remote sensing

MDPI

Article

An Instrument Error Correlation Model for Global Navigation Satellite System Reflectometry

C. E. Powell [1,*], Christopher S. Ruf [1], Darren S. McKague [1], Tianlin Wang [2] and Anthony Russel [1]

[1] Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA; russelan@umich.edu (A.R.)
[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91011, USA
* Correspondence: cepowell@umich.edu

Abstract: All sensing systems have some inherent error. Often, these errors are systematic, and observations taken within a similar region of space and time can have correlated error structure. However, the data from these systems are frequently assumed to have completely independent and uncorrelated error. This work introduces a correlated error model for GNSS reflectometry (GNSS-R) using observations from NASA's Cyclone Global Navigation Satellite System (CYGNSS). We validate our model against near-simultaneous observations between two CYGNSS satellites and double-difference our results with modeled observables to extract the correlated error structure due to the observing system itself. Our results are useful to catalog for future GNSS-R missions and can be applied to construct an error covariance matrix for weather data assimilation.

Keywords: GNSS; reflectometry; observation error; correlated error

check for updates

Citation: Powell, C.E.; Ruf, C.S.; McKague, D.S.; Wang, T.; Russel, A. An Instrument Error Correlation Model for Global Navigation Satellite System Reflectometry. *Remote Sens.* **2024**, *16*, 742. https://doi.org/10.3390/rs16050742

Academic Editor: Chung-yen Kuo

Received: 5 January 2024
Revised: 13 February 2024
Accepted: 16 February 2024
Published: 20 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Launched in 2016, the Cyclone Global Navigation Satellite System (CYGNSS) mission provides a suite of land and ocean surface sensing products through a technique known as Global Navigation Satellite System reflectometry (GNSS-R). With GNSS-R, the eight CYGNSS satellites act as receivers in a bistatic radar system, where GPS satellites are the transmitters, and the Earth's surface is the target of interest.

The CYGNSS mission aims to improve tropical cyclone prediction by measuring ocean surface wind speeds over the tropics [1]. Other satellite observations are impaired by the difficulty in collecting wind speeds underneath thick cloud layers in tropical cyclones, and in situ measurements are often challenging and dangerous to collect due to extreme conditions within tropical cyclones. By exploiting GPS broadcasts in the L-band, where there is limited attenuation from thick rain and cloud layers, CYGNSS can collect observations through mature tropical cyclones.

With eight satellites in equatorial low-Earth orbit, CYGNSS can collect global observations with a median revisit time of 3.8 h. However, the sampling characteristics of CYGNSS differ from many other space-based weather observations. Instead of imaging, CYGNSS data are comprised of tracks of specular points that streak across the Earth's surface. Images are reconstructed from composites of multiple satellites. This unique sampling mechanism poses challenges for assimilation into numerical weather prediction (NWP) models.

Data assimilation (DA) is the component of an NWP system that incorporates observations with evolving forecast models. A common DA technique poses the problem as a variational equation that minimizes a cost function [2]:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \mathbf{J}(\mathbf{x}) = (\mathbf{x}-\mathbf{x_b})^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x_b}) + (\mathbf{y}-H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}]), \quad (1)$$

where $\mathbf{J}(\mathbf{x})$ represents the cost function, $\mathbf{x}$ is the variational argument, $\mathbf{x_b}$ is the background model state of dimension n, $\mathbf{B}$ is the n-by-n background error covariance matrix, $\mathbf{y}$ is the

observation vector of dimension p, H is the forward model operator for transforming from model space to observation space, and **R** is the p-by-p observation error covariance matrix.

In practice, a great deal of focus is placed on estimating **B** [3,4] because incorrectly specified **B** can have significant impacts on the performance of the model. Famously, introducing perfect observations to a perfect model can still lead to forecast degradations if **B** is poorly constructed [5].

On the other hand, **R** is comparatively overlooked. NWP data assimilation (DA) schemes often assume that each observation error is completely uncorrelated [6–8] and, as such, treat **R** operationally as a diagonal matrix. This is performed for three main reasons: to simplify the assumptions needed for minimizing the cost function in Equation (1) and reduce the computational resources required to solve the forecast problem; to reduce the data handling and storage requirements for observation data at low latencies; and because frequently the information required to specify correlated observation errors was unavailable or difficult to ascertain. Regardless of the rationale, the under-specification of correlated observation error in **R** leads to the suboptimal use of the information content in observations.

There are some major recent advancements on this front. Discussions of correlated instrument errors are generally limited to explorations for interchannel correlated errors in infrared [9,10] and microwave [11] sounders, with some physical models of correlation proposed [12,13]. Spatial observation error correlation is discussed in [11,14] but is generally restricted to estimates of total observation error correlation as estimated from a weather model [15]. More recently, instrument error inventories for microwave sounders have been established [13].

Nevertheless, the assumption that observation errors are independent and uncorrelated is not valid for any observatory. In practice, observation error correlations are minimized by thinning the dataset or 'super-obbed' samples, which improves overall model skill by mitigating the effects of correlated error, which may be interpreted by the model as a real signal [16–18].

The premise of not fully exploiting observation data is unappealing, considering that the process of acquiring these data usually requires the considerable capital expense of building and flying remote sensing satellites. The highest utility of dense observations in space and time occurs when the application (i.e., the model) resolution matches the observation resolution. For many imagers and sounders, there may be a reasonable case that thinning is an appropriate way to scale dense observation data to match model gridding.

For CYGNSS, however, thinning the data is an especially unpalatable solution, as the peak utility of CYGNSS data is during the relatively uncommon occurrence that a specular point passes through the eyewall of a tropical cyclone. Thinning may miss this observation or misrepresent the structure of the storm in the model. By a similar token, 'super-obbing' is not a practicable remedy because tracks are one dimensional and collocated observations are potentially hours apart. Blending data from a wide temporal window risks misrepresenting the dynamics of the tropical cyclone and negates the fast-revisit utility of the observatory.

Estimating the correlated error structure for observations can improve model skills in NWP [6,14,19–21]. A number of methods have been demonstrated to estimate the observation error correlation matrix [22,23], most notably Desroziers' diagnostic [24].

Our work does not estimate the full observation error correlation matrix that is typical of these prior works. Instead, we provide a first-principle, bottoms-up, tunable engineering model for how the CYGNSS instrumentation itself can cause observations to contain correlated errors. This correlation model is validated and tuned against empirical observation data during a period when two CYGNSS assets were in a specific orbital geometry where observations nearly coincided in space and time. We further discuss the challenges and opportunities that result from this correlated error model, as well as highlight the potential for applicability to future assimilation investigations.

2. Materials and Methods

2.1. The CYGNSS Observatory

The CYGNSS main observable is normalized bistatic radar cross section (NBRCS), σ^o. This is derived from the radar-range equation [25]:

$$\sigma^o = \frac{P^g (4\pi)^3 L^{atm}}{P^T G^T \lambda^2 G^R R^{tot} \bar{A}}, \quad (2)$$

where P^g is the calibrated power received by the nadir science antenna; L^{atm} are the losses due to attenuation in the atmosphere; P^T is the power of the transmitter, which, in this case, is a GPS satellite; G^T is the antenna gain of the GPS satellite in the direction of the Earth specular point; λ is the wavelength of the GPS signal, which, for CYGNSS, is the GPS L1 signal at 1575 MHz; G^R is the gain of the CYGNSS nadir science antenna in the direction of the Earth specular point; R^{tot} is the total range loss of the transmission via the Earth specular point; and $\bar{A}$ is the effective scattering area of the specular point.

NBRCS is directly related to the mean square slope (MSS) of ocean wave spectra via [25,26]:

$$\sigma^o(\theta) = \frac{|\Re(\theta)|^2}{MSS}, \quad (3)$$

where θ is the angle of incidence of the scattering geometry and $\Re(\theta)$ is the Fresnel electric field reflection coefficient for the ocean surface at the specular point. CYGNSS retrieves wind speed from σ^o via an empirical Geophysical Model Function (GMF) [27] that is processed operationally via a stored look-up table (LUT).

In practice, the un-normalized bistatic radar cross section σ is calculated for every delay and Doppler, the native coordinates of the delay–Doppler Mapping Instrument (DDMI). σ^o is calculated from the average of σ over a 3-by-5 bin window centered at the specular point and is divided by the effective area $\bar{A}$ corresponding to the surface area bounded by the 3-by-5 window [m^2/m^2].

After the launch of the CYGNSS mission, it became clear that uncertainty in the GPS effective isotropic radiated power (EIRP) would significantly impact calibration quality [28–30]. The EIRP is simply the product $P^T G^T$ from Equation (2). Errors are present in both terms: not only is there general uncertainty with the actual (versus published) transmit power of GPS satellites, but newer generations of the constellation operate with a flexible power mode and dynamically change transmit power levels across their respective orbits [31]. Further, only a subset of the GPS antenna patterns was published [32], and they were not sufficiently detailed to constrain the error in EIRP for CYGNSS calibration.

To remedy this, Wang et al. developed a novel calibration technique that uses the onboard zenith CYGNSS antenna and empirically derived relationships from measured GPS antenna patterns to estimate GPS EIRP in the direction of the Earth's specular point at the time of observation [33]. This new technique has been adopted in the CYGNSS Level 1 calibration algorithm since version 3.0 and modifies the radar equation to become the following [34]:

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

where all the terms are the same as in Equation (2), with the additions of L^Z, the transmission range loss from the GPS source to the CYGNSS zenith antenna (written as a loss in the numerator to distinguish from the R^{tot} range losses); the CYGNSS zenith antenna gain G^Z in the direction of the GPS satellite; the power received from the CYGNSS zenith receiver P^Z; and the specular–zenith ratio ζ^E for the GPS EIRP as described in [33].

Each of the terms in Equation (4) can potentially be a source of correlated error, but for practical reasons, this study only evaluates those sources with the largest error magnitude and, therefore, the largest potential utility for future data assimilation users. We omitted consideration of the smallest error sources that, when combined as a root sum of squares,

contribute to less than 1% of the total error magnitude. The five largest sources of error are explored, as shown in Table 1, representing a 1-sigma error magnitude for a reference wind speed of 10 m/s. The following sections evaluate each of these terms in depth.

Table 1. The magnitudes of 1-sigma errors for each term in Equation (4). The shaded rows indicate that the absolute magnitudes of these terms are negligible compared to the dominant error terms and are neglected in the construction of the error model in this work.

Error Term	Error Magnitude [dB]
$E(G^R)$	0.43 [25]
$E(P^g)$	0.23 [35]
$E(G^Z)$	0.20 [33]
$E(P^Z)$	0.18 [33]
$E(\zeta^E)$	0.15 [33]
$E(\overline{A})$	0.05 [34]
$E(L^{atm})$	0.04 [34]
$E(R^{tot})$	<0.01 [34]
$E(L^Z)$	<0.01 (assumed $\sim E(R^{tot})$)

2.2. The Bottoms-Up Correlated Error Model

The CYGNSS Level 1 correlated error model is constructed to represent the major error sources based on the physical and operational characteristics of the CYGNSS observatories. These errors are correlated over both space and time, and depend on several variables relating to the operation of the CYGNSS constellation.

