_c(\mathbf{x}) \tag{3}$$

composed of three terms: J_b, representing the difference between the wind field and the background,

$$J_b(\mathbf{x}) = \lambda_b \frac{1}{\sigma_b^2}(\mathbf{x} - \mathbf{x}_b)^T(\mathbf{x} - \mathbf{x}_b), \tag{4}$$

J_o, representing the difference between the wind field and the observation,

$$J_o(\mathbf{x}) = \lambda_{ddm}(\mathbf{h}(\mathbf{x}) - \mathbf{Y})^T\mathbf{R}^{-1}(\mathbf{h}(\mathbf{x}) - \mathbf{Y}), \tag{5}$$

and J_c, the constraint term,

$$J_c(\mathbf{x}) = \lambda_{lap}J_{lap}(\mathbf{x}) + \lambda_{div}J_{div}(\mathbf{x}) + \lambda_{vor}J_{vor}(\mathbf{x}) \tag{6}$$

where x_b is the background wind vector field, σ_b^2 is the standard deviation of the background wind components. $\mathbf{h}()$ is the DDM forward operator. **Y** is the DDM observation. **R** is the observation error covariance matrix. J_{lap}, J_{div} and J_{vor} are the Laplacian, divergence and vorticity of the increment. λ_b, λ_{ddm}, λ_{lap}, λ_{div} and λ_{vor} are the weights of each term. Details for calculation of the constraint terms are given in Hoffman et al. (2003).

In the term J_b, the background error is characterized by a single constant value, σ_b^2. The background error correlations are characterized by the constraint term J_c, which is derived from the Navier-Stokes equations for viscous fluid motion. The combination of J_b and J_c takes the place of the usual background term in a traditional DA variational DA that includes a full background covariance matrix, as explained in Hoffman et al. (2003). Rather than assimilating all DDMs in one cost function, the DDMs on each CYGNSS specular point track are assimilated sequentially to reduce the computation and memory requirements. One DDM will be assimilated at a time and the analysis wind field will be updated after processing each DDM until all observations within the DA cycle have been assimilated. The observation error covariance matrix, **R**, represents the errors and correlations of all DDM bins at the same time (i.e., in one **Y** vector). The matrix **R** will be characterized in section 4. Tuning the background and observation weights λ_b and λ_{ddm} can improve the a priori estimated errors of the background and observation in the VAM as these errors are usually based on limited information (Hoffman et al., 2003). The constraint weights λ_{lap}, λ_{div} and λ_{vor} should be large enough to correctly shape the error correlations of the background wind field. They are set to ensure that the influence of the observations spreads out to the scale of the effective model resolution. The weight values used in the experiments will be specified in section 5.2.

3.2 | DDM forward operator and Jacobian

A numerical forward operator and its Jacobian, which represent the measurement physics, are required in any DA system. In the case of DDM assimilation, the forward operator projects the discrete wind field into the DDM measurement space. The DDM forward operator has been presented in Huang et al. (2020a). It is based on a Kirchoff Approximation and Geometric Optics (KA-GO) surface scattering model (Zavorotny and Voronovich, 2000). The ocean surface slope probability density function (PDF), a key parameter of the KA-GO model, is assumed to be an isotropic normal distribution defined by a single parameter, the omni-directional mean square slope (MSS). An empirical model derived from aircraft experiments (Katzberg et al., 2006) gives a monotonic relationship between MSS and wind speed. Waves driven by nonlocal winds (e.g., swell) are not considered in the forward model, but could be considered in the future if ancillary data such as significant wave height from a wave model were available. The wind field around the specular point within an area of 120 km × 120 km is gridded into 0.125° spacing for input to the forward operator. The forward operator takes in the satellite geometries, transmitter Equivalent Isotropically Radiated Power (EIRP), specular bin indices from the CYGNSS Level 1 product, receiver antenna patterns, as well as the gridded wind field to produce a modeled DDM in the same delay-Doppler coordinates as the measured one. The Jacobian represents the sensitivity of each DDM bin with respect to the wind speed of each surface grid point. It is computed analytically by linearizing the forward operator. Details of the computation in the forward operator and Jacobian are described in Huang et al. (2020a). The assessment of the forward operator in Huang et al. (2020a) shows that it performs well at a certain range of wind speed under adequate bias correction and quality control. In particular, in Huang et al. (2020a), we have assessed that the impact of swell is negligible except under very low wind speeds (e.g., < 2 m/s) and these cases are removed in this study during quality control (QC).

3.3 | Bias correction

It is crucial to have unbiased observations in order to obtain the Best Linear Unbiased Estimator (BLUE) in DA (Bouttier and Courtier, 2002). Bias can arise in the measurement or the forward operator and should be removed before assimilating the observations. The DDM forward operator requires an estimate of the transmitter Effective Isotropic Radiated Power (EIRP) and the receiver antenna patterns for each CYGNSS satellite. The CYGNSS mission uses a ground-based power monitor to estimate the EIRP, which is provided in the Level 1 data (Wang et al., 2019). Receiver antenna patterns are estimated by pre-launch measurements and on-orbit corrections (Gleason et al., 2018). These patterns were made available to us by the CYGNSS project and are distributed as part of the forward model code (Huang et al., 2020b). Previous studies have found bias in the CYGNSS observations which largely resides in the estimated transmitter EIRP with some contribution from the receiver antenna patterns (Ruf et al., 2018; Huang et al., 2019b). In order to remove this bias, we assume that the GPS transmitter EIRP remains constant for all observations along the same CYGNSS specular point track. This is a reasonable assumption, given that the duration of a track is generally less than 20 minutes. This suggests a "track-wise" DDM bias correction scheme, similar to that used by Said et al. (2019) for correcting bias on the retrieved wind speed. In our DA approach, however, a bias correction will be applied to the DDM power.

Our basic assumption is that the background wind field from a global NWP model (e.g., ECMWF) is globally unbiased (Stoffelen and Vogelzang, 2018). Thus, comparing the average of a large sample of measurements against model predictions from a background reference can be used to correct the observation bias. In this scheme, DDMs on a continuous specular point track formed by one specific pair of GPS transmitter and CYGNSS receiver are first identified. Assuming both the transmitter EIRP and uncertainty in the receiver antenna gain patterns are multiplicative error sources, a scaling term is computed as the mean proportion between the M measured DDMs and the corresponding modeled DDM computed from the background along the specular point track.

$$\Phi = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{K_m}\sum_{i=1}^{K_m}\frac{Y_i(t_m)}{h_i(\mathbf{x}, t_m)} \tag{7}$$

where t_m is the time of the m-th DDM; K_m is the number of informative bins of the m-th DDM; $h_i(\mathbf{x}, t_m)$ is the power of the i-th modeled DDM bin at time t_m, computed from the background wind field using the forward operator.

When assimilating DDMs on the track, each modeled DDM from the forward model is multiplied by the scaling term Φ, such that the cost function (5) becomes

$$J_o(\mathbf{x}) = \lambda_{ddm}(\Phi\mathbf{h}(\mathbf{x}) - \mathbf{Y})^T\mathbf{R}^{-1}(\Phi\mathbf{h}(\mathbf{x}) - \mathbf{Y}) \tag{8}$$

Figure 2 shows the specular bin power of DDM observations, DDM and bias-corrected DDM forward model estimates from an example track. The systematic bias between the models and observations is significantly reduced by the bias correction.

3.4 | Quality control

The following QC tests are applied to filter CYGNSS Level 1 DDM observations before DA.

1) The netCDF variable "quality_flags" values in the CYGNSS L1 data are required to be zero. This discards cases in which the observation is over or close to land, the spacecraft has attitude rotation larger than 1°, the transmitter power has a high uncertainty or there are some calibration issues.

2) All data with signal-to-noise ratio (SNR) less than 3 dB are discarded. Small SNR indicates high noise power, making it difficult to extract informative DDM bins.

3) All data with incidence angle larger than 60° are discarded. DDMs observed under large incidence angle can have a glistening zone larger than 120 km × 120 km, which cannot be modeled accurately by the forward operator.

4) All data with background wind speed at the specular point less than 2 m/s or larger than 35 m/s are discarded. The swell at very low wind speed cases and the complicated sea state at very high wind speed cases cannot be modeled well by the

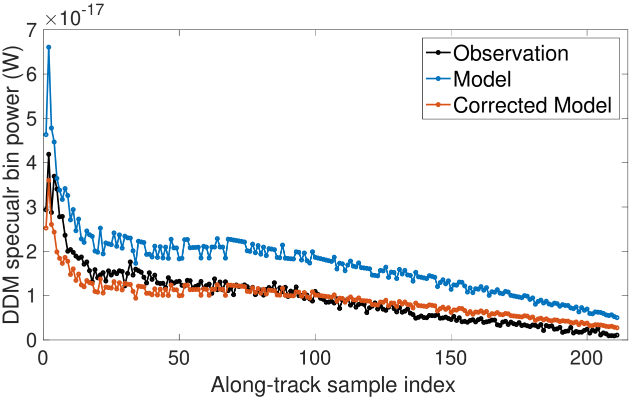


FIGURE 2 Specular bin power (units in watt) of DDM observations (**Y**), DDM forward model estimates (**h**(**x**)) and bias-corrected DDM forward model estimates (Φ**h**(**x**)) from an example track. Observations of the track were collected within 22:22:49 to 22:28:27 UTC on 1 June 2017 with CYGNSS Space Vehicle (SV) 4 and GPS Psuedo Random Noise (PRN) 17.

forward operator (Huang et al., 2020a). The reduced sensitivity of the DDM to high wind speed is also well known.

5) Relative power difference and correlation coefficient between the observed DDM and modeled DDM from the background are used to identify additional observation data quality issues and avoid model representativeness errors. They are discussed in detail in Huang et al. (2020a). Data with relative power difference larger than 100% and correlation coefficient less than 0.9 are discarded.

The QC tests and the yield (percent passing) for each one are summarized in Table 1.

Observation Characteristic	Must be	Yield
CYGNSS L1 “quality_flags” variable	0	25%
SNR	> 3 dB	57%
Incidence angle	$< 60°$	88%
Wind speed at specular points	2–35 m/s	87%
Relative power difference	$< 100\%$	99.9%
Correlation coefficient	> 0.9	99%

TABLE 1 QC tests and yields for the assimilation of CYGNSS DDMs.