The construction of the error model is intended to be realistic enough to represent measurable and plausible correlated errors, yet flexible enough to allow for tuning and parameterization. The general modeled structure of correlated error takes the form:

$$R_{mod}(i,j) = \frac{1}{\mathcal{N}} \sum_n K_n(i,j), \tag{5}$$

where $R_{mod}(i,j)$ is the normalized correlated error between any two samples i and j with a domain $-1 \leq R_{mod} \leq 1$. There are n component terms that are added together, and each of the nth term corresponds to a unique source of error. $K_n(i,j)$ represents the error covariance between any two samples i and j for a specific component n. $\mathcal{N}$ is a normalization constant to ensure that the diagonal terms of the R_{mod} matrix is 1.

Each of the K_n terms can be further generalized as follows:

$$K_n(i,j) = (E(n))^2 \cdot Ecorr_n(i,j)\ , \tag{6}$$

where $E(n)$ represents the magnitude of each error component n as calculated in [33,34,36] and can be thought of as the variance of the error, and $Ecorr_n(i,j)$ is a derived function with a domain $-1 \leq Ecorr \leq 1$ that represents the correlation in error between samples i and j for each component n.

Each *Ecorr* function depends on different arguments depending on the nature of the error source. The individual errors are explored in depth in the following sections. Further, several tuning parameters have been added to assist with validation. The initial values of the tuning parameters are 1 and essentially leave the model output unmodified. However, this process assumes that the relative magnitude of the error components could be incorrectly specified, and by varying the tuning parameters, the model can be shaped to match a validation dataset.

2.2.1. Model Assumptions

Several assumptions are made regarding the overall model and each individual *Ecorr* term. The first is that the decorrelation scales of interest are on the order of seconds to minutes. This restricts the problem to errors that decorrelate on timescales that would be of relevance to numerical weather prediction users. CYGNSS observations may have a correlated error structure that evolves during longer timescales, say, daily or seasonally. However, this structure would probably be better resolved using other calibration and processing techniques and would not add significant value to a correlated error matrix. As a result, all errors between CYGNSS samples i and j are assumed to be 100% uncorrelated if there is more than 10 min of separation in observation time. This corresponds to roughly 4500 km of distance, where observations could reasonably be treated as wholly uncorrelated.

Further, all error terms $Ecorr_n$ are assumed to be uncorrelated with each other. This assumption simplifies the implementation and calculation of the model but is less supported theoretically. There are a number of plausible rationales for why many of these error sources are correlated. However, this model hopes to correct for these empirically with the implementation of tuning parameters.

Finally, there is a general assumption that all the error sources in the model exhibit wide-sense stationarity during the timescales of interest. Therefore, the error magnitude values $E(n)$ are treated as constants. In practice, the magnitude of errors and their correlation time and space scale sizes may well vary (e.g., due to seasonal dependence). A fully operational implementation of the methodology developed here would take these dependencies into account, e.g., by an appropriate parameterization of the properties of the errors. Here, we only consider stationary error properties in order to illustrate how they are determined and how they would be combined to be used by a DA scheme.

2.2.2. The Full Correlated Error Model

Considering the terms in Table 1 and applying Equation (5), the Level 1 correlated error model is constructed:

$$R_{mod}(i,j) = \frac{1}{\mathcal{N}} \begin{pmatrix} R_{P\mathcal{g}}(i,j) + \\ R_{PZ}(i,j) + \\ R_{G^R}(i,j) + \\ R_{G^Z}(i,j) + \\ R_{\zeta}(i,j) \end{pmatrix}, \tag{7}$$

where $R_{P\mathcal{g}}$ is the correlated error component from the calibrated nadir receiver power and is discussed in Appendix B; R_{PZ} is the correlated error component from the zenith receiver and is discussed in Appendix C; R_{G^R} and R_{G^Z} are the correlated error components from the nadir and zenith CYGNSS antenna patterns, respectively, and are discussed in Appendix D; and R_{ζ} is the correlated error component due to the zenith–specular ratio used for dynamic EIRP estimation and is discussed in Appendix E.

2.3. Verification Techniques

One of the primary challenges in constructing the CYGNSS Level 1 correlated error model R_{mod} is identifying a plausible validation scenario. In practice, it can be difficult to disentangle the various sources of correlated observation error structure, e.g., those caused by the inherent behavior and calibration of the instrument, by the geophysical retrieval and inversion process, and by the representativity errors imposed when observations are gridded and ingested into models. Further, the choice of ground truth may impose an additional source of correlated error, such as when using reanalysis data or another observation source.

In an effort to disentangle these correlated errors and isolate only those due to instrumental sources, this work validates the correlated error structure R_{mod} by matching up near-simultaneous collocated observations made by two different observatories at nearly

identical geometries, generating modeled NBRCS σ^o_{mod} from reanalysis data, and single- and double-differencing the results.

2.3.1. Curating Matchup Observations

As a constellation of eight small satellites in the same orbital plane, CYGNSS is generally unable to collect collocated, near-simultaneous observations from different satellites. If the CYGNSS observatories were equally distributed across the orbital plane, each asset would follow the next at approximately a 10 min lag. However, in that time, both of the uniquely defining features of the CYGNSS observation changes: (1) the surface of the Earth rotates underneath the observatories, changing the ground track, and (2) the GPS satellites that serve as the source of radar signal advance in their orbit. Therefore, in 10 min, it can be quite challenging to develop one-to-one matchup conditions suitable for investigating the correlated error structure at timescales of seconds to minutes.

CYGNSS has no onboard thrusters, and orbit phasing is controlled solely through differential drag. At several junctures during the CYGNSS mission, one of the observatories advanced within the plane to be nearly overlapping with another yet at a slightly different altitude. Generally, the greater the altitude difference between the observatories, the faster the relative precession within the orbit plane.

An exhaustive review of each observatory's ephemerides for the life of the mission identified a few matchup opportunities. Matchups near the beginning of the mission were preferred, as the orbit planes of the satellites tend to drift apart over the lifetime of the mission. After further filtering by the operational status of each observatory to ensure that they were in similar attitude configurations and operating in similar science modes, a 24 h period starting 0Z on 11 SEP 2019 was chosen, where FM1 and FM5 were in a nearly identical orbital phase for a sustained period. A representation of the ground sampling during this period is shown in Figure 1.

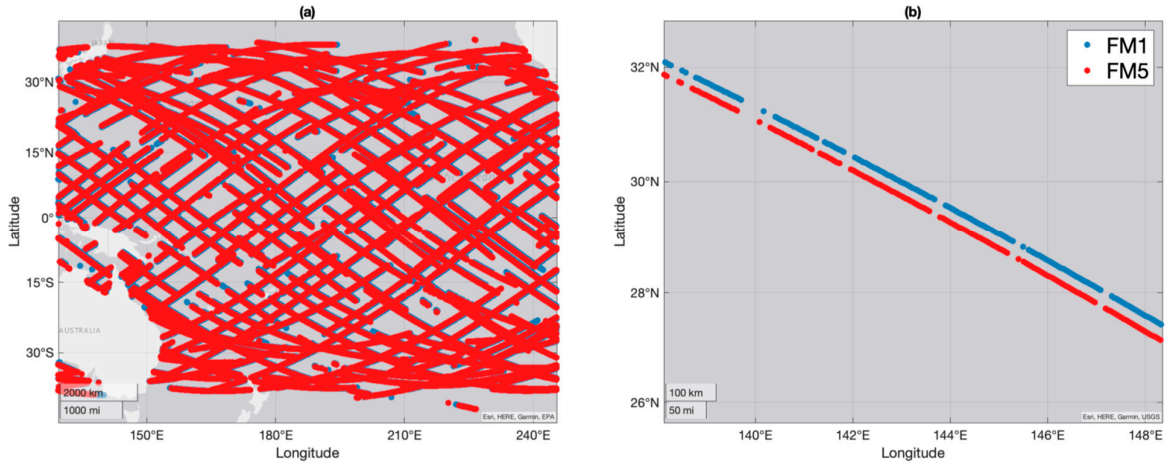


Figure 1. Comparison of FM1 and FM5 observations for 11 SEP 2019. (**a**) Rendering of the ground samples captured during the 24 h period in the near-overlap condition; (**b**) A single track of samples for both FM1 and FM5 that has been matched sample-for-sample to facilitate one-to-one comparisons of the observations. The samples from each observatory are no more than 0.5 deg apart in distance and were acquired approximately 3 s apart in time.

Each 'track' of observations (when a series of observations are made in close succession sharing a CYGNSS receiver and GPS transmitter) was matched sample-for-sample between FM1 and FM5 by first minimizing the distance between observations and then quality controlling for several factors:

- Matched tracks must both contain more than 300 samples;
- Individual sample matchups are valid if samples are within 0.5 degrees (great circle distance);

- Individual samples are screened to ensure no quality control flags apply; and
- The matched track is only valid if 60% of the data remains after all other matchup criteria apply.

The resulting matchup conditions produced 103 matched tracks within the 24 h period of near-simultaneous observations.

2.3.2. Generating Model NBRCS

A forward model for the CYGNSS observatory (generating σ^o_{mod} from wind data) can be challenging. Operationally, CYGNSS uses a GMF LUT that maps between the two quantities. However, for the purposes of this analysis, we prefer to use a physically representative forward operator that represents the physical dynamics of the roughening of the ocean surface due to locally driven winds.

For each sample matched up, σ^o_{mod} is generated by computing a spectrally corrected ocean surface MSS from an ERA5-forced [37] WaveWatch III wave model as described in [38]. Because the resolution of ERA5 is much coarser than CYGNSS Level 1 observations in both space and time, the σ^o_{mod} is matched up with σ^o via a tri-linear interpolation across three dimensions (latitude, longitude, and time).

Therefore, for every track of data, there are observations from two different CYGNSS observatories and two modeled NBRCS from reanalysis data.

2.3.3. Estimating Total Correlated Error

Single- and double-differencing are common techniques to calibrate remote sensing instruments, especially radiometers [39,40] and radars [41]. Both are useful for quantifying the correlated error structures in CYGNSS observations.

For each track of data, there are two single-differenced datasets,

$$SD_{obs} = \sigma^o_1 - \sigma^o_5 \tag{8a}$$

$$SD_{mod} = \sigma^o_{mod,1} - \sigma^o_{mod,5} \tag{8b}$$

where SD_{obs} is the single-differenced data from the two matched-up observations from FM1 and FM5, and SD_{mod} is the single-differenced data of the model-generated NBRCS for the specific coordinates of FM1 and FM5. The double-difference is computed by differencing these two quantities:

$$DD = SD_{obs} - SD_{mod} \tag{9}$$

Both the SDs and DD are useful for our analysis. The SD_{obs} term represents the difference in σ^o measured by two different CYGNSS satellites. Despite the strict matchup conditions, these assets are still measuring different fields of view at different times. These differences may result in some residual correlated structure. In addition, any correlated structure in SD_{obs} may not reveal structure imposed from fundamental properties of the earth system. Therefore, SD_{mod} allows us to identify the correlated structure of any systematic differences in observation target. An example of the utility of these differencing techniques is shown in Figure 2.