4 | ERROR CHARACTERISTICS OF DDM OBSERVATIONS

In addition to unbiased observations, an accurate observation error covariance matrix, **R**, is required for optimal estimation in DA. Observation errors usually include measurement error (error related to the instrument and measurement technique) and representation error (e.g., error related to the forward operator and differences in scales between the observation and the analysis)

(Janjić et al., 2018). This section will only focus on the statistics of the measurement error as the representation error can be accounted by varying the weights λ_{ddm} in equation (5) as described in section 3.1.

As stated earlier, the VAM assimilates one DDM each time and the observation error covariance matrix, $\mathbf{R}$, represents the errors and correlations of all K informative bins in one measured DDM

$$\mathbf{R} = \mathbb{E}\left\{(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\})(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\})^T\right\} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1K} \\ \vdots & \ddots & \vdots \\ \sigma_{K1} & \cdots & \sigma_K^2 \end{bmatrix}. \tag{9}$$

The observation $\mathbf{Y}$ is a vector assumed to follow a Gaussian distribution by the central limit theorem as it is an average value over a large number, N by equation (1).

Measurement error is assumed to come from both background noise and speckle noise. Background noise includes thermal emission from the ocean, correlation of the signal with that from other GNSS transmitters, and receiver thermal noise (Gleason et al., 2019). In this study, the background noise is assumed to be stationary white Gaussian, as the impact of the correlation from ambient signals is negligible, as discussed in Gleason et al. (2019). Speckle is the result of distructive and constructive interference of random scattered signals during the coherent integration time. The background noise is additive while the speckle noise is multiplicative (Gleason et al., 2010). In previous studies, analytical models for second order statistics of the DDM complex voltage signal in the delay dimension, $X(t, \tau, 0)$, were derived by considering both thermal noise and speckle (Martín-Neira et al., 2011; Germain and Ruffini, 2006; Martín et al., 2014; Garrison, 2016). A detailed analytical model of the covariance matrix of the averaged DDM power in the delay dimension was derived and validated using actual data (Li et al., 2018). Analytical models, however, have practical limitations for direct use in DDM assimilation. First, those models require knowledge of the thermal noise statistics (equivalent thermal noise temperature) which is not estimated accurately for the CYGNSS mission. Second, present models only consider the correlations between measurements at different delays, while the correlations in the Doppler dimension and between the delay and Doppler are not characterised. Finally, analytical models require computation of a surface integral and convolution with the Woodward ambiguity function, which is computationally expensive and thus not practical for large scale DA. Another approach often used in NWP applications is to compute the error covariance directly from a large number of observation samples (Desroziers et al., 2005; Waller et al., 2016; Cordoba et al., 2017). This method has a very low computational cost at the expense of requiring a large ensemble of observations with the same error statistics. In the spaceborne GNSS-R application, however, the relatively low sampling frequency (1 Hz for CYGNSS) and high receiver speed (resulting in fast changes in the geometry, antenna gain and observed wind field), limits the set of observations with similar statistics to a number too small to give a good estimation of the covariance matrix.

In this section, two methods to compute the DDM error covariance matrix are proposed. One method assumes it to be a diagonal matrix with error proportional to the observation and another method uses an empirical model which includes the error correlations.

4.1 | Scale method

In the NWP data assimilation, it is common to use a diagonal observation error covariance matrix as the error correlations are generally difficult to estimate. The use of a diagonal matrix $\mathbf{R}$ has simple implementation and low computational cost but may lose information from the observation error correlations (Hoffman, 2018).

Gleason et al. (2016) estimated the error in CYGNSS Level 1 DDM power to be 0.50 dB (12%) and 0.23 dB (5%) for wind speed below and above 20 m/s, respectively, by analyzing each error source in the Level 1 calibration (Table II in Gleason et al. 2016). With this in mind, we simply model the error as proportional to the observation magnitude. We used a constant of proportionality of 10% (in between the two values in (Gleason et al., 2016)) to create a diagonal covariance matrix with

$$R_{ii} = (0.1Y_i)^2 \tag{10}$$

4.2 | An empirical model

In this section, a parametric model for the DDM error covariance matrix, incorporating off-diagonal elements, is empirically developed from a large set of CYGNSS Level 1 observations. We will show that this model provides a good representation of the DDM error statistics with a low computational cost. The Appendix provides a more detailed description of the model development.

In this model, the diagonal elements (variance) and the off-diagonal elements (covariance) of the matrix are modeled separately by parametric fitting to sample covariance matrices computed from actual DDM observations. For each observation at a specific delay-Dopler coordinate of the DDM, $Y(t, \tau_i, f_i)$, the variance is modeled as the sum of that from speckle, $\sigma_{i,s}^2$, and a background noise, σ_n^2, assumed constant and independent of the delay-Dopler coordinate.

$$\sigma_i^2 = \sigma_{i,s}^2 + \sigma_n^2 \tag{11}$$

Speckle noise for a single observation (before averaging) is proportional to the signal magnitude. Modeling variance of the incoherently-averaged observation, $\sigma_{i,s}^2$, however, would require accounting for the correlation between sequential waveforms (Li et al., 2018). We attempted to approximate this with a simpler functional dependence, by assuming a general power law relationship,

$$\sigma_{i,s} = p[i] Y_i^{q[i]}, \tag{12}$$

Coefficients, $p[i]$, and exponents, $q[i]$, are indepdendently estimated for each of the discrete 11×17 delay-Doppler bins, from a large set of data spanning a wide range of surface wind speeds and other conditions.

The off-diagonal elements, σ_{ij}, represent correlation between a pair of bins from the same DDM, at different delay-Doppler coordinates, $(\tau, f)_i$ and $(\tau, f)_j$. This can be normalized to define the correlation coefficient, ρ_{ij}

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}. \tag{13}$$

We have observed that ρ has a dependence on wind speed , which could also be explained by several analytical models (listed in Appendix). An empirical parametric model for the dependence of the correlation coefficient on wind speed is assumed to take the form of

$$\rho_{ij} = a[i,j] + b[i,j]u^{-1} + c[i,j]u^{-2} \tag{14}$$

where u is the background wind speed at the specular point. Please refer to the Appendix for details of the development of the model and computation of the parameters a, b, c, p, q.

Figure 3 presents a typical example for the comparsion of the different covariance matrix models. Figure 3(a) is the DDM observation collected by CYGNSS SV 2 with GPS PRN 20 at 22:58:59 UTC on 1 June 2017. Figure 3(b) is the corresponding diagonal covariance matrix computed by the scale method in section 4.1. Figure 3(c) is the corresponding non-diagonal covariance matrix computed by the empirical model developed in section 4.2. Figure 3(d) is the sample covariance matrix computed from

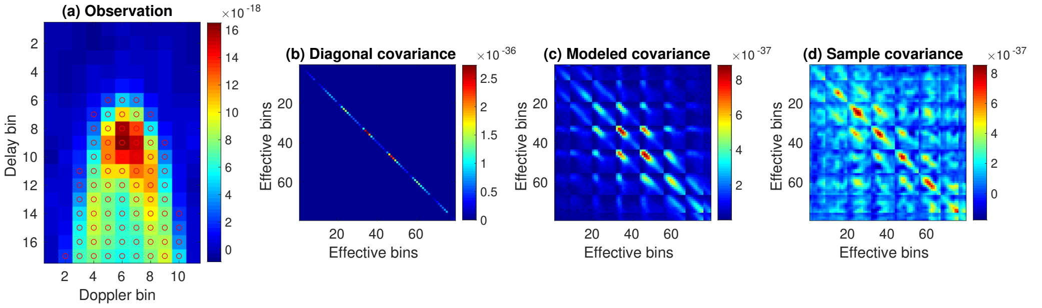


FIGURE 3 Comparison between the modeled DDM error covariance matrices and DDM sample covariance matrix for the CYGNSS mission. (a) DDM observation with informative bins as red circles. (b) Diagonal covariance matrix computed by the scale method in section 4.1. (c) Non-diagonal covariance matrix computed by the model in section 4.2. (d) Sample covariance matrix computed from sequential 25 DDMs.

DDM observations between 23:58:47 and 23:59:11 UTC with CYGNSS SV 2 and GPS PRN 20. Note that the sample covariance matrix can be noisy because it is computed using only 25 samples. It can be observed that the empirical non-diagonal covariance matrix model captures much of the structures of the sample covariance matrix. Note that the covariances in panels (b)–(d) appear patchy because we have to "unroll" the 2D DDM (e.g., panel (a)) into a vector for DA (equation (2)). The covariances presented here are computed for that "unrolled" vector.

The inverse of the covariance matrix, $\mathbf{R}^{-1}$, is required in the VAM cost function (5). It is found that the covariance matrix computed by the empirical model is often ill-conditioned, making it difficult to compute an accurate inverse. Ridge regression (Tabeart et al., 2020), a reconditioning method, is applied to reduce the condition number of the matrix to ~100. This method increases the diagonal values of the matrix by a fixed number and thus will also increase the modeled variances of the observations.

5 | GLOBAL DATA ASSIMILATION RESULTS

5.1 | Data description and experimental design

5.1.1 | CYGNSS DDM observations

CYGNSS version 2.1 Level 1 DDM data from 1 June 2017 to 30 June 2017 were used as observations. Details about the CYGNSS DDM observations were introduced in section 2. Level 1 data also include the transmitter EIRP and satellite geometries, estimated by the CYGNSS SOC. Receiver antenna patterns were separately provided by the SOC as well.

5.1.2 | ECMWF background

ECMWF is an independent intergovernmental organisation aiming to provide accurate medium-range global weather forecasts supported by most European countries (Owens and Hewson, 2018). Zonal and meridional (u, v) components of the 10-meter ocean surface winds provided by the ECMWF operational short-range forecast (background in the four-dimensional variational analysis system) from 1 June 2017 to 30 June 2017 were used for the background wind field. The ocean surface winds in ECMWF are hourly forecasts initiated from analysis times at 00UTC and 12UTC on a grid spacing of 18 km.