Since we are primarily interested in correlated error, and not absolute error, our primary investigative tool is the autocorrelation function. The correlated error for each track can be computed by autocorrelating the DD using the standard formulation:

$$\rho(\tau) = \frac{1}{(N-\tau)} \frac{1}{\sigma_{x_i}} \frac{1}{\sigma_{x_{i+\tau}}} \sum_{i=1}^{N-\tau} (X_i - \bar{X})(X_{i+\tau} - \bar{X}) \tag{10}$$

where τ is the lag, N is the total number of lags observed, X_i is the value of timeseries X at time i, $\bar{X}$ is the mean of timeseries X, and the standard deviation is formulated as normal, $\sigma_i = \sqrt{\frac{\sum_{i=1}^{N-\tau}(X_i - \bar{X})^2}{N-\tau}}$.

Because the error correlation behavior can vary from track-to-track, we also construct a bulk autocorrelation which approximates the autocorrelation behavior for all tracks sampled. For this we consolidate M tracks, the bulk autocorrelation is:

$$\mathcal{R}(\tau) = \frac{1}{M(N-\tau)} \frac{1}{\sigma_{y_i}} \frac{1}{\sigma_{y_{i+\tau}}} \sum_{i=1}^{N-\tau} (Y_i - \bar{Y})(Y_{i+\tau} - \bar{Y}) \tag{11}$$

where $Y_i = \sum_{j=1}^{M} X_{i,j}$, j is the index for track number, and the standard deviation and mean are calculated as before, but with the consolidated track series Y.

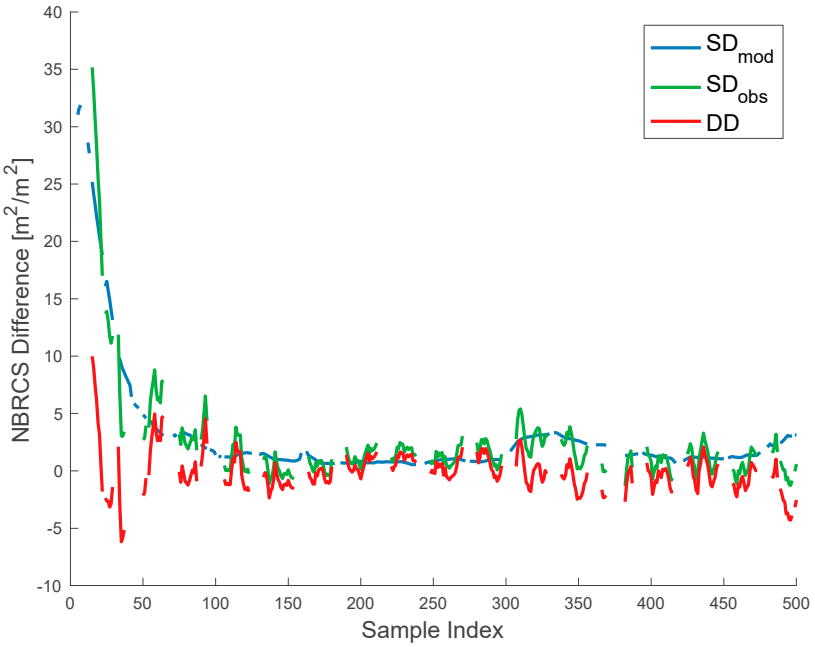


Figure 2. A display of model (blue) and observed (green) single-differenced matchup data for a single CYGNSS track, with the double-differenced data (red) overlaid. Note that both the model and observed single differences are similarly high early in the track, suggesting that the differences in the observations may be because the samples are observing fundamentally different targets, captured in SD_{mod}. The double-difference accounts for these types of errors, and for the entire track, the double-difference is quite stable. The sparsity of data is caused by quality control parameters, such as when the CYGNSS observatory is performing its onboard calibration procedure, which occurs once every minute. Because the two matched observatories do not have synchronized calibration clocks, the matched dataset typically flags out two of these cadences for every minute of sampling.

2.3.4. Model Tuning

The constructed correlated error model $R_{mod}(i,j)$ is designed to estimate the correlated error between arbitrary samples i and j. Validating the correlated error with empirical matches is challenging since the empirical error correlation can only be estimated in a broader, statistical sense. To create an appropriate comparison, we introduce the *modeled* error autocorrelation $\widetilde{\rho}_{mod}$, which is a function of lag τ:

$$\widetilde{\rho}_{mod}(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} R_{mod}(i, i+\tau), \tag{12}$$

where R_{mod} is calculated using Equation (7), τ is the lag, and N is the total number of lags observed. The quantity $\widetilde{\rho}_{mod}$ is reasonably comparable to the autocorrelation $\rho(\tau)$ of observed error for a single track of samples. A similar analog can be made for the *bulk*

modeled error autocorrelation $\widetilde{\mathcal{R}}_{mod}(\tau)$, which estimates the autocorrelation behavior of the error model across M tracks:

$$\widetilde{\mathcal{R}}_{mod}(\tau) = \frac{1}{M}\sum_{j=1}^{M} \widetilde{\rho}_{mod,j}(\tau), \quad (13)$$

where $\widetilde{\rho}_{mod,j}(\tau)$ is the modeled autocorrelation for track j at lag τ.

A parametrized version of R_{mod} is constructed using the following form:

$$R_{mod}(\alpha,\beta,\gamma,\delta)|_{i,j} = \frac{1}{\mathcal{N}}\left[R_{P\mathcal{S}}(\alpha,\beta) + R_{PZ}(\alpha,\beta) + R_{G^R}(\gamma,\delta) + R_{G^Z}(\gamma,\delta) + R_{\zeta}(\beta)\right], \quad (14)$$

where the following is true:

- α represents the relative magnitude of the white noise component of the error, which decorrelates at $\tau = 1$;
- β represents the relative magnitude long-decay pedestal or any residual correlated errors at the edge of our timescales of interest;
- γ represents the relative magnitude of the correlated error caused by the terms R_{G^R} and R_{G^Z}, which exhibit smooth decay as samples spread apart when projected through the nadir and zenith antenna coordinates, respectively [see Appendix D for an in-depth discussion]; and
- δ represents the relative decorrelation roll-off in terms R_{G^R} and R_{G^Z}.

With an appropriate benchmark, the tuning parameters α, β, γ, and δ are iterated such that $\widetilde{\mathcal{R}}_{mod}$ matches a target signal. The tuning parameters are applied such that they can be modified to change specific characteristics of the modeled behavior:

The parameters modify R_{mod} directly and are described in depth in Appendices B–E. The target for matching the modeled autocorrelation $\widetilde{\mathcal{R}}_{mod}$ is the bulk autocorrelation of the double-differenced data DD:

$$\mathcal{R}[DD](\tau) \cong \widetilde{\mathcal{R}}_{mod}(\tau) \quad (15)$$

Because this is an under-determined system, there is no unique solution to optimizing the tuning parameters. Further, these parameters all relate to one another, and changing one will impact the others. Instead, R_{mod} is tuned heuristically such that the modeled behavior matches the empirical data at key points: to match the white noise component at lag $\tau = 1$; to match the rolloff at lags $\tau = 5$ and $\tau = 30$; and to match the endpoint behavior at lag $\tau = 100$.

3. Results

3.1. Bulk Behavior

The constructed model R_{mod} is first tuned to match the overall behavior of $\mathcal{R}[DD](\tau)$, as described in Section 2.3.4. With the appropriate tuning parameters, the bulk-modeled autocorrelation $\widetilde{\mathcal{R}}_{mod}$ closely resembles $\mathcal{R}[DD](\tau)$ in most important aspects, including the relative magnitude of the white noise component, the relative rate of decorrelation roll-off, and endpoint behavior at large lags. The final tuned $\widetilde{\mathcal{R}}_{mod}$ is shown in Figure 3 with $\mathcal{R}[SD_{obs}]$, $\mathcal{R}[SD_{mod}]$, and $\mathcal{R}[DD]$.

The behavior in Figure 3 is worth discussing in detail. The single-differenced model generated σ^2_{mod} decays at a much slower rate than the single-differenced observations, and double-differenced data suggests that the errors of slight mismatch in observation location and time are generating errors on a fundamentally different scale than the instrument errors. Further, the fact that the double-differenced decorrelation is almost identical to the observation single-differencing suggests that single-differencing near-simultaneous observations are a reasonable approximation for bulk error correlation.

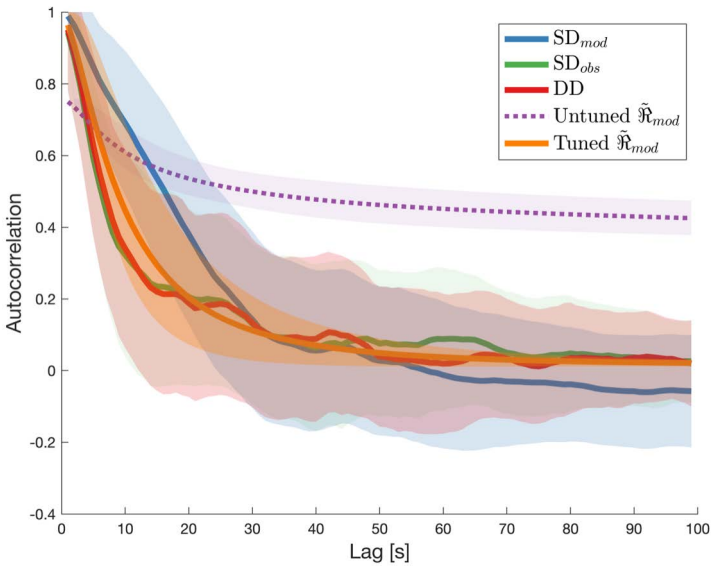


Figure 3. Comparison of the bulk error autocorrelation behavior for all matched tracks. The single difference of modeled σ^o_{mod} (blue), the single difference of observed σ^o (green), the double-difference DD (red), the untuned bulk model error correlation $\widetilde{\mathcal{R}}_{mod}$ (dotted purple), and the final tuned bulk model error correlation $\widetilde{\mathcal{R}}_{mod}$ (orange) are shown. The shading indicates the estimated 1-sigma standard deviation of the population decorrelation behavior for each individual trace. The lags are computed in seconds, which correspond to single samples at 1 Hz sampling for the CYGNSS observatory. Note that DD generally follows the SD_{obs} trace, suggesting that single-differencing for this particular use case may be a reasonable representation of bulk error correlation behavior.

Evaluating the double-difference trace in Figure 3 also reveals important qualities of the CYGNSS error correlation structure. First, the uncorrelated error component accounts for roughly 5% of the total system error. Therefore, assumptions that samples may be treated as independent with uncorrelated error are empirically refuted. Further, the error decorrelation roll-off is swift: about 50% of the error decorrelates within 7 s of observation. Using a flat-plane approximation of Earth, this corresponds to a distance of approximately 50 km, which is generally the scale of the gridding of modern global weather models but much coarser than the resolution of state-of-the-art regional models. Finally, the endpoint behavior at large lag times suggests a small, positive correlated error (~2%) at larger timescales. The statistical properties of the lag correlation become less stable at larger lags, but the existence of a long-timescale correlation is not surprising, considering all observations from a single receiver share a common electronics and processing chain.

The fact that our tuned model R_{mod} can approximate the bulk error correlation structure $\widetilde{\mathcal{R}}_{mod}$ with similar features as $\mathcal{R}[DD]$ validates the assumption that the overall instrument-correlated error can be modeled from the fundamental components of the instrument observable: in our case, the radar-range equation (Equation (2)). Further, that R_{mod} can be generated between arbitrary samples suggests that an observation error correlation matrix $\mathbf{R}$ can be generated dynamically from first principles given appropriate knowledge about the instrument and retrieval.

The value of the chosen tuning parameters is also worth investigating. The untuned model is defined by having the parameters $\alpha = \beta = \gamma = \delta = 1$. Figure 3 demonstrates that the modification of tuning parameters can change the overall model behavior significantly to match observed behavior. Figure 3 further suggests that the untuned R_{mod} generally overestimates both the relative magnitudes of uncorrelated error (i.e., white noise, tuned by α) and totally correlated error (i.e., endpoint correlation, tuned by β). The chosen parameters for tuning are shown in Table 2.

Table 2. The chosen magnitudes of model tuning parameters for R_{mod}.

Tuning Parameter	Magnitude	Function
α	0.005	Relative magnitude of uncorrelated error
β	0.01	Relative magnitude of endpoint correlated error
γ	1	Relative magnitude of nearby roll-off
δ	1	Steepness of roll-off component

The tuning parameter γ is used to vary the relative magnitude of the near-sample correlated error roll-off. In our tuned model, this parameter remains at 1. The tuning parameter δ is used to enhance or reduce the steepness of the roll-off in correlation caused by two observations moving farther apart in the nadir and zenith coordinate systems. If $\delta > 1$, it enhances the steepness of decay. The fact that our optimized tuning maintains $\delta = 1$ suggests that the filtering theory discussed in Lemma A1 is a reasonable model of decay.

3.2. Single-Track Comparisons

The tuned model can be compared to single-track autocorrelations $\rho(\tau)$ of matched-sample double-differences using the modeled error autocorrelation $\widetilde{\rho}_{mod}$ as described in Equation (12). We compare three exemplar tracks in Figure 4.

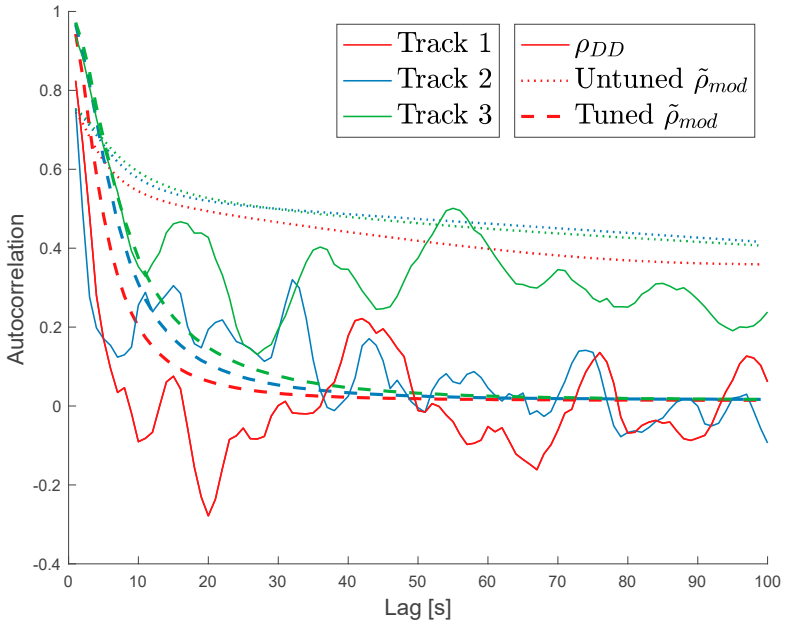


Figure 4. Comparison of the single-track error autocorrelation behavior for three selected tracks. The autocorrelation of the NBRCS double-difference (solid lines), untuned modeled error correlation $\widetilde{\rho}_{mod}$, and tuned modeled error autocorrelation $\widetilde{\rho}_{mod}$ (dashed lines) for each of the three tracks are shown. Each track is painted a different color. This figure demonstrates the wide variability of single-track autocorrelations of double-differences, arising from the variability of the observable, the limited amount of data in a single track, and the challenges in quality controlling sufficient observation data. It is possible that the correlated error does vary this much from track to track. In contrast, both the tuned and untuned modeled error autocorrelation $\widetilde{\rho}_{mod}$ are much more stable from track to track.

The single-track autocorrelation behavior shown in Figure 4 is also revealing of the limitations of this work. The single-track error autocorrelation for double-differenced data is highly variable. This may be due to both artifacts in the data (processing, quality control, insufficient data), as well as the real behavior of the observation. It is worth articulating that the autocorrelation of data is generally less stable with smaller datasets and at longer lags. With our quality control parameters, we flag out significant quantities of data, which

decreases the stability of autocorrelations from track to track. We further note the difficulty in exploring the dynamics of this behavior in a statistical sense: our aperture of observation where two CYGNSS assets are in a near-overlap condition is quite rare, and as the mission progresses, the orbit planes of the satellites have drifted, making further matchup scenarios less representative. These behaviors may evolve differently day-to-day or as a function of observed wind speed, but the paucity of data in this configuration makes it challenging to establish sufficient baselines to test for significance.

3.3. Dynamic Correlated Error Estimation and Impact of Tuning

The model R_{mod} can produce plausible dynamics, suggesting the broader importance of a realistic dynamic correlated error model. Figure 5 plots two nearby tracks captured nearly simultaneously by the same receiver (but different GPS transmitters).

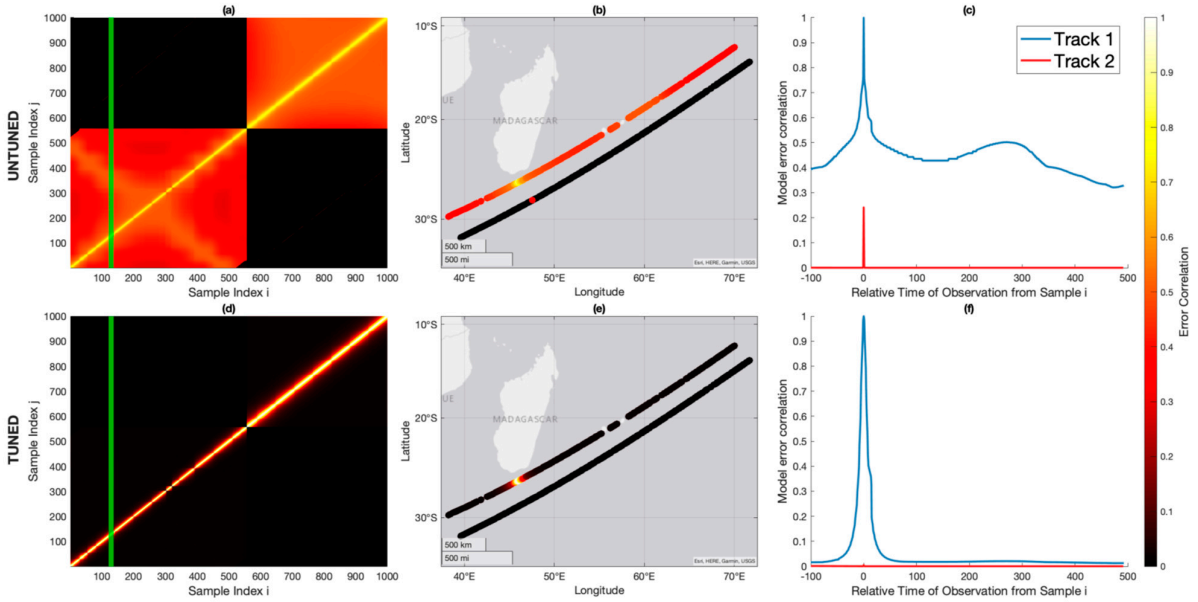


Figure 5. Generated untuned and tuned R_{mod} for two nearby tracks captured by the same CYGNSS receiver, each with three different representations. (**a**) Untuned R_{mod} as represented by sample index. The two tracks are concatenated in a vector, and the modeled error correlation is calculated for each combination of indices. The green vertical line represents an exemplar index (i = 129) along the track, where the correlated error is estimated for every other observation in the neighborhood. (**b**) Untuned R_{mod} for the same exemplar index as represented by the physical location of sample acquisition. (**c**) Untuned R_{mod} for the same exemplar index represented by time of acquisition. The traces for the two different tracks are plotted in different colors. The error correlation along the same track is in blue, while the error correlation for the adjacent track is in red. (**d**) Tuned R_{mod} as represented by sample index for the same two tracks as in (**a**). (**e**) Tuned R_{mod} for the same tracks represented in the physical location. (**f**) Tuned R_{mod} for the same exemplar index represented by the time of acquisition. Note that for both the untuned and tuned cases, the correlated error for the adjacent track is non-zero but generally very small compared to the error correlation along the track.

Because the correlated error R_{mod} can be generated for any arbitrary observation using the bottoms-up model, the dynamics of how instrument errors decorrelate can be explored. To illustrate, we choose an arbitrary exemplar index (i = 129) to demonstrate that this calculation can be performed for any sample within a track. If the observations were treated as completely independent without any correlated error, the matrix in Figure 5a would contain non-zero elements exclusively along the main diagonal. However, we observe several structural elements. The 'pixelation' pattern is largely a result of the R_ζ term and originates from the coarse mapping of the estimated GPS antenna gain pattern as a function of scattering incidence θ_{inc}. This phenomenon is explored in depth in Appendix D.

The smooth decorrelation roll-off is from the R_{G^R} and R_{G^Z} components of the model and represent the direct and reflected signals moving about the nadir and zenith antenna patterns in CYGNSS receivers as the observation is collected along the track. We note that the model allows for cross-track correlation where the two different tracks share a CYGNSS receiver but have a different GPS transmitter. In both the untuned and tuned cases, these tracks appear to have uncorrelated cross-track error. There are residual correlations between the tracks when the scattering geometry is such that the two tracks share similar incidence angles or are near similar antenna coordinates, in addition to the shared receiver noise at lag $\tau = 0$.

We can also determine R_{mod} for our matchup tracks where CYGNSS observes similar track geometries with the same GPS transmitter but with two different receivers. The correlated structure for one of the tracks in the previous example (but matched by different receivers) is shown in Figure 6.