5.1.3 | Scatterometer winds

A scatterometer is an instrument to measure the roughness of a surface using radar backscatter. Spaceborne scatterometers have provided accurate wind field information for meteorology and climate over the past decades. Scatterometer (SCAT) 10-meter ocean surface winds from ASCAT aboard the Metop satellites (Metop-A and Metop-B) and OSCAT aboard the ScatSat-1 satellite (OSI SAF/EARS Winds Team, 2019; OSI SAF Winds Team, 2018) were used for validation in this study. The Metop satellites were developed by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and the ScatSat-1 satellite was developed by the Indian Space Research Organisation (ISRO). ASCAT has two sets of three antennas measuring ocean surface winds in two 550-km-wide swaths on both sides of the satellite ground track. It provides 10-meter wind products with 25-km and 12.5-km cell spacing. OSCAT uses a dish rotating antenna measuring ocean surface winds in an 1800-km-wide swath, providing 10-meter wind products in 50-km and 25-km cell spacing. The 25-km products from both instruments were used in this study to evaluate the result of DDM assimilation.

The 25-km zonal and meridional wind components measured by both instruments have been validated to have error standard deviation less than 1 m/s by a triple collocation method compared to buoy wind measurements and NWP models (Stoffelen et al., 2017; Verhoef et al., 2018). We never know the true wind speed in the real world. Buoy data are useful for validation, but have very limited spatial sampling compared to the satellite observations and can be affected by swell and wave reflections in coastal areas. SCAT (ASCAT especially, and OSCAT to a lesser degree) are very well-known and characterized systems for the last 30 years that provide accurate observation and excellent sampling of those parts of the ocean that CYGNSS measures. Given the high availability and accuracy of SCAT data (less than 1 m/s), SCAT data are used as “ground truth” in this study. It should be kept in mind that all validation statistics presented are statistics of differences, not errors.

5.1.4 | CYGNSS wind products

The CYGNSS Level 2 product, CYGNSS Climate Data Record (CDR) product and NOAA CYGNSS wind product are three different wind speed products retrieved from the CYGNSS Level 1 product using different algorithms. They will be compared to results of the DDM assimilation at the specular points.

- *CYGNSS Level 2 product v2.1:* Two observables, NBRCS and LES are first computed from a 3×5 window of the Level 1 DDM BRCS around the specular point. Two GMFs are developed to retrieve the 25-km surface wind speed at the specular point from these two observables. The two resulting wind speeds are then optimally combined to derive the minimum variance (MV) wind speed (Clarizia and Ruf, 2016).
- *CYGNSS Level 2 CDR product v1.0:* This is a new wind product released by CYGNSS SOC in 2020 (Ruf and Twigg, 2020). It is similar to the CYGNSS Level 2 product except that the observables NBRCS and LES are track-wise corrected using NASA’s MERRA-2 wind product to calibrate the GPS transmitter EIRP. Additional QC is also applied to the observables.
- *NOAA CYGNSS wind product:* Prepared by the National Oceanic and Atmospheric Administration (NOAA), this product is a 25-km surface wind speed at the CYGNSS specular points (Said et al., 2019). A new GMF was derived that expresses the CYGNSS NBRCS observable as a function of wind speed, incidence angle and significant wave height. The NBRCS observables are also track-wise corrected using the ECMWF model. 25-km gridding is implemented along each track to avoid overlapping observations. Additional rigorous QC is applied to the data.

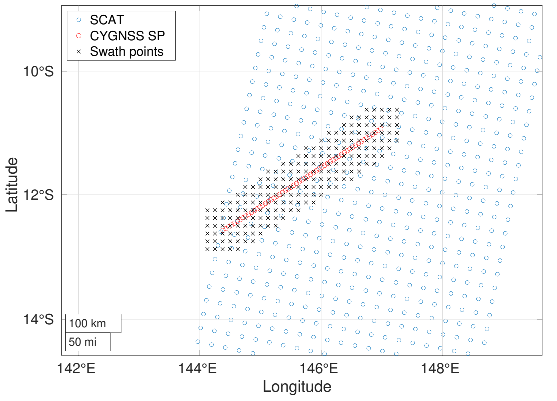


FIGURE 4 An example of the collocation for CYGNSS specular points, 0.125° grid points of the CYGNSS 80-km-wide swath, and 25-km WVCs of the SCAT swath in the period 00:00–00:20 UTC on 10 June 2017. The CYGNSS observations are measured by CYGNSS SV 5 from GPS PRN 14 signals. The SCAT measurement locations are for ASCAT-A.

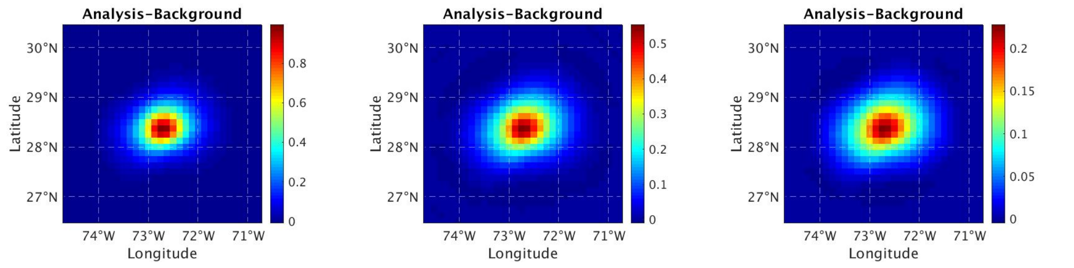
5.1.5 | Experimental design

The CYGNSS specular points were collocated with the SCAT wind vector cells (WVC) for all data from 1 June 2017 to 30 June 2017. Maximum differences of 40 minutes in time and 25 km in distance were used as criteria for collocation. If a CYGNSS specular point is collocated with several WVCs from different satellites (Metop-A, Metrop-B or ScatSat-1) then the average value of the wind speeds in all collocated WVCs was used. The DA experiment was done using a 20 minute cycle (0–20, 20–40, 40–60 minutes in each hour). In each 20-minute period, the analysis time is at the center of each cycle and the wind field is assumed to be constant. Hourly ECMWF surface winds were quadratically interpolated to the center time of each cycle from 0000 UTC on 1 June 2017 to 2400 UTC on 30 June 2017 and used as the background. The original ECMWF surface winds were also bilinearly interpolated to 0.125° grid spacing to match the working resolution of the DDM forward operator. In each cycle, all CYGNSS DDMs that were measured within the time period, passed the QC described in section 3.4, and were collocated with the SCAT WVCs were assimilated with the background using the VAM to produce the analysis on a 0.125° grid.

Two comparisons were made between the analysis winds and the reference SCAT winds.

- *Comparison at the specular points:* Wind vectors from the background and analysis wind field are linearly interpolated to the CYGNSS specular points and then compared to the collocated SCAT winds.
- *Comparison over a swath along the specular point track:* In order to evaluate the extent of the impact of assimilating DDMs, the wind vectors are compared over a much larger area than the one grid cell located at the specular point. CYGNSS data are first separated into different tracks corresponding to a specific pair of GPS transmitter and CYGNSS receiver. Along each track, background and analysis wind vectors on the 0.125° grid within a swath of a certain width are compared with collocated SCAT observations. Wind speeds at SCAT WVCs are linearly interpolated to the 0.125° grid of VAM wind field for the comparison. Figure 4 shows an example of the collocation for CYGNSS specular points, an 80-km-wide swath of the VAM gridded wind field, and 25-km SCAT WVCs.

The results of using three different DDM error covariance matrices are also compared: (a) a diagonal matrix using the scale method presented in section 4.1 (R-scale); (b) a diagonal matrix whose diagonal values are computed using the model presented



(a) (b) (c)

FIGURE 5 Wind speed increments (analysis−background) of assimilating a single DDM using different constraint weights. (λ_{div}, λ_{vor}, λ_{lap}) = (a) (50, 100, 25); (b) (200, 400, 100); (c) (800, 1600, 400). Higher weights increase the extent of the impact of new observations and reduce the increment's intensity. The DDM is observed by CYGNSS SV 4 and GPS PRN 2 at 1:18:43 UTC on 1 June 2017. The background and observation weights are 4 and 1/4 in all three cases.

in section 4.2 (R-model-diagonal); (c) a non-diagonal matrix computed using the model presented in section 4.2 (R-model).

5.2 | Tuning the weights

As introduced in section 3.1, there are a number of coefficients that can be used to weight the relative importance of the background winds vs. the new information. The constraint term and its weights describe background error correlations. In the study, the weight and standard deviation of the background wind components were fixed to be σ_b = 1 m/s and λ_b = 4. Only the ratio between these weights is important. The observation weight, λ_{ddm}, and constraint weights, λ_{lap}, λ_{div}, λ_{vor} were then determined by a series of sensitivity tests. In general, increasing the observation weight increases the intensity of the DA response, making the analysis closer to the observation, but does not change the shape of the response. Increasing the constraint weights increases the spatial scale of the response and decreases the intensity. λ_{lap} controls the smoothness of the response. λ_{div} and λ_{vor} control the shape of the response. Increasing the observation and constraint weights will also increase the number of iterations and computation cost in the minimization.

The constraint weights were first determined by a sensitivity test. Since they describe the background error correlations, the spatial scale of the response should be similar to the scale of the background effective resolution. It is important to note that the NWP grid spacing size and the model's effective resolution are different. In previous studies, the effective NWP model resolution was found to be 4-8 times larger than the grid spacing size (Skamarock, 2004; Abdalla et al., 2013). In our case, the effective model resolution of the ECMWF background is expected to be around 150 km (Stoffelen et al., 2018). Figure 5 shows the responses of assimilating a single DDM observation using three different sets of constraint weights. This example clearly show that increasing the constraint weight increases the area over which observations would have an effect. The DDM covariance is computed by the scale method and the observation weight λ_{ddm} is 1/4 in all three cases.