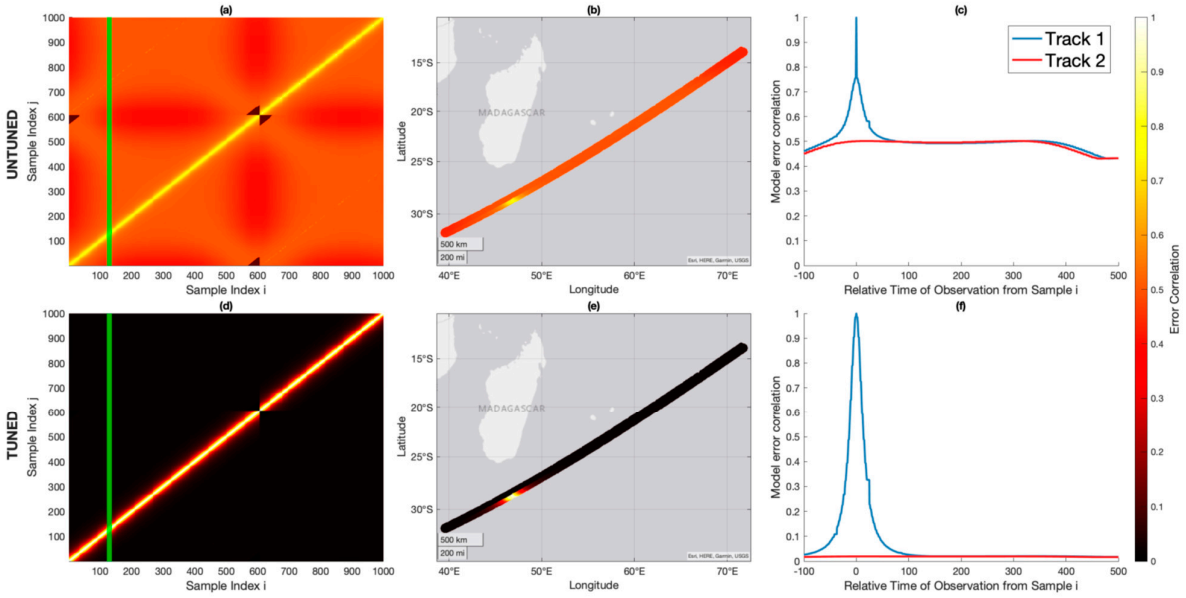


Figure 6. Generated untuned and tuned R_{mod} for two nearby tracks captured by two different CYGNSS receivers during the overlap period, each with three different representations. (**a**) Untuned R_{mod} as represented by sample index. The two tracks are concatenated in a vector, and the modeled error correlation is calculated for each combination of indices. The green vertical line represents an exemplar index (i = 129) along the track where the correlated error is estimated for every other observation in the neighborhood. The correlated error model allows for correlated errors across receivers. (**b**) Untuned R_{mod} for the same exemplar index as represented by the physical location of sample acquisition. (**c**) Untuned R_{mod} for the same exemplar index represented by the time of acquisition. The traces for the two different tracks are plotted in different colors. The error correlation along the same track is in blue, while the error correlation for the adjacent track is in red. (**d**) Tuned R_{mod} as represented by the sample index for the same two tracks as in (**a**). (**e**) Tuned R_{mod} for the same tracks represented by the physical location. (**f**) Tuned R_{mod} for the same exemplar index represented by the time of acquisition. Note that the tuning has virtually eliminated cross-track error correlation.

The model R_{mod} allows for correlated error between two different receivers that share a GPS transmitter. We note that observations sharing the same GPS transmitter may contain correlated error due to the correlated misestimation of GPS transmit power. This may be an incomplete articulation of the full cross-receiver error correlation. For example, CYGNSS assets in similar orbital regimes may experience similar thermal and environmental conditions that cause correlated errors during our timescales of interest. We assume that correlated error due to misestimation of GPS transmit power is nearly constant for the timescales of interest.

As demonstrated in Figure 6, the cross-track correlation virtually disappears after tuning the model R_{mod}. Taken with the data presented in Figure 5, there appears to be negligible cross-track error correlation, either when two tracks share a CYGNSS receiver or when two tracks share a GPS transmitter. This is a novel result but challenging to validate in practice, as the double-differencing may eliminate any shared error correlation between two tracks.

4. Discussion

R_{mod} was designed to maintain several key features. First, R_{mod} can be generated for arbitrary samples in any dataset given the appropriate inputs. Second, R_{mod} is designed to maintain traceability from component error sources with reasonable physical assumptions. Third, R_{mod} is designed to be tuned to allow for calibration and validation as more data becomes available. Finally, R_{mod} is intended to be implementable in a dynamic error model without significant computational expense.

4.1. Limitations

The correlated error model R_{mod} exhibits complex dynamics that are generally plausible but difficult to validate outside of the large-scale statistical estimation discussed in Section 3.1. The fact that the bulk statistical behavior between the estimated autocorrelation model averaged over all tracks is consistent with the observed double-difference autocorrelation (cf. Figure 3) is reassuring, but the disparity between the individual track model estimated autocorrelations and the single-track double-difference estimations (c.f. Figure 4) suggests that either the model is insufficient at capturing the dynamics track-to-track or that the correlated error may not be as stationary as assumed.

With only about a day of data during the overlap period, there remains the possibility that the validation dataset is unrepresentative of the overall statistical behavior of the observation. R_{mod} is designed only to account for errors at short spatial and temporal scales relevant for data assimilation purposes; therefore, this model may not necessarily account for changes of the underlying statistical distributions at larger scales due to seasonality or orbital precession.

The error model assumes that errors are generally isotropic and that $R_{mod}(i,j) = R_{mod}(j,i)$. This assumption has important practical utility for simplifying the implementation of R_{mod} and the conditioning of the required matrix inversion. Anisotropies may exist in any of the components of the error model but are assumed to be small compared to the modeled behavior.

Further, this model generally assumes no spatial dependence. Spatial dependence of the correlated error is primarily driven by the errors in the retrieval and not the instrumentation, which orbits Earth every 95 min. The 'spatial' dependence of error from CYGNSS transiting its orbit has largely to do with the dynamics of the thermal loading of the receiver as lighting conditions evolve in orbit. This is accounted for in the assumptions of the component models R_{PS} and R_{PZ} as discussed in Appendices B and C, respectively. The correlated error resulting from the evolution of the observing geometries as both GPS and CYGNSS propagate in their orbits is captured in the correlated structure of the antenna gain maps and the zenith–specular ratio explored in Appendices D and E.

4.2. Impact of Tuning Parameters

The application of tuning parameters is a design decision to capture the fact that initial estimates of the bottom's up error magnitudes may be incomplete and to facilitate rapid adjustments as new calibration and validation information becomes available. The tuning parameters attempt to weigh the value of individual component errors in comparison to each other based on aggregate data from one day of observations during the overlap period.

The specific values of the tuning parameters recovered suggest that the correlated error model R_{mod} overestimates both the uncorrelated and highly correlated components of instrument error. This is useful information for future studies because, as an opportunistic

measurement, GNSS-R has limited insight into the error structures in the source signals from GPS.

For observations within a single track, these parameters provide evidence that nearby samples in space and time can add significant new information to a forecast. The tuning parameter $\alpha = 0.005$ indicates that the uncorrelated component of the error is significantly overestimated, suggesting that the constituent error magnitudes in [36] may be overestimated. Further, the fact that the tuning parameters γ and $\delta = 1$ suggest that the overall structure of the error correlation decay is consistent with the theory posited in Appendix D (that the primary source of error correlation is from the application of smoothing filters in the production of antenna gain patterns) and that the application of the Filtering Lemma (Lemma A1) is reasonable.

For observations between tracks, the tuned error model R_{mod} indicates that errors are nearly uncorrelated between tracks (cf. Figures 5 and 6). This is driven primarily by the tuning of the parameter $\beta = 0.01$. This has significant practical utility to future assimilation strategies for CYGNSS, suggesting that two nearby tracks are essentially independent measurements with independent instrument error (correlated error may still result from correlated errors in the retrieval or representation).

Finally, it should be noted that the overall effect of the tuning somewhat obviates the need for a dynamic correlated error model, as the overall behavior of the tuned model is quite stable from track to track (cf. Figure 4). This suggests a simplified instrument-correlated error model could be derived from the bulk behavior of $\widetilde{\rho}_{mod}$, reducing many of the required input data.

5. Conclusions

This work produces a first-principles estimate of correlated instrument error with results that approximate observed statistical behavior. We believe that this model presents a significant advancement in the estimation of the spatial and temporal correlation structure of instrument error for remote sensing systems and, in particular, for the unique considerations of GNSS-R measurements by CYGNSS.

In essence, this work answers a theoretical exercise for enumerating the plausible engineering reasons why two data points from the same observing constellation can share correlated sources of error. We evaluate the correlated error structure for CYGNSS by examining the potential plausible sources of correlated structure from individual components of the instrumentation, combining these sources from first principles as a tuned engineering model, and evaluating the efficacy of the model via a robust validation during a period when two satellites with nearly identical observing geometry captured near-simultaneous and near-collocated samples.

The instrument-originating sources of correlated error is likely to be a small component of the overall correlated structures of observation error. For instance, a likely significant source of correlated error structure is the Geophysical Model Function retrieval that converts observed NBRCS to surface wind speed. These errors can be multifaceted, both encompassing representation error, as the ground truth for training this retrieval is reanalysis data, which may not capture the spatial or temporal dynamics of wind speed [27]. A companion work will explore how the retrieval that maps from NBRCS to wind speed produces correlated error structures in space and time.

Further, the fact that the single-differenced observations generally drive the double-differenced autocorrelation behavior suggests that Powell et al.'s metric of simultaneity and collocation [42] generally applies for CYGNSS when satellites are in a near overlap condition. This interesting result suggests that double-differencing may not be required for statistical estimations of GNSS-R errors given a sufficiently large dataset of observations that nearly overlap. This implication may relax further validation requirements of R_{mod} and may enable near-real-time calibration strategies when samples are sufficiently close without the need to generate reanalysis-driven forward model NBRCS, which introduces significant latencies.

Finally, while the correlated error model R_{mod} was designed to compute the estimated correlated error from arbitrary CYGNSS samples, this may be unnecessary in practice as the tuning makes the model quite stable between samples and tracks. Further, given sufficient assumptions about the stationarity of the instrument-correlated error and the overall magnitude of the instrument-originating sources of correlated error, this information could be conveyed as a static look-up table.

Author Contributions: Conceptualization, C.E.P. and C.S.R.; methodology, C.E.P.; software, C.E.P., D.S.M., T.W. and A.R.; validation, C.E.P.; formal analysis, C.E.P.; investigation, C.E.P.; resources, C.S.R.; data curation, C.E.P. and A.R.; writing—original draft preparation, C.E.P.; writing—review and editing, all authors; visualization, C.E.P.; supervision, C.S.R.; project administration, C.E.P. and C.S.R.; funding acquisition, C.S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Aeronautics and Space Administration Science Mission Directorate under Contract NNL13AQ00C with the University of Michigan and Contract 80NM0018D0004 with the California Institute of Technology Jet Propulsion Laboratory.

Data Availability Statement: Publicly available datasets were analyzed in this study. CYGNSS Level 1 data can be found at NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC) [43]. ERA5 reanalysis can be found at [44].