Considering that the footprint of a DDM observation is around 100 km and the model's effective resolution is around 150 km, the scale of the response should be about 250 km. By the sensitivity test, the constraint weights were chosen to be

$$(\lambda_{div}, \lambda_{vor}, \lambda_{lap}) = (200, 400, 100). \tag{15}$$

After determining the constraint weights, the observation weight λ_{ddm} is determined by another sensitivity test. As the CYGNSS specular point moves at about 6 km/s on the earth surface and the impact area of a DDM is about 250 km, the analysis

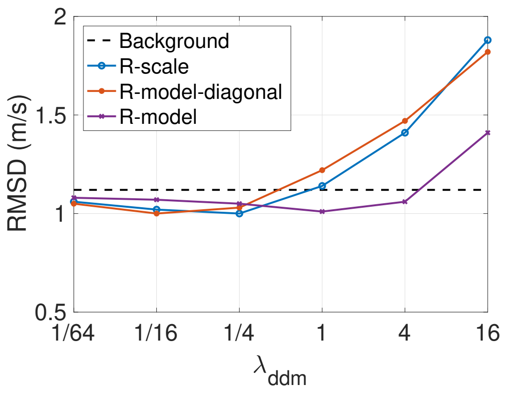


FIGURE 6 Wind speed RMSD at CYGNSS specular points versus observation weight in the VAM for different DDM error covariance matrices (R-scale, R-model-diagonal, R-model). The background wind speed RMSD at specular points is shown as the black dash line. Results are computed using data of one day on 10 June 2017.

wind speed at a point on the ocean surface can be impacted by 35–40 DDMs. Since the area impacted by a DDM through DA (~250 km) is larger than the area of its glistening zone (~100 km), the analysis wind speed at one point on the ocean surface can be affected by DDMs that, by themselves, are not sensitive to winds at that point. Due to this feature of overlapping measurements, in general λ_{ddm} should be much smaller than λ_b as a "deweighting" or equivalent "thining" of the observations. A total of ~25,000 DDMs from one day (10 June 2017) are processed by the VAM using a set of different observation weights, $\lambda_{ddm} = (1/64, 1/16, 1/4, 1, 4, 16)$, for each of the three DDM error covariance matrices. In each case, the Root Mean Square Difference (RMSD) between the VAM and SCAT wind speeds, evaluated at the specular point, was computed. Figure 6 shows the RMSD for all cases in the sensitivity test. The optimal λ_{ddm} for each DDM covariance matrix can be found by choosing the one with the minimal RMSD.

This result shows that the best observation weights λ_{ddm} for the three DDM error covariance matrices (R-scale, R-model-diagonal, R-model) are 1/4, 1/16 and 1, respectively. The optimal weight for the non-diagonal matrix (R-model) is larger than that for a diagonal matrix (R-model-diagonal) because adding error correlations and reconditioning the covariance matrix will reduce the weight of the observation (Tabeart et al., 2020). When λ_{ddm} decreases, the analysis wind field approaches that of the background, so it is expected that the RMSD in each case would likewise approach the background RMSD. When λ_{ddm} increases beyond its optimal value, the RMSD increases dramatically due to overfitting. Therefore, if the optimal λ_{ddm} cannot be precisely decided in an experiment, it is generally preferable to use a smaller one.

5.3 | Use of observation error covariance matrix

Results from our study using one day of data (Figure 6) show that, if the optimal λ_{ddm} is selected, there is little difference in the RMSD from using either of the three DDM error covariance matrices. To additionally validate the performance of using the three matrices, a total of ~170,000 DDMs from 5 days data (10 June 2017 to 14 June 2017) were processed using the three matrices combined with the corresponding optimal weights. The comparison was made both at the specular points and over swaths with two different widths. The results are listed in Table 2.

The conclusion of this study is that there is no significant difference in the accuracy of DA results, from comparisons at either the specular points or over a swath, using either of the three observation error covariance matrices. The slight differences

DDM error covariance matrix	Specular	80-km swath	120-km swath
R-scale	1.03	1.05	1.07
R-model-diagonal	1.04	1.07	1.10
R-model	1.06	1.08	1.10

TABLE 2 Wind speed RMSD compared to SCAT at CYGNSS specular points, over 80-km swath, and over 120-km swath. Comparison of results using different error covariance matrices. 5 days (10 June 2017 to 14 June 2017) of data. All units in m/s.

in the results of using the three matrices are possibly the results of testing only a single set of discrete values of λ_{ddm}. Similar performance for all three covariance matrics could be explained by the following reasons:

1) The VAM is heuristic. The observation error covariance matrix and the λ_{ddm} weight together determine the relative contribution of the observation in the analysis. Error in modeling the observation covariance matrix is compensated by choosing the optimal weight in the sensitivity study. This explains why the optimal λ_{ddm} for the three different covariance matrices are different whereas their final RMSD results are almost the same.

2) Each DDM bin observes an area defined by its delay and Doppler coordinate. This area on the ocean surface is usually 10–50 km across, which is much smaller than the ECMWF effective model resolution (150 km). Although the error correlations between each DDM bin may provide extra information, this small-scale information is smoothed out by the constraint terms in the VAM which are controlled by the effective model resolution of the background.

3) The reconditioning method used to decrease the large condition number of the non-diagonal error covariance matrix could add extra noise to the DA process, counteracting the benefit of additional information contained in the off-diagonal elements.

It is valuable to note in Figure 6, that the RMSD for R-model increases more slowly than the RMSD for R-scale when λ_{ddm} increases beyond its optimal value. This means that results from using R-model would be less sensitive to the choice of λ_{ddm}. One possible reason for this effect could be that the performance of DDM assimilation is mainly dependent on the error variances of DDM bins near the specular point and the weight λ_{ddm}. So if λ_{ddm} is selected to accurately correct the observation error covariance, the result is not sensitive to the method computing the covariance matrix. Whereas, if λ_{ddm} is not optimal, more accurately estimated covariances of DDM bins away from the specular point (from R-model) could mitigate the effect of sub-optimal weighting.

Our conclusion is that the three DDM error covariance matrices should give similar results when the optimal λ_{ddm} is selected. For the remainder of this study, the covariance matrix R-scale with its optimal weight will be applied, due to its simplicity. The non-diagonal covariance matrix R-model accounting for error correlations in the DDM could be valuable if DDMs are assimilated into DA systems at mesoscale or smaller spatial scales, e.g., a regional weather forecast model.

5.4 | Assimilation results

One month of CYGNSS Level 1 data from 1 June 2017 to 30 June 2017 (~663,000 DDMs, after applying the QC in section 3.4) was assimilated with the ECMWF background into the VAM to produce the analysis wind field (ECMWF-CY-DDM). The R-scale covariance matrix was used with weights determined in section 5.2 ($\lambda_{ddm} = 1/4$). Figure 7 shows an example of the wind field background, analysis, and increment for the 20-minute period from 6:40–7:00 UTC on 1 June 2017. This figure demonstrates that the impact of assimilating a track of DDMs extends over a 200–250 km wide swath, which is consistent with the 150 km background correlation length scale seen in Figure 5(b), given that the footprint of a DDM observation is around 100 km. Figure 8 shows the wind vectors on the contour maps of the background, analysis and increment for a closer look at a

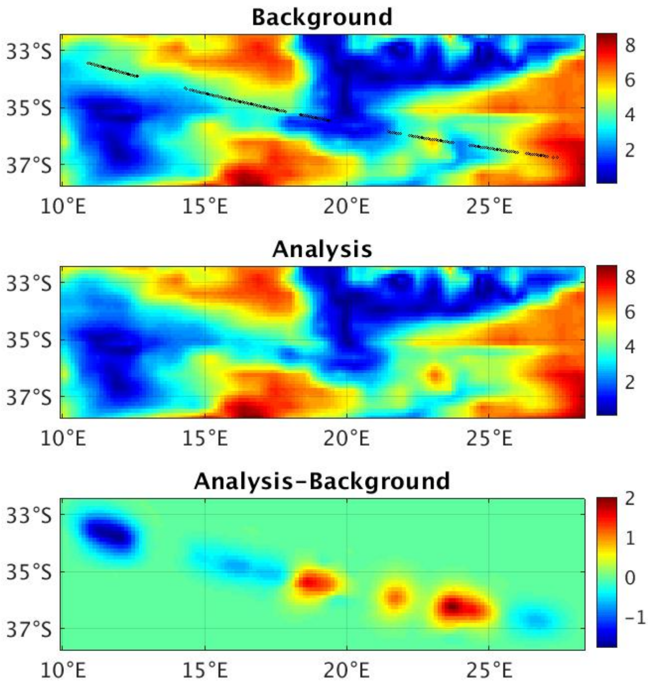


FIGURE 7 Wind field maps (m/s) of the ECMWF background, VAM analysis and increment (analysis−background) at 6:50 UTC on 1 June 2017. The CYGNSS specular point track is shown as the black circles on the background map.

region in the same time period. Since an isotropic slope PDF is assumed, with MSS a monotonic function of wind speed, the DDM observations will contain essentially no wind direction information. Analysis wind directions from the VAM are almost the same as those in the background, except for some negligibly slight changes due to the flow-dependent constraint terms.