Acknowledgments: We acknowledge the insightful feedback of two anonymous reviewers who improved the quality of this manuscript. All remaining errors and deficiencies are the authors' own.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The construction of R_{mod} makes a few important assumptions that are worth examining in depth. First, we will describe the general mathematical framework. In general, if X and Y are random vectors of length N, then the covariance matrices of X and Y are constructed:

$$K_{XX} = \langle X - \langle X \rangle \rangle \langle X - \langle X \rangle \rangle^T, \tag{A1a}$$

$$K_{YY} = \langle Y - \langle Y \rangle \rangle \langle Y - \langle Y \rangle \rangle^T, \tag{A1b}$$

where K_{XX} and K_{YY} are the N-by-N covariance matrices of X and Y, respectively, the bracket operator denotes the expectation operation, and the superscript T denotes the transpose vector. Notably, the covariance matrix is symmetric and positive semi-definite. The covariance matrices can also be constructed in the following manner:

$$K_{XX} = S_X R_{XX} S_X, \tag{A2}$$

where S_X is a diagonal N-by-N matrix with the standard deviations of X_i:

$$S_X = \begin{bmatrix} \sigma_{X1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{XN} \end{bmatrix}, \tag{A3}$$

and R_{XX} is the correlation matrix with correlation coefficients $\rho_{i,j}$:

$$R_{XX} = \begin{bmatrix} 1 & \rho_{X1,X2} & \cdots & \rho_{X1,XN} \\ \rho_{X2,X1} & 1 & & \\ \vdots & & \ddots & \\ \rho_{XN,X1} & & & 1 \end{bmatrix} \tag{A4}$$

As standard, $-1 \leq \rho_{i,j} \leq 1$, and the diagonals all must equal 1. The *cross*-covariance matrices can also be defined in a similar fashion:

$$K_{XY} = \langle X - \langle X \rangle \rangle \langle Y - \langle Y \rangle \rangle^T, \tag{A5a}$$

$$K_{YX} = \langle Y - \langle Y \rangle \rangle \langle X - \langle X \rangle \rangle^T, \tag{A5b}$$

where K_{XY} and K_{YX} are the cross-covariance matrices and are not generally identical. The error magnitudes in Table 1 are calculated using a standard error propagation technique as highlighted in Gleason et al. [25]:

$$E_n = \left| \frac{\partial F}{\partial n} \right| \Delta n \tag{A6}$$

E_n is the error magnitude of term n, Δn is the estimated 1-sigma dynamic range of n and F is an arbitrary function. We can propagate these errors as the sums of covariance matrices with the derivation below, inspired by Chapter 9 in Taylor [45].

For an arbitrary function F with random vector arguments X and Y, $F(X, Y)$, where $X = (X_1, X_2, \ldots, X_N)$ and $Y = (Y_1, Y_2, \ldots, Y_N)$, we can approximate F by its first order Taylor Series expansion about the mean value, assuming that the errors are generally small compared to the arguments:

$$F_i = F(X_i, Y_i) \tag{A7}$$

$$F_i \approx F(\langle X \rangle, \langle Y \rangle) + \frac{\partial F}{\partial X}(X_i - \langle X \rangle) + \frac{\partial F}{\partial Y}(Y_i - \langle Y \rangle) \tag{A8}$$

Noting that $\langle F \rangle = F(\langle X \rangle, \langle Y \rangle)$, and using $F \approx F(\langle X \rangle, \langle Y \rangle) + \frac{\partial F}{\partial X}(X - \langle X \rangle) + \frac{\partial F}{\partial Y}(Y - \langle Y \rangle)$, then,

$$K_{FF} = \langle F - \langle F \rangle \rangle \langle F - \langle F \rangle \rangle^T, \tag{A9}$$

and expanding using Equations (A5), (A7), and (A8),

$$\begin{aligned}
K_{FF} &= \left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) + \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) + \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle^T \\
&= \left[\left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) \right\rangle + \left\langle \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \right] \left[\left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) \right\rangle + \left\langle \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \right]^T \\
&= \left[\left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) \right\rangle \right] \left[\left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) \right\rangle \right]^T + \left[\left\langle \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \right] \left[\left\langle \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \right]^T \\
&\quad + \left[\left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) \right\rangle \right] \left[\left\langle \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \right]^T + \left[\left\langle \tfrac{\partial F}{\partial Y}(Y - \langle Y \rangle) \right\rangle \right] \left[\left\langle \tfrac{\partial F}{\partial X}(X - \langle X \rangle) \right\rangle \right]^T \\
&= \left(\tfrac{\partial F}{\partial X} \right)^2 K_{XX} + \left(\tfrac{\partial F}{\partial Y} \right)^2 K_{YY} + \left(\tfrac{\partial F}{\partial X} \right) \left(\tfrac{\partial F}{\partial Y} \right) K_{XY} + \left(\tfrac{\partial F}{\partial X} \right) \left(\tfrac{\partial F}{\partial Y} \right) K_{YX} \#
\end{aligned} \tag{A10}$$

Substituting $E_n^2 = \left(\frac{\partial F}{\partial n} \right)^2 \sigma_n^2$ from Equation (A6), and noting the decomposition in Equation (A2),

$$K_{FF} = E_x^2 R_{XX} + E_y^2 R_{YY} + E_x E_y R_{XY} + E_x E_y R_{YX} \tag{A11}$$

The construction of R_{mod} in Equation (7) ignores the cross-correlation between component terms, i.e., we assert $R_{XY} = R_{YX} = 0$. This is for two main reasons:

1. The terms K_n, as described in Appendices B–E, are not constructed from random variables but rather through analytic specification to emulate the expected correlated behavior. We generally have insufficient knowledge to measure or estimate the cross-correlation *between* error components. Instead, this model simply estimates the cross-correlation of error *within* individual components, which are then added independently.
2. Any residual cross-correlation between error components can be tuned per our tuning parameters.

This formulation assumes that each component term of the error E_n comes from a wide-sense stationary distribution where the error magnitudes are treated as constants. This means that for each component of K_{FF}, and noting Equation (A2),

$$\left(\frac{\partial F}{\partial X}\right)^2 K_{XX} = \left(\frac{\partial F}{\partial X}\right)^2 S_X R_{XX} S_X = \left(\frac{\partial F}{\partial X}\right)^2 \sigma_X^2 R_{XX} = E_x^2 R_{XX} \tag{A12}$$

For our error correlation model R_{mod}, we simply construct R_{XX} analytically for each term, for which we adopt the nomenclature *Ecorr* to emphasize that it is an error correlation function. Therefore, from Equation (5) (reproduced here),

$$R_{mod}(i,j) = \frac{1}{\mathcal{N}} \sum_n K_n(i,j),$$

where $K_n(i,j) = E_n^2 \cdot Ecorr_n(i,j)$.

For this model, F is drawn from Equation (4) (reproduced here):

$$\sigma^o = \frac{P^g 4\pi L^{atm} L^Z G^Z}{P^Z \zeta^E G^R R^{tot} \bar{A}},$$

where we use the parameters with the largest error magnitude, $\sigma^o = F(P^g, P^Z, G^Z, G^R, \zeta^E)$. The relevant partial derivatives $\frac{\partial F}{\partial N}$ from Equation (A10) become the following:

$$\left|\frac{\partial F}{\partial P^g}\right| = \frac{4\pi L^{atm} L^Z G^Z}{P^Z \zeta^E G^R R^{tot} \bar{A}} \tag{A13a}$$

$$\left|\frac{\partial F}{\partial P^z}\right| = \frac{P^g 4\pi L^{atm} L^Z G^Z}{(P^Z)^2 \zeta^E G^R R^{tot} \bar{A}} \tag{A13b}$$

$$\left|\frac{\partial F}{\partial G^z}\right| = \frac{P^g 4\pi L^{atm} L^Z}{P^Z \zeta^E G^R R^{tot} \bar{A}} \tag{A13c}$$

$$\left|\frac{\partial F}{\partial G^R}\right| = \frac{P^g 4\pi L^{atm} L^Z G^Z}{P^Z \zeta^E (G^R)^2 R^{tot} \bar{A}} \tag{A13d}$$

$$\left|\frac{\partial F}{\partial \zeta^E}\right| = \frac{P^g 4\pi L^{atm} L^Z G^Z}{P^Z (\zeta^E)^2 G^R R^{tot} \bar{A}} \tag{A13e}$$

To estimate E_n, as shown in Table 1, these partial derivatives are evaluated at the 1-sigma value for a reference 10 m/s wind speed using Equation (A6). The total error model becomes the following:

$$R_{mod}(i,j) = \frac{1}{\mathcal{N}} \left[K_{p^g}(i,j) + K_P^z(i,j) + K_{G^Z}(i,j) + K_{G^R}(i,j) + K_{\zeta^E}(i,j) \right], \tag{A14}$$

where K_{p^g} is described in Appendix B, K_P^z is described in Appendix C, K_{G^Z} and K_{G^R} are described in Appendix D, and K_{ζ^E} is described in Appendix E. $\mathcal{N}$ is a normalization constant that forces R_{mod} to behave like a correlation such that $-1 \leq R_{mod} \leq 1$ and can be thought of as an estimate for the rolled-up variance. Because this model has tunable parameters, it is not necessarily representative of the true variance of σ^o but rather of the modeled variance from our bottoms-up model construction.

Appendix B

The term P^g represents calibrated received power from GPS signals reflected from the Earth's surface. P^g is calibrated both from pre-launch characterizations as well as on an on-orbit blackbody at a known temperature, which, as of 2022, takes a reading every

10 min to re-compute gain that may have changed due to the dynamic thermal environment in orbit [35].

For CYGNSS, P^g is computed by the Level 1A algorithm:

$$P^g = \frac{C - C_N}{G}, \tag{A15}$$

where C is the raw counts at delay–Doppler bins, where a scattered signal is present and is the measured parameter for an observation, C_N is an estimate of the background noise without any scattered signal, and G is the receiver gain in units of counts/watt. In current processing, C_N is an average of the raw counts at in the delay–Doppler map at coordinates at lower delay values than the specular point [25]. Further, gain G is measured in orbit by performing readings from an onboard blackbody at a known temperature and calibrated via the following:

$$G = \frac{C_B}{P_B + P_r}, \tag{A16}$$

where C_B is the counts measured while looking at the blackbody, P_B is the power in watts from the blackbody as estimated from a thermocouple located near the receiver's low noise amplifier, and P_r is the receiver noise power in watts estimated from a pre-launch parameterization. Equations (A15) and (A16) can be combined:

$$P^g = \frac{(C - C_N)}{C_B}(P_B + P_r), \tag{A17}$$

The magnitude of the errors for the components calculating P^g is displayed in Table A1 and has a similar calculation as before.

Table A1. The magnitudes of 1-sigma errors for each term in Equation (A17). The shaded rows indicate that the absolute magnitude is negligible compared to the dominant error terms and is neglected in the construction of the error model in this work.

Error Term	Error Magnitude [dB]
$E(C_N)$	0.14 [36]
$E(P_r)$	0.14 [36]
$E(C)$	0.10 [36]
$E(C_B)$	0.07 [35]
$E(P_B)$	~0.04 [36]

C_B, P_B, and P_r all vary with temperature, and the instrument gain G will fluctuate as the satellite enters different thermal conditions in orbit. The most significant errors will occur just as the satellite crosses the terminator. At that point, CYGNSS will go from a nearly steady-state thermal environment, such as approximately half an orbit of illumination or eclipse, and then quickly enter the opposite state. The fraction of orbit spent illuminated is determined by the orbit beta angle, which varies on scales of weeks to months.