A pair of density scatterplots showing a comparison of background and analysis wind speeds at CYGNSS specular points to SCAT winds is shown in Figure 9. The symmetric distribution of the samples with respect to the 1:1 line in both subfigures shows that both background and analysis are almost unbiased. The total wind speed RMSD at the specular points decreases from 1.17 to 1.07 m/s and the mean difference (bias) decreases from -0.14 to -0.08 m/s as a result of assimilating the DDMs. Wind speeds from both the background and analysis are smaller than SCAT wind speeds in general. The reduction of this bias, therefore, implies that the assimilation of CYGNSS DDMs increase the wind speeds from the ECMWF background on average. The wind speed RMSD and bias at the specular points for the background and analysis at different ranges of SCAT wind speed are shown in Table 3. Both the RMSD and bias of the background are significantly decreased by the assimilation of CYGNSS DDMs for wind speed less than 15 m/s, while the statistics almost remain the same for wind speed larger than 15 m/s. The decrease of the performance on high wind speed cases is mainly related to the decrease in sensitivity of the DDM measurements (surface slope PDF) to wind speed at high wind speeds, which is an intrinsic limitation of the physics of GNSS-R (Ruf et al., 2018). Also, the impact of wave age and fetch length at high wind speeds, which are not considered in the forward operator, could be another source of error. Nevertheless, the bias correction scheme prevents the assimilation of DDMs from introducing

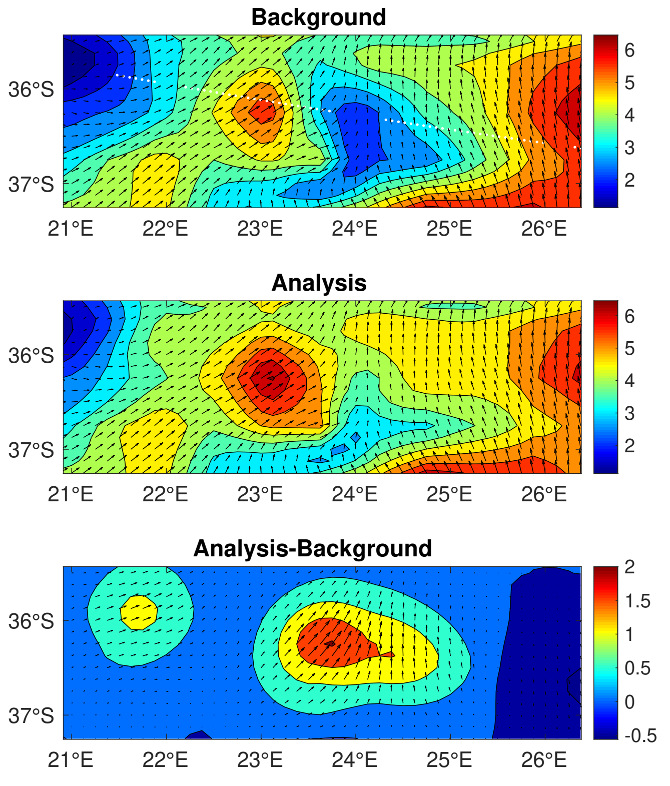


FIGURE 8 Wind contour maps and wind vector fields of the ECMWF background, VAM analysis and increment (analysis−background) at 6:50 UTC on 1 June 2017. Only a small part of the region plotted in Figure 7 is presented here. The CYGNSS specular point track is shown as the white circles in the background map. Wind vectors on the increment map are shown at a scale 5 times larger than that used on the Background and Analysis maps.

additional errors into the analysis relative to the background at high wind speeds. In the comparison of wind directions, data with collocated SCAT wind speeds less than 4 m/s are excluded because SCAT wind directions are less accurate at low wind speeds (Singh et al., 2011). The wind direction RMSDs of the background and analysis at specular points for the one month of data are 20.73° and 20.70°, the biases are 0.011° and 0.003°, respectively, compared to SCAT wind directions. Thus, the analysis retains the wind direction information from the background while the wind speeds are changed by the DDM assimilation.

Wind speed statistics are also computed over swaths of various widths (80, 120, and 150 km) along the CYGNSS specular point tracks. These results are listed in Table 4. Assimilation of CYGNSS DDMs is shown to improve the wind field accuracy, both at the specular point and over all swath widths. This improvement decreases as the swath width increases, which we interpret to be a consequence of the reduced sensitivity of the DDM away from the specular points. These results demonstrate the capability of CYGNSS DDM assimilation to improve the analyses of global NWP systems. The reduction of RMSD and bias of the ECMWF background is comparable to results from assimilating conventional scatterometer winds at global NWP centers

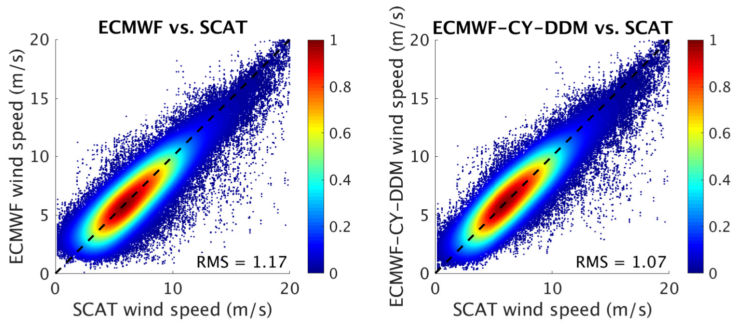


FIGURE 9 Density scatterplots for ECMWF background wind speeds (ECMWF), left panel and VAM analysis wind speeds (ECMWF-CY-DDM), right panel versus SCAT wind speeds at the CYGNSS specular points for one month of data (June 2017). The color scale indicates the density (normalized number) of the samples.

Wind speed range	< 5 m/s	5–10 m/s	10–15 m/s	> 15 m/s	Total
Nobs	178,498	393,826	80,918	9,425	663,909
ECMWF RMSD	1.14	1.08	1.39	2.45	1.17
ECMWF-CY-DDM RMSD	0.98	0.99	1.34	2.45	1.07
ECMWF Bias	0.33	-0.21	-0.66	-1.48	-0.14
ECMWF-CY-DDM Bias	0.22	-0.07	-0.62	-1.50	-0.08

TABLE 3 Wind speed RMSD (m/s) and mean difference (bias, m/s) of ECMWF background and VAM analysis (ECMWF-CY-DDM) compared to SCAT wind speeds over different ranges of SCAT wind speeds. The number of observations (Nobs) in each wind speed range is listed as well.

(Singh et al., 2011; Laloyaux et al., 2016).

	Specular	80-km swath	120-km swath	150-km swath
ECMWF	1.17	1.18	1.19	1.20
ECMWF-CY-DDM	1.07	1.10	1.11	1.13

TABLE 4 Wind speed RMSD (m/s) of the ECMWF background and VAM analysis (ECMWF-CY-DDM) at the CYGNSS specular points and over a swath with different widths (80-km, 120-km and 150-km) compared to SCAT wind speeds.

Another benefit of DDM assimilation is that the interpolated wind vectors from the VAM analyses can subsequently be used in other systems, provided it is recognized that the result is a combination of the DDM observation and ECMWF background information—essentially a wind retrieval from the DDM observable using the ECMWF background as a prior. To evaluate the performance of those wind speed retrievals, the interpolated wind speeds at the specular points from ECMWF-CY-DDM are compared to several other CYGNSS wind products: CYGNSS Level 2, CYGNSS Level 2 CDR, and NOAA-CYGNSS, which are described in section 5.1.4. All three products are 25-km wind speeds at the CYGNSS specular points retrieved from the CYGNSS Level 1 product. Both the CYGNSS Level 2 CDR product and the NOAA CYGNSS wind product apply a track-wise correction on the retrieved wind speeds using referenced NWP models. Wind speeds in the three products retrieved from the same CYGNSS Level 1 product for the one month of data in this study are compared to collocated SCAT winds. Note that all the three products apply some additional QCs and the NOAA CYGNSS wind product implements 25-km gridding along the track. Therefore, there are fewer collocated wind speeds from these three products (especially in the case of the NOAA product) than the number of CYGNSS Level 1 observations used in the DDM assimilation. RMSD and bias of all four products are compared in Table 5. The wind speeds from ECMWF-CY-DDM are shown to have smaller RMSD and bias than any of the other CYGNSS products. Another advantage of those retrievals is that a wind direction is assigned to each specular point, which might be beneficial to DA systems.

	Nobs	RMSD (m/s)	Bias (m/s)
CYGNSS-L2	661,230	1.50	-0.45
CYGNSS-CDR	520,432	1.57	-0.44
NOAA-CYGNSS	135,931	1.20	-0.33
ECMWF	663,909	1.17	-0.14
ECMWF-CY-DDM	663,909	1.07	-0.08

TABLE 5 Wind speed RMSD and bias at CYGNSS specular points compared to collocated SCAT wind speeds for CYGNSS Level 2 product (CYGNSS-L2), CYGNSS CDR product (CYGNSS-CDR), NOAA CYGNSS wind product (NOAA-CYGNSS), ECMWF background and VAM analysis (ECMWF-CY-DDM), for one month of data (June 2017).

6 | COMPUTATIONAL EFFICIENCY

Although DDM assimilation has been shown to improve global NWP analyses and produce wind speed estimates at a higher accuracy than conventional Level 2 products, it does come with a significant computational cost. The DDM forward operator is evaluated at each iteration of the optimization. The cost function in the VAM is minimized by a Quasi-Newton algorithm (Bonnans et al., 2006), using the convergence criteria listed in Table 6. About 30–50 function evaluations (including the forward operator) are generally required to reach the minimum.

Maximum infinity norm for the gradient of the cost function	10^{-6}
Maximum infinity norm for the change of the state between two iterations	10^{-6}
Maximum number of iterations	30
Maximum number of function evaluations	50

TABLE 6 The convergence criterion in VAM's minimization.

The experiment tasks in this study were run in parallel on two servers using Intel Xeon processors (one with 10 cores at 3.10-GHz, another with 12 cores at 2.53-GHz). Running the forward operator one time to compute a simulated DDM and a Jacobian matrix takes 0.4–0.5 CPU seconds on either server. Assimilating one DDM in the VAM takes about 20–30 CPU seconds. In total, it takes about 20 days elapsed time to process one month of data with ~663,000 DDMs using both servers running in parallel by *GNU parallel* (Tange, 2018).

The wind field grid size in this study is small (0.125°), which makes the computational cost of the DDM assimilation relatively high. The computational cost can be reduced by using a larger grid size of the wind field. Computing the forward operator by GPUs should also significantly improve the computational efficiency (Cervelló et al., 2020). Alternatively, a machine learning (ML) model could be trained to emulate results from the physically-based model.

7 | CONCLUSIONS

A variational analysis method (VAM) for assimilating CYGNSS Level 1 DDM power into global NWP analyses has been demonstrated, validated and assessed. A track-wise bias correction scheme was found to be necessary. The best results were obtained using a simple diagonal observation covariance matrix combined with optimal selection of the cost function weights. However, we did find a lower sensitivity to the observation weight when a non-diagonal covariance matrix was used. Our explanation for this effect is that the observation weight can counteract an inaccurate covariance matrix and the small-scale information in the error correlations is smoothed out by the constraint terms in the VAM. For some applications, such as regional forecast models, a full observation covariance matrix accounting for correlation between delay-Doppler bins may be beneficial.

We demonstrated our approach on one month (June 2017) of CYGNSS data collocated with SCAT observations, consisting of ~663,000 Level 1 DDMs. The VAM used ECMWF background winds in a cycle of 20 minutes to produce analysis winds on a 0.125° grid. Assimilation of a track of DDMs was shown to have an impact over a 200–250 km wide swath, corresponding approximately to the total extent of the DDM footprint (~100 km) plus the ECMWF effective model resolution (~150 km). These results also showed a reduction of the RMSD from 1.17 to 1.07 m/s and bias from -0.14 to -0.08 m/s as compared to reference scatterometer wind speeds. Wind directions were not changed significantly in the analyses, with an RMSD of 20.7° and bias of 0.0° compared to scatterometer data. DDM assimilation was also shown to improve the background wind field over a swath up to 150 km wide, reducing the wind speed RMSD from 1.20 to 1.13 m/s. These improvements in RMSD and bias are small, but are statistically significant considering the large sample of observations in our one-month study period. Because the ECMWF background we use has high accuracy and very small bias, there is not much room for improvement. Furthermore, we show that we can avoid overfitting with the proper setting of the weights in the VAM. Overall, these results indicate that assimilation of GNSS-R DDMs can have a positive impact on NWP analyses. The impact of GNSS-R DDM assimilation on regional weather forecast is the subject of a future study. We found that improvement was mostly limited to wind speeds below 15 m/s, however, probably as a result of the lower sensitivity of DDM observations to higher winds.

Wind vectors interpolated to the CYGNSS specular points from the VAM analysis can also be considered to be wind retrievals from the Level 1 DDM observables using the ECMWF background as a prior, essentially a Level 2 product with complicated error characteristics. These retrievals were compared to wind speeds from other CYGNSS wind products (Level 2, Level 2 CDR and NOAA). The RMSD and bias of VAM retrieved wind speeds were found to be lower than those of these other three products, as compared to scatterometer data.

Results presented here show substantial potential for assimilating DDMs directly into more complex DA systems. Future improvements and enhancements include streamlining the implementation of the forward model (possibly using ML) to improve computational efficiency, implementing the forward model within more complex DA systems, and accounting for the wave components driven by nonlocal winds in coupled DA systems.

REFERENCES

S Abdalla, Lars Isaksen, PAEM Janssen, and Nils Wedi. Effective spectral resolution of ECMWF atmospheric forecast models. *ECMWF Newsletter*, 137:19–22, 2013. doi: 10.21957/rue4o7ac. URL https://www.ecmwf.int/node/17358.

The authors declare no conflict of interest.

E Andersson, J Pailleux, J-N Thépaut, JR Eyre, AP McNally, GA Kelly, and P Courtier. Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 120(517):627–653, 1994.

Bachir Annane, Brian McNoldy, S Mark Leidner, Ross Hoffman, Robert Atlas, and Sharanya J Majumdar. A study of the HWRF analysis and forecast impact of realistically simulated CYGNSS observations assimilated as scalar wind speeds and as VAM wind vectors. *Monthly Weather Review*, 146(7):2221–2236, 2018. doi: 10.1175/MWR-D-17-0240.1.

Robert Atlas, Ross N Hoffman, Joseph Ardizzone, S Mark Leidner, Juan Carlos Jusem, Deborah K Smith, and Daniel Gombos. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bulletin of the American Meteorological Society*, 92(2):157–174, 2011. doi: 10.1175/2010BAMS2946.1.

Joseph-Frédéric Bonnans, Jean Charles Gilbert, Claude Lemaréchal, and Claudia A Sagastizábal. *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006. doi: 10.1007/978-3-540-35447-5.

F Bouttier and P Courtier. Data assimilation concepts and methods march 1999. *Meteorological training course lecture series. ECMWF*, 718:59, 2002. URL http://msi.ttu.ee/~elken/Assim_concepts.pdf.

Oriol Cervelló, Daniel Pascual, Raul Onrubia, and Adriano Camps. Advanced GNSS-R signals processing with GPUs. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:1158–1163, 2020. doi: 10.1109/JSTARS.2020.2975109.

Maria Paola Clarizia and Christopher S Ruf. Wind speed retrieval algorithm for the cyclone global navigation satellite system (CYGNSS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4419–4432, 2016. doi: 10.1109/TGRS.2016.2541343.

Maria Paola Clarizia and Christopher S. Ruf. Bayesian wind speed estimation conditioned on significant wave height for GNSS-R ocean observations. *Journal of Atmospheric and Oceanic Technology*, 34(6):1193–1202, 2017. doi: 10.1175/JTECH-D-16-0196.1.

Maria Paola Clarizia and Christopher S Ruf. Statistical derivation of wind speeds from CYGNSS data. *IEEE Transactions on Geoscience and Remote Sensing*, 2020. doi: 10.1109/TGRS.2019.2959715.

Maria Paola Clarizia, Christopher S Ruf, Philip Jales, and Christine Gommenginger. Spaceborne GNSS-R minimum variance wind speed estimator. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):6829–6843, nov 2014. doi: 10.1109/TGRS.2014.2303831.

Maria Paola Clarizia, Valery Zavorotny, and Christopher Ruf. CYGNSS algorithm theoretical basis document Level 2 wind speed retrieval, document 148-0138. Technical report, University of Michigan, Ann Arbor, MI, USA, 2018. URL https://clasp-research.engin.umich.edu/missions/cygnss/reference/148-0138%20ATBD%20L2%20Wind%20Speed%20Retrieval_Rev5_Aug2018_release.pdf.

MP Clarizia, CP Gommenginger, ST Gleason, MA Srokosz, C Galdi, and M Di Bisceglie. Analysis of GNSS-R delay-doppler maps from the UK-DMC satellite over the ocean. *Geophysical Research Letters*, 36(2), 2009. doi: 10.1029/2008GL036292.

M Cordoba, Sarah L Dance, GA Kelly, Nancy K Nichols, and Joanne A Waller. Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 143(702): 333–341, 2017. doi: 10.1002/qj.2925.

L Cucurull, JC Derber, and RJ Purser. A bending angle forward operator for global positioning system radio occultation measurements. *Journal of Geophysical Research: Atmospheres*, 118(1):14–28, 2013. doi: 10.1029/2012JD017782.

Zhiqiang Cui, Zhaoxia Pu, Vijay Tallapragada, Robert Atlas, and Christopher S Ruf. A preliminary impact study of CYGNSS ocean surface wind speeds on numerical simulations of hurricanes. *Geophysical Research Letters*, 46(5):2984–2992, 2019. doi: 10.1029/2019GL082236.

Gérald Desroziers, Loic Berre, Bernard Chapnik, and Paul Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613):3385–3396, 2005. doi: 10.1256/qj.05.108.

ECMWF. Copernicus climate change service (c3s) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, 2020. URL https://cds.climate.copernicus.eu/cdsapp#!/home.

Giuseppe Foti, Christine Gommenginger, Philip Jales, Martin Unwin, Andrew Shaw, Colette Robertson, and Josep Rosello. Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission. *Geophysical Research Letters*, 42(13): 5435–5441, 2015. doi: 10.1002/2015GL064204.

James L Garrison. A statistical model and simulator for ocean-reflected GNSS signals. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6007–6019, 2016. doi: 10.1109/TGRS.2016.2579504.

James L Garrison, Attila Komjathy, Valery U Zavorotny, and Stephen J Katzberg. Wind speed measurement using forward scattered GPS signals. *IEEE Transactions on Geoscience and Remote Sensing*, 40(1):50–65, 2002. doi: 10.1109/36.981349.

O Germain and G Ruffini. A revisit to the GNSS-R code range precision. *GNSS-R'06 Workshop*, 2006.

Scott Gleason, Stephen Hodgart, Yiping Sun, Christine Gommenginger, Stephen Mackin, Mounir Adjrad, and Martin Unwin. Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1229–1241, 2005. doi: 10.1109/TGRS.2005.845643.

Scott Gleason, Christine Gommenginger, and David Cromwell. Fading statistics and sensing accuracy of ocean scattered GNSS and altimetry signals. *Advances in space research*, 46(2):208–220, 2010. doi: 10.1016/j.asr.2010.03.023.

Scott Gleason, Christopher S Ruf, Maria Paola Clarizia, and Andrew J O'Brien. Calibration and unwrapping of the normalized scattering cross section for the cyclone global navigation satellite system. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):2495–2509, 2016. doi: 10.1109/TGRS.2015.2502245.

Scott Gleason, Christopher S Ruf, Andrew J O'Brien, and Darren S McKague. The CYGNSS Level 1 calibration algorithm and error analysis based on on-orbit measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12 (1):37–49, 2018. doi: 10.1109/JSTARS.2018.2832981.

Scott Gleason, Joel Johnson, Chris Ruf, and Charles Bussy-Virat. Characterizing background signals and noise in spaceborne GNSS reflection ocean observations. *IEEE Geoscience and Remote Sensing Letters*, 2019. doi: 10.1109/LGRS.2019.2926695.

RN Hoffman, SM Leidner, JM Henderson, R Atlas, JV Ardizzone, and SC Bloom. A two-dimensional variational analysis method for NSCAT ambiguity removal: Methodology, sensitivity, and tuning. *Journal of Atmospheric and Oceanic Technology*, 20(5):585–605, 2003. doi: 10.1175/1520-0426(2003)20\%3C585:ATDVAM\%3E2.0.CO;2.

Ross N Hoffman. SASS wind ambiguity removal by direct minimization. *Monthly Weather Review*, 110(5):434–445, 1982. doi: 10.1175/1520-0493(1982)110\%3C0434:SWARBD\%3E2.0.CO;2.

Ross N Hoffman. SASS wind ambiguity removal by direct minimization. part II: Use of smoothness and dynamical constraints. *Monthly Weather Review*, 112(9):1829–1852, 1984. doi: 10.1175/1520-0493(1984)112\%3C1829:SWARBD\%3E2.0.CO;2.