The dominant error term in the Level 1A algorithm is C_N, which also varies with temperature. The calibration sequence is designed to correct for this, and we assume that the errors vary slowly with the timescale of interest, which is defined to be on orders of seconds to minutes. Therefore, all errors from C_N are assumed to be 100% correlated in time within a given track of CYGNSS observations, i.e., when a series of samples adjacent in space and time share a GPS transmitter and a CYGNSS receiver. C_N is also very sensitive to radio-frequency interference (RFI), which will present as non-physical signals above the specular point in a delay–Doppler map. We do not aim to model the complex phenomenologies of RFI in this work and assume there are no correlated error structures from RFI.

Errors in P_r occur because of a variety of reasons. The low noise amplifiers were all characterized on the ground prior to launch to establish the relationship of the noise figure with respect to temperature. The values of this relationship were stored in a look-up table (LUT) for processing science data. However, as the amplifiers age, the noise floor characteristics may have evolved, producing errors in this mapping. Further, the thermal environment of the thermocouple may not be exactly the same as experienced by the amplifier itself. For the purposes of this model, we assume all errors due to incomplete or erroneous knowledge of the true receiver noise power are 100% correlated with each other for a given track, as we assume that the errors evolve slowly compared to the timescales of interest.

Therefore, the error correlation terms for C_N and P_r for any arbitrary CYGNSS samples x_i and x_j are the following:

$$Ecorr_{C_N}(x_i, x_j) = Ecorr_{Pr}(x_i, x_j) = \begin{cases} 1, & conditions\ met \\ 0, & else \end{cases} \tag{A18}$$

The conditions that must be met are the following: x_i and x_j must share a CYGNSS receiver and a GPS transmitter and be observed within 10 min of each other.

C is the measured parameter, the raw counts of power from a science observation near the region of the specular point. The analog-to-digital processing chain is the primary source of errors, such as quantization errors and non-common-mode interference. We assume that these error terms are 100% uncorrelated with each other; that is, for every sample, it can be treated as white noise. The error correlation term for C is then the following:

$$Ecorr_C(x_i, x_j) = \begin{cases} 1, & x_i = x_j \\ 0, & x_i \neq x_j \end{cases} \tag{A19}$$

The error term for counts measured during a blackbody sample C_B is a source of analytically defined correlated error. Every 10 min (earlier in the mission, every 1 min), the receiver is switched from the nadir science antenna to look at the onboard blackbody source for a period of 4–6 s. Science observation processing linearly interpolates the counts between the nearest blackbody looks. When errors are made in estimating C_B, those errors are correlated linearly with all adjacent samples due to this interpolation. Correlation due to linear interpolation has an analytical form. Assuming a blackbody look happens at timesteps 0 and n, then, the correlation between any two samples x_i and x_j at arbitrary timesteps i and j where $0 \leq i < j \leq n$ is the following:

$$Ecorr_{C_B}(x_i, x_j) = \frac{1}{\sigma_i \sigma_j}\left(\frac{(n-i)(n-j)}{n^2} + \frac{ij}{n^2}\right), \tag{A20}$$

where $\sigma_i = \sqrt{\frac{(n-i)^2+i^2}{n^2}}$ and $\sigma_j = \sqrt{\frac{(n-j)^2+j^2}{n^2}}$. The actual values of the sampled blackbodies do not matter, as the correlated error is simply a function of how far the samples are from the blackbody looks in time.

Errors in P_B are due to misestimations of the blackbody's true noise power, which may be because the thermocouple is measuring incorrectly. We assume that the errors of this nature not only are slowly varying compared to the timescales of interest but, because of the marginal absolute magnitude, factor a negligible and unmeasurable amount in the overall correlated error structure. As such, the model ignores this term.

The rolled-up correlated error model from the sources in P^g between any arbitrary samples i and j can be expressed as follows:

$$K_{P^g}(i,j) = \begin{bmatrix} \alpha \cdot E(C)^2 \cdot Ecorr_C(i,j) + \\ \beta \cdot E(C_N)^2 \cdot Ecorr_{C_N}(i,j) + \\ \beta \cdot E(P_r)^2 \cdot Ecorr_{P_r}(i,j) + \\ E(C_B)^2 \cdot Ecorr_{C_B}(i,j) \end{bmatrix} \tag{A21}$$

Because this model estimates the correlated error in each term that calculates P^g, this contribution to the overall correlated error is not multiplied by the rolled-up error magnitude $E(P^g)$. As such, we do not normalize this construction, as it will be normalized when combined with the other constituent terms in R_{mod}. R_{Pg} contains two tuning parameters: α is used to size uncorrelated white-noise error, and β is used to size the magnitude of totally correlated errors.

Appendix C

The zenith receiver on CYGNSS works much the same way as the nadir science receiver. However, the zenith receiver does not have an onboard calibration system, and data are processed from the pre-flight characterizations of the electronics. The counts of the receiver are converted to watts via a quadratic regression. Because the satellite is subject to the same thermal dynamics as the nadir receiver, one can expect that the zenith power estimate contains errors due to thermally driven gain variations. We model the $Ecorr_{PZ}$ similarly to the nadir receiver but with some important distinctions. Because the zenith receiver operates without an onboard blackbody to calibrate against, a number of simplifying assumptions are made. Errors are broken up into just two components: correlated $Ecorr_{P_1^Z}$ and uncorrelated $Ecorr_{P_2^Z}$. We further assume that correlated error due to the absence of an onboard blackbody are outside the timescale of interest. The correlated error term $Ecorr_{P_1^Z}$ is assumed to be totally correlated, provided that the matching conditions are met:

$$Ecorr_{P_1^Z}(x_i, x_j) = \begin{cases} 1, & conditions\ met \\ 0, & else \end{cases} \tag{A22}$$

The conditions for $Ecorr_{P_1^Z}$ are that samples x_i and x_j are made with the same CYGNSS observatory and within 10 min of each other. In addition, an uncorrelated component is allowed:

$$Ecorr_{P_2^Z}(x_i, x_j) = \begin{cases} 1, & x_i = x_j \\ 0, & x_i \neq x_j \end{cases} \tag{A23}$$

The rolled-up correlated error model from the sources in P^Z between any arbitrary samples i and j can be expressed as follows:

$$K_{P^Z}(i,j) = \left[\beta \cdot E\left(P_1^Z\right)^2 \cdot Ecorr_{P_1^Z}(i,j)\right] + \left[\alpha \cdot E\left(P_2^Z\right)^2 \cdot Ecorr_{P_2^Z}(i,j)\right], \tag{A24}$$

with the same tuning parameters α and β, as in Appendix A. Note that we assume $E\left(P_1^Z\right) = 0.18$ dB, as suggested in [33]. $E\left(P_2^Z\right)$ is estimated directly from a 24 h of CYGNSS data as approximately 1% of the magnitude of the signal P^Z; therefore, this model assumes $E\left(P_2^Z\right) = 0.04$ dB.

Appendix D

The CYGNSS observatory has three antennas: one zenith antenna that is used for direct GPS-to-CYGNSS signal tracking, as well as two nadir science antennas that are used to capture the scattered signal from Earth's surface. Each of the eight spacecraft had all

three antennas characterized pre-launch, and values were stored in a look-up table for science processing.

Errors in the antenna gain pattern can arise for a variety of reasons. First, the measurement equipment on the ground is essentially a receiver but in controlled conditions. This means that while systematic and correlated errors are likely well-constrained, uncorrelated speckle-type error can still occur. To produce realistic antenna gain patterns, the results of the ground characterization were smoothed with various filters and techniques.

Another source of error is the fact that CYGNSS antennas were not characterized while integrated with the spacecraft. This was a cost-saving measure decided by the mission management team. However, the electromagnetic properties of the antenna couple in some fashion with the spacecraft bus, and that will inevitably change the gain patterns.

Initial analysis of CYGNSS data shortly after launch showed significant retrieval performance dependence on the observation azimuthal angle with respect to the CYGNSS body frame, which was later hypothesized to originate in errors in the CYGNSS antenna patterns. To compensate for this deviation from measured patterns, the CYGNSS antenna patterns have been updated at several instances over the mission life via empirical calibration. The nadir antenna patterns are updated by comparing a climatology of CYGNSS measurements of σ^o (>2 years) with model-generate σ^o_{mod} and plotting a scaling factor in the antenna reference coordinate system. σ^o_{mod} is generated by using modeled reanalysis winds to generate mean-squared slope with the L-band spectrum extension model, as described in [38]. However, during the generation of these updated patterns, a number of smoothing filters are applied.

This prompts a discussion of a conjecture used extensively for this section:

Conjecture A1. *Uncorrelated errors can become correlated by post-processing with averaging and filters.*

This insight drives much of this section's analysis. Smoothing and filtering will necessarily impose a correlation in error between previously uncorrelated errors. For white noise, that implies that the choice of filter will add color and structure to the noise.

In particular, a handy lemma allows us to demonstrate that for white noise, the information required to capture correlated error structure is the filtering kernel itself. We will explore this behavior for a one-dimensional case, but it is generalizable to higher dimensions in our application, as the two-dimensional filters used for antenna smoothing are separable by construction.

Lemma A1. *For a filtered signal $F(t) = K(t) * D(t)$, where $K(t)$ is a filtering kernel and $D(t)$ is an arbitrary data signal, the autocorrelation $F \star F$ is the convolution of the autocorrelated kernel $K \star K$ and the autocorrelated signal $D \star D$.*

Proof of Lemma A1. Assume convolution and cross-correlation have the standard definitions for two real-valued timeseries $K(t)$ and $D(t)$, that is,

$$K(t) * D(t)[n] = \sum_{t=-\infty}^{\infty} K(t)D(n-t) = \sum_{t=-\infty}^{\infty} K(n-t)D(t) \tag{A25}$$

and

$$K(t) \star D(t)[n] = \sum_{t=-\infty}^{\infty} K(t)D(n+t) = \sum_{t=-\infty}^{\infty} K(t-n)D(t), \tag{A26}$$

where $*$ is the convolution operator, $\star$ is the cross-correlation operator, and n is the lag argument. Observe that convolution operations are commutative and, further, that the cross-correlation can be written as a convolution by exploiting its symmetry:

$$K(t) \star D(t)[n] = K(-t) * D(t)[n] \tag{A27}$$

Therefore, to evaluate the correlated error imposed by kernel $K(t)$,

$$\begin{aligned} F(t) \star F(t) &= \{K(t) * D(t)\} \star \{K(t) * D(t)\} \\ &= K(-t) * D(-t) * K(t) * D(t) \\ &= \{K(-t) * K(t)\} * \{D(-t) * D(t)\} \\ &= \{K(t) \star K(t)\} * \{D(t) \star D(t)\} \end{aligned}$$

If the arbitrary signal $D(t)$ happens to be white noise, it is completely uncorrelated, and its autocorrelation collapses to a Dirac delta function centered at $t = 0$. Therefore, the entire structure of the correlated error is from the filter itself:

$$F(t) \star F(t) = K(t) \star K(t) \tag{A28}$$

□

For the purposes of this work's error model, we have no knowledge of the potential correlated structure in the actual errors in the gain pattern. The nadir antenna patterns are updated after applying a 6-degree boxcar averaging filter in both the azimuthal and elevation in the spacecraft coordinate frame and then an additional 10-degree two-dimensional smoothing window. We assume that zenith antenna patterns use a similar post-processing technique during their generation.