Ross N Hoffman. The effect of thinning and superobservations in a simple one-dimensional data analysis with mischaracterized error. *Monthly Weather Review*, 146(4):1181–1195, 2018.

Feixiong Huang, James L. Garrison, Nereida Rodriguez-Alvarez, Andrew J. O'Brien, Kaitie M. Schoenfeldt, Soon Chye Ho, and Han Zhang. Sequential processing of GNSS-R delay-doppler maps to estimate the ocean surface wind field. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):10202–10217, Dec 2019a. ISSN 0196-2892. doi: 10.1109/TGRS.2019.2931847.

Feixiong Huang, Giuseppe Grieco, and Ad Stoffelen. GNSS-R processing and NWP assimilation. *Eumetsat OSI SAF*, 2019b. URL http://www.osi-saf.org/sites/default/files/dynamic/page_with_files/file/osisaf_vs19_01_Huang_Report_v1.0.pdf.

Feixiong Huang, James L Garrison, S Mark Leidner, Bachir Annane, Ross N Hoffman, Giuseppe Grieco, and Ad Stoffelen. A forward model for data assimilation of gnss ocean reflectometry delay-doppler maps. *IEEE Transactions on Geoscience and Remote Sensing*, 2020a. doi: 10.1109/TGRS.2020.3002801.

Feixiong Huang, Andrew O'Brien, Nereida Rodriguez-Alvarez, and James Garrison. GNSS-R DDM forward model and variational analysis method for data assimilation. https://www.codeocean.com/, June 2020b.

T Janjić, N Bormann, M Bocquet, JA Carton, SE Cohn, SL Dance, SN Losa, NK Nichols, R Potthast, JAa Waller, et al. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278, 2018.

Stephen J Katzberg, Omar Torres, and George Ganoe. Calibration of reflected GPS for tropical storm wind speed retrievals. *Geophysical Research Letters*, 33(18), 2006. doi: 10.1029/2006GL026825.

Patrick Laloyaux, Jean-Noël Thépaut, and Dick Dee. Impact of scatterometer surface wind data in the ECMWF coupled assimilation system. *Monthly Weather Review*, 144(3):1203–1217, 2016. doi: 10.1175/MWR-D-15-0084.1.

S Mark Leidner, Bachir Annane, Brian McNoldy, Ross Hoffman, and Robert Atlas. Variational analysis of simulated ocean surface winds from the cyclone global navigation satellite system (CYGNSS) and evaluation using a regional OSSE. *Journal of Atmospheric and Oceanic Technology*, 35(8):1571–1584, 2018. doi: 10.1175/JTECH-D-17-0136.1.

S Mark Leidner, SJ Majumdar, J Hegarty, and Brian D McNoldy. CYGNSS data impact on global analyses of ocean surface winds. In *100th American Meteorological Society Annual Meeting*. AMS, 2020.

Weiqiang Li, Antonio Rius, Fran Fabra, Estel Cardellach, Serni Ribo, and Manuel Martín-Neira. Revisiting the GNSS-R waveform statistics and its impact on altimetric retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5):2854–2871, 2018. doi: 10.1109/TGRS.2017.2785343.

Xuanli Li, John R Mecikalski, and Timothy J Lang. A study on assimilation of CYGNSS wind speed data for tropical convection during 2018 January MJO. *Remote Sensing*, 12(8):1243, 2020. doi: 10.1029/2019GL082236.

Hans Lievens, Brecht Martens, NEC Verhoest, S Hahn, RH Reichle, and D Gonzalez Miralles. Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates. *Remote Sensing of Environment*, 189:194–210, 2017. doi: 10.1016/j.rse.2016.11.022.

Francisco Martín, Salvatore D'Addio, Adriano Camps, and Manuel Martín-Neira. Modeling and analysis of GNSS-R waveforms sample-to-sample correlation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(5):1545–1559, 2014. doi: 10.1109/JSTARS.2014.2308982.

Manuel Martín-Neira, Salvatore D'Addio, Christopher Buck, Nicolas Floury, and Roberto Prieto-Cerdeira. The PARIS ocean altimeter in-orbit demonstrator. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2209–2237, 2011. doi: 10.1109/TGRS.2010.2092431.

OSI SAF Winds Team. ScatSat-1 wind product user manual, version 1.3, 2018. URL http://projects.knmi.nl/scatterometer/publications/pdf/osisaf_cdop2_ss3_pum_scatsat1_winds.pdf.

OSI SAF/EARS Winds Team. ASCAT wind product user manual, version 1.16, 2019. URL https://projects.knmi.nl/scatterometer/publications/pdf/ASCAT_Product_Manual.pdf.

RG Owens and TD Hewson. ECMWF forecast user guide. *Reading: ECMWF*, 2018. doi: 10.21957/m1cs7h.

Jennifer Reynolds, Maria Paola Clarizia, and Emanuele Santi. Wind speed estimation from CYGNSS using artificial neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:708–716, 2020. doi: 10.1109/JSTARS.2020.2968156.

Nereida Rodriguez-Alvarez and James L. Garrison. Generalized linear observables for ocean wind retrieval from calibrated GNSS-R delay–doppler maps. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2):1142–1155, Feb 2016. doi: 10.1109/TGRS.2015.2475317.

C Ruf, M Unwin, J Dickinson, R Rose, D Rose, M Vincent, and A Lyons. CYGNSS: Enabling the future of hurricane prediction [remote sensing satellites]. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):52–67, 2013. doi: 10.1109/MGRS.2013.2260911.

C Ruf, P Chang, MP Clarizia, S Gleason, Z Jelenak, J Murray, M Morris, S Musko, D Posselt, D Provost, et al. CYGNSS handbook. *Ann Arbor, MI, Michigan Pub*, 154, 2016. URL https://clasp-research.engin.umich.edu/missions/cygnss/reference/cygnss-mission/CYGNSS_Handbook_April2016.pdf.

Chris Ruf and Dorina Twigg. Algorithm theoretical basis document Level 1 & 2 trackwise corrected climate data record. Technical report, University of Michigan, Ann Arbor, MI, USA, 2020. URL https://podaac-tools.jpl.nasa.gov/drive/files/allData/cygnss/L2/docs/148-0389-1_ATBD_Trackwise_Corrected_CDR.pdf.

Christopher S Ruf and Rajeswari Balasubramaniam. Development of the CYGNSS geophysical model function for wind speed. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(1):66–77, 2018. doi: 10.1109/JSTARS.2018.2833075.

Christopher S Ruf, Scott Gleason, and Darren S McKague. Assessment of CYGNSS wind speed retrieval uncertainty. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(1):87–97, 2018. doi: 10.1109/JSTARS.2018.2825948.

Faozi Said, Zorana Jelenak, Jeonghwang Park, Seubson Soisuvarn, and Paul S Chang. A 'track-wise'wind retrieval algorithm for the CYGNSS mission. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 8711–8714. IEEE, 2019. doi: 10.1109/IGARSS.2019.8898099.

Randhir Singh, Prashant Kumar, and Pradip Kumar Pal. Assimilation of Oceansat-2-scatterometer-derived surface winds in the weather research and forecasting model. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1015–1021, 2011. doi: 10.1109/TGRS.2011.2164410.

William C Skamarock. Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132(12):3019–3032, 2004. doi: 10.1175/MWR2830.1.

F Soulat, M. Caparrini, O. Germain, P. Lopez-Dekker, M. Taani, and Giulio Ruffini. Sea state monitoring using coastal GNSS-R. *Geophysical Research Letters*, 31(21):L21303, nov 2004. doi: 10.1029/2004GL020680.

Ad Stoffelen and Jur Vogelzang. Wind bias correction guide. *EUMETSAT, Darmstadt, Germany, Tech. Rep. NWPSAF-KN-UD-007, v1.3*, 2018. URL http://projects.knmi.nl/publications/fulltexts/wind_bias_correction_guide_v1.3_def.pdf.

Ad Stoffelen, Jeroen Adriaan Verspeek, Jur Vogelzang, and Anton Verhoef. The CMOD7 geophysical model function for ASCAT and ERS wind retrievals. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2123–2134, 2017. doi: 10.1109/JSTARS.2017.2681806.

Ad Stoffelen, Jur Vogelzang, and Gert-Jan Marseille. High resolution data assimilation guide. *EUMETSAT, Darmstadt, Germany, Tech. Rep. NWPSAF-KNUD-008, v1.2*, 2018. URL https://nwp-saf.eumetsat.int/site/download/documentation/scatterometer/reports/High_Resolution_Data_Assimilation_Guide_1.2.pdf.

Jemima M Tabeart, Sarah L Dance, Amos S Lawless, Nancy K Nichols, and Joanne A Waller. Improving the condition number of estimated covariance matrices. *Tellus A: Dynamic Meteorology and Oceanography*, 72(1):1–19, 2020. doi: 10.1080/16000870.2019.1696646.

Ole Tange. *GNU Parallel 2018*. Ole Tange, March 2018. ISBN 9781387509881. doi: 10.5281/zenodo.1146014. URL https://doi.org/10.5281/zenodo.1146014.

Anton Verhoef, Jur Vogelzang, and Ad Stoffelen. *ScatSat-1 wind validation report*. OSI SAF, 2018. URL http://projects.knmi.nl/scatterometer/publications/pdf/osisaf_cdop3_ss3_valrep_scatsat1_winds.pdf.

Joanne A Waller, Susan P Ballard, Sarah L Dance, Graeme Kelly, Nancy K Nichols, and David Simonin. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote sensing*, 8(7):581, 2016. doi: 10.3390/rs8070581.

Tianlin Wang, Christopher S Ruf, Bruce Block, Darren S. McKague, and Scott Gleason. Design and performance of a GPS constellation power monitor system for improved CYGNSS L1B calibration. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(1):26–36, jan 2019. doi: 10.1109/JSTARS.2018.2867773.

Yue Ying and Fuqing Zhang. Potentials in improving predictability of multiscale tropical weather systems evaluated through ensemble assimilation of simulated satellite-based observations. *Journal of the Atmospheric Sciences*, 75(5):1675–1698, 2018. doi: 10.1175/JAS-D-17-0245.1.

Huaizu You, James L Garrison, Gregory Heckler, and Dino Smajlovic. The autocorrelation of waveforms generated from ocean-scattered GPS signals. *IEEE Geoscience and Remote Sensing Letters*, 3(1):78–82, 2006. doi: 10.1109/LGRS.2005.856704.

Valery U Zavorotny and Alexander G Voronovich. Scattering of GPS signals from the ocean with wind remote sensing application. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2):951–964, 2000. doi: 10.1109/36.841977.

Valery U Zavorotny, Scott Gleason, Estel Cardellach, and Adriano Camps. Tutorial on remote sensing using GNSS bistatic radar of opportunity. *IEEE Geoscience and Remote Sensing Magazine*, 2(4):8–45, 2014. doi: 10.1109/MGRS.2014.2374220.

Shixuan Zhang, Zhaoxia Pu, Derek J Posselt, and Robert Atlas. Impact of CYGNSS ocean surface wind speeds on numerical simulations of a hurricane in observing system simulation experiments. *Journal of Atmospheric and Oceanic Technology*, 34(2):375–383, 2017. doi: 10.1175/JTECH-D-16-0144.1.

Cinzia Zuffada, Tanos Elfouhaily, and Stephen Lowe. Sensitivity analysis of wind vector measurements from ocean reflected GPS signals. *Remote Sensing of Environment*, 88(3):341–350, 2003. doi: 10.1016/S0034-4257(03)00175-5.

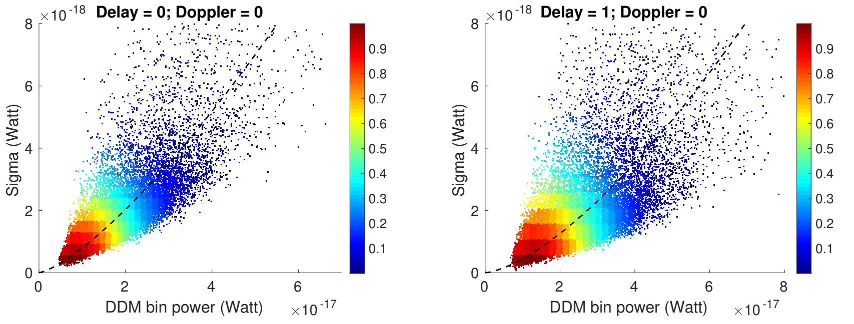
A | APPENDIX

Additional details concerning the development of an empirical model for the DDM error covariance, described in section 4.2, are presented here.

Recall that standard deviation of the speckle component of the DDM at each delay-Doppler coordinate is assumed to follow a power-law dependence on the DDM magnitude in equation (12). This assumption is justified from knowledge that speckle (before averaging) is a multiplicative noise having an exponential distribution in which the standard deviation is proportional to signal power (Gleason et al., 2010). The actual DDM, however, is formed from the incoherent average of 1000 cross-correlations every second. Correlation time of the DDM observation from a spaceborne receiver is typically a few milliseconds (Li et al., 2018), resulting in an incoherent average containing fewer than 1000 equivalent independent samples. The correlation time depends on the geometry, delay, and Doppler of the corresponding DDM bin (Zuffada et al., 2003). The noise distribution will therefore be a function of the delay and Doppler coordinates. Generally, correlation time decreases with longer delays (You et al., 2006). A nonlinear model for the standard deviation of the speckle noise as a function of signal expectation and correlation time was given in Clarizia et al. (2018). Our empirical model is an attempt to account for this variation through assigning unique coefficients in (12) at each delay-Doppler coordinate.

Similarly, the correlation between DDM observations at different delay-Doppler pairs, defined by a correlation coefficient (13), is modelled as a polynomial function of the inverse wind speed (14). This dependence on wind speed was found to fit the data well and could be explained by the structure of models for the bin-bin ("fast time") covariance (e.g., equation (41) in Garrison (2016) or equation (29) in Martín et al. (2014)).

Our basic approach is to estimate arrays of coefficients, p, q, a, b, and c, which best fit the functions (12) and (14) to a month of CYGNSS Level 1 v2.1 DDM data (June 2017), encompassing the expected range of geometry and surface conditions. 10-meter ocean surface wind speeds provided by the ECMWF ERA5 reanalysis (ECMWF, 2020) in a 0.25° latitude-longitude grid were used as the reference. The ECMWF ERA5 reanalysis winds were interpolated linearly in time and space to the specular point of each DDM. Given the approximate velocity of a CYGNSS specular point on the earth surface of around 6 km/s (Ruf et al., 2016), and approximating the DDM covariance matrix as constant over scales equal to the effective ECMWF model resolution (150 km, (Stoffelen et al., 2018)), batches of 25 sequential DDMs were used to compute the sample covariance. The satellite geometries, transmitter power, and antenna patterns were also assumed to remain constant within the corresponding 25



(a) (b)

FIGURE 10 Speckle variance, $\sigma_{i,s}$ vs. DDM power magnitude at two different delay-Doppler coordinates. The color scale indicates the density of the points. In the titles of the two figures, "Delay" and "Doppler" are relative to that of the specular point in units of bins (0.25 chip, 500 Hz). Black dashed lines on both figures show the best fit of equation (12).

second time period. Such small batches of data will result in a large uncertainty in the individual covariance estimates. However, combining a large number of these batches together to estimate a small number of parameters defining the empirical model in (12) and (14) is expected to average out the uncertainty in the individual sample covariances.

The following quality control (QC) tests were applied to the data used to compute the covariance matrices:

- The "quality_flags" variable in the CYGNSS Level 1 data for each DDM is zero.
- The signal-to-noise ratio (SNR) for each DDM is larger than 3 dB.
- The minimum of wind speeds for each batch is larger than 3 m/s. This is to avoid the impact of the swell and coherent scattering (Huang et al., 2020a).
- The range of wind speeds for each batch is less than 10% of the average wind speed for the batch. This is to confirm that the wind speed almost remains the same during the time of a batch, in case there is a high variational wind condition.

In contrast to the QC approach defined in section 3.4 for DA, we did not set requirements on the relative power difference or correlation coefficient. A total of 119193 DDM batches in June 2017 passed these QC tests.

The contribution of thermal noise was assumed constant in time and independent of the delay-Doppler coordinate. An average of the sample variances for the first two rows (assumed not to contain any reflected signal) was used to compute a value of $\hat{\sigma}_n^2 = 9.576 \times 10^{-38}$ W^2.

The sample variance for the i-th delay-Doppler coordinate of the DDM, $\hat{\sigma}_i^2$ was computed for each batch as well. The thermal noise contribution was then subtracted to produce an estimate of the speckle contribution to the standard deviation,

$$\hat{\sigma}_{i,s} = \sqrt{\hat{\sigma}_i^2 - \hat{\sigma}_n^2}. \tag{16}$$

Figure 10 shows scatterplots for the speckle noise contribution, $\hat{\sigma}_{i,s}$ vs. the DDM magnitude from all batches for two different delay-Doppler coordinates. Although there is large scattering on both figures due to the small sample size in each batch, a clear trend with DDM magnitude is visible. The best fit of equation (12), through estimating p and q, is shown as the dashed black line on these figures. This model fitting was applied to all DDMs over discrete delay range [-1,10] and Doppler range [-3,3], in bins defined relative to the specular point delay and Doppler. This provides 12×7 matrices, $\mathbf{P}$ and $\mathbf{Q}$, containing values of p and

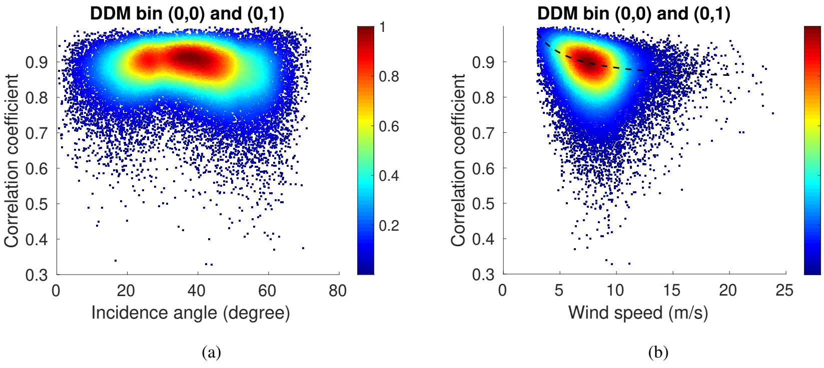


FIGURE 11 Correlation coefficient between the DDM at the specular point (0,0) and that sampled at (1,0), versus incidence angle (a) and wind speed (b). The color scale indicates the density of the points. The black dashed line shows the best fit of equation (14).

q for each bin of the DDM in delay-Doppler space.

A similar approach was applied to determine numerical values in the correlation coefficient model (14). The correlation coefficient at two different delay-Doppler coordinates, $(\tau, f)_i$ and $(\tau, f)_j$, was computed as

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j} \tag{17}$$

where $\hat{\sigma}_{ij}$ is the sample covariance of the DDM at $(\tau, f)_i$ and $(\tau, f)_j$, computed from the same 25-member batch as $\hat{\sigma}_i$ and $\hat{\sigma}_j$. Figure 11 shows scatterplots of the correlation coefficient between at the (0,0) and (1,0) delay-Doppler coordinate vs incidence angle (a) and wind speed (b). These figures show little dependence on the incidence angle, but an evident dependence on the wind speed. Scatterplots generated at different delay-Doppler coordinates all show similar patterns, supporting our assumption that the correlation coefficient does not strongly depend on SNR, DDM power magnitude, transmitter EIRP, or receiver antenna gain (not shown), but does exhibit some dependence on wind speed. These sensitivity studies were used to determine the form of (14). The black dashed line on Figure 11(b) shows the fitting of this function to the data. This approach was applied to every pair of DDM observables over the delay range [-1,10] and Doppler range [-3,3]. Fitting the model produces (84×84) symmetric matrices, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ containing the three coefficients defining the model in (14). Diagonal values of A are all ones and diagonal values of $\mathbf{B}$ and $\mathbf{C}$ are all zeros.

Numerical values for matrices, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are provided as supplemental material.