These filtering kernels act like low-pass filters. All correlated structure on scales ~5 degrees and smaller and uncorrelated error will be strongly influenced by the filtering process, and the correlated structure can be estimated from the Filter Lemma. This model assumes that there is no residual larger-scale structure in correlated error in the antenna gain patterns.

The generated filter kernel, which applies to both the nadir and zenith antenna patterns, can be shown in Figure A1. The correlated error is a function of how close any two observations are with respect to the relevant antenna gain pattern coordinates.

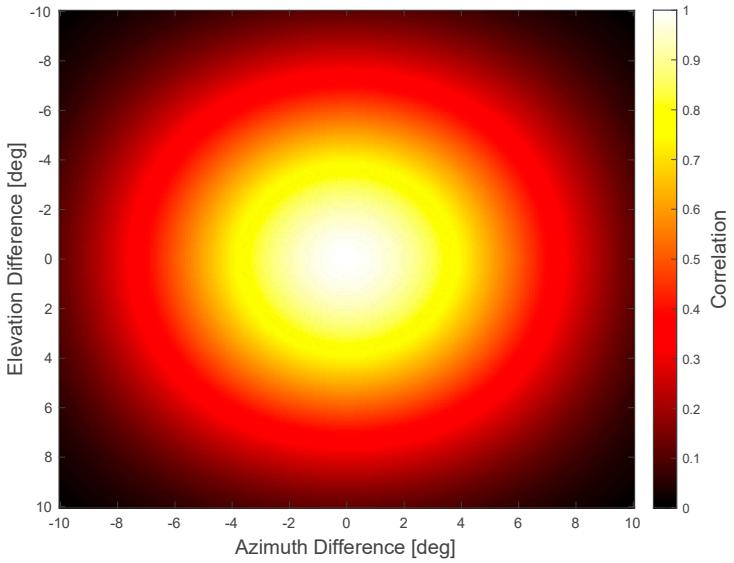


Figure A1. The filtering kernel used to smooth nadir and zenith antenna gain patterns. This kernel imposes correlated error structure onto the antenna gain patterns. The coordinate system should be read as distance in the relevant antenna reference frame. Therefore, if two observations are nearby in the antenna pattern, they will have strongly correlated errors. However, if two observations are far apart in the pattern, the correlated structure decays.

For any arbitrary samples i and j, we compute the gain pattern coordinates (θ_i, ϕ_i) and (θ_j, ϕ_j) in the relevant antenna reference frame. To retrieve how correlated the error is, we compute the distance between the two observations in the reference frame:

$$\Delta\theta = \theta_i - \theta_j$$

$$\Delta\phi = \phi_i - \phi_j$$

The error correlation function is computed via a LUT of the filter kernel K:

$$Ecorr_{G^z} = K(\Delta\theta_Z, \Delta\phi_Z) \tag{A29}$$

$$Ecorr_{G^R} = K(\Delta\theta_R, \Delta\phi_R) \tag{A30}$$

This error correlation holds if the samples i and j share the same antenna and are on the same spacecraft. If they are on separate antennas or spacecrafts, the correlated error is zero. The rolled-up correlated errors for the gain patterns can be expressed as follows:

$$K_{G^Z}(i,j) = \gamma \cdot E\left(G^Z\right)^2 \cdot \frac{Ecorr_{G^Z}(i,j)}{\delta \cdot (1 + \Delta\phi^2 + \Delta\theta^2)}, \tag{A31}$$

$$K_{G^R}(i,j) = \gamma \cdot E\left(G^R\right)^2 \cdot \frac{Ecorr_{G^R}(i,j)}{\delta \cdot (1 + \Delta\phi^2 + \Delta\theta^2)}, \tag{A32}$$

where two new tuning parameters have been introduced. γ is used to the tune the overall magnitude of the correlated error from these components, and δ is used to scale the decorrelation roll-off rate as the samples spread in antenna coordinates.

Appendix E

Errors in the zenith–specular ratio ζ are defined as a function of specular incidence angle θ_{inc}, which is a function of the geometry of a given GPS transmitter, a CYGNSS receiver at any given sample time.

ζ is used to estimate GPS EIRP and is derived via the following, as described in [33]:

$$\zeta \equiv \frac{EIRP_z}{EIRP_S} = \frac{EIRP(t, \theta_z, \phi_z)}{EIRP(t, \theta_S, \phi_S)} = \frac{P^T(t)G^T(\theta_z, \phi_z)}{P^T(t)G^T(\theta_S, \phi_S)} = \frac{G^T(\theta_z, \phi_z)}{G^T(\theta_S, \phi_S)}, \tag{A33}$$

where the angles are defined in the GPS reference frame. For specular geometries, the azimuthal angles in the zenith direction are nearly identical to the specular direction, so $\phi_z = \phi_S = \phi$. In addition, the elevation angles in the GPS antenna reference frame θ_z and θ_S can be estimated from the angle of incidence of specular reflection from Earth θ_{inc}:

$$\theta_z \cong \theta_z(\theta_{inc}) \tag{A34a}$$

$$\theta_S \cong \theta_S(\theta_{inc}) \tag{A34b}$$

As a result, ζ can be expressed as a function of the specular incidence angle and azimuthal angle in the GPS antenna reference frame. While GPS antenna patterns are known to exhibit azimuthal dependence, this variation is less significant than the elevation angle, and CYGNSS uses the azimuthal average for its EIRP estimate:

$$\zeta(\theta_{inc}) \equiv \frac{1}{2\pi}\int_0^{2\pi} \frac{G_T(\theta_z(\theta_{inc}), \phi)}{G_T(\theta_S(\theta_{inc}), \phi)} d\phi \tag{A35}$$

The estimated correlated error in ζ, however, comes with two steps of this processing. First is the mapping of Earth scattering incidence angle θ_{inc} to GPS antenna elevation angles θ_z and θ_S in Equations (A34a) and (A34b). This particular mapping is coarse, as even the high-fidelity-derived GPS antenna maps are plotted to 0.5-degree increments. Because

the dynamic range of θ_z only extends to about 15 degrees, that only leaves ~30 data points to map the full dynamic range of scattering incidence angles.

The second aspect has to do with the way in which ζ is processed and generated and invokes the same logic as in the Filter Lemma. For every GPS satellite, a ζ LUT is generated as a function of observation incidence angle θ_{inc}. To minimize discontinuities, a fourth-order power series is fit. We argue that this smoothing is the predominant source of correlated error structure. An example of this is demonstrated in Figure A2.

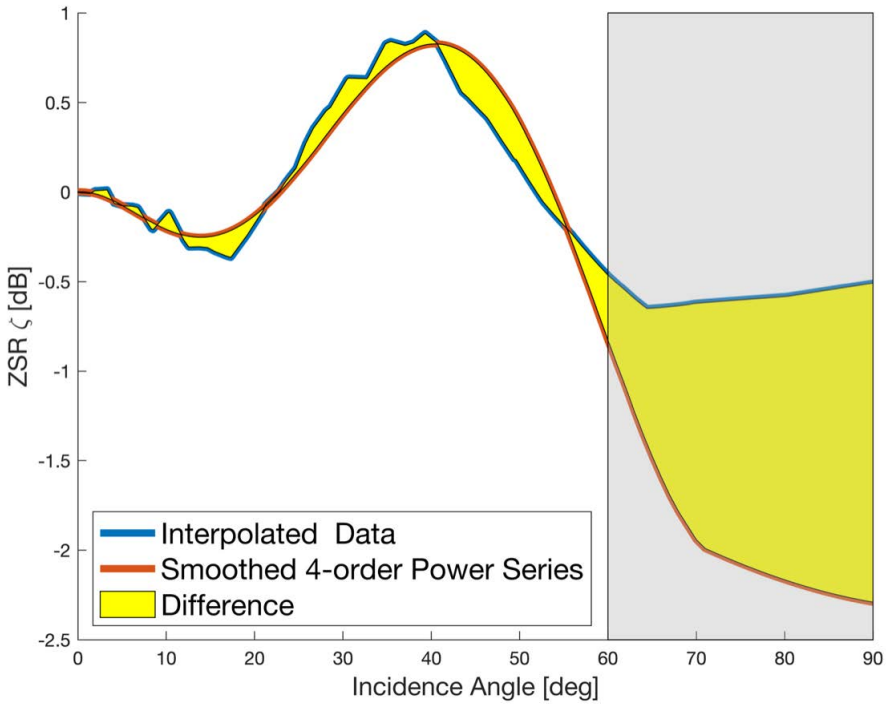


Figure A2. This figure illustrates a calculated zenith–specular ratio ζ as a function of observation incidence angle θ_{inc} at a fixed GPS antenna azimuth for GPS PRN 2. The blue trace is interpolated from raw observations over a two-year period at each of the elevation gridpoints in the GPS antenna pattern for a single azimuthal cut of PRN 2. The red trace is a generated smoothed zenith–specular ratio ζ that would be similar to the ones used in the operational LUTs using a 4th-order power series fit. Note that at large incidence angles, i.e., grazing observations, there is a great deal of uncertainty in ζ because there are few valid observations in those regions. In practice, only data at incidence angles < 60 degrees constrain error in ζ.

While linear interpolation itself imparts some degree of error structure, we believe it is the most representative way to express 'raw' data in a continuous series for the purposes of exploring correlated error due to the power series smoothing. For each GPS PRN, we calculate the difference between these estimates:

$$\Delta\zeta_\phi(\theta_{inc}) = \left(\frac{G_T(\theta_z(\theta_{inc}),\phi)}{G_T(\theta_S(\theta_{inc}),\phi)}\right)_{interp} - \left(\frac{G_T(\theta_z(\theta_{inc}),\phi)}{G_T(\theta_S(\theta_{inc}),\phi)}\right)_{smooth}$$

Then, the error correlation is simply the following:

$$Ecorr_\zeta(\theta_i,\theta_j) = corr\left(\Delta\zeta_\phi(\theta_i),\ \Delta\zeta_\phi(\theta_j)\right), \tag{A36}$$

where θ_i is the incidence angle of the observation at sample i, and θ_j is the incidence angle of the observation at sample j. In practice, the correlation is computed by using each azimuthal cut as an instance and building a LUT of correlation as a function of incidence

angles for samples i and j. An example of this LUT for GPS PRN 2 is shown in Figure A3. The rolled-up correlated error for ζ can then be expressed as follows:

$$R_{\zeta}(i,j) = \left(\beta \cdot E(\zeta)^2 \cdot Ecorr_{\zeta}(i,j)\right)^{\frac{1}{2}}, \tag{A37}$$

with the same tuning parameter β as introduced in Appendix B.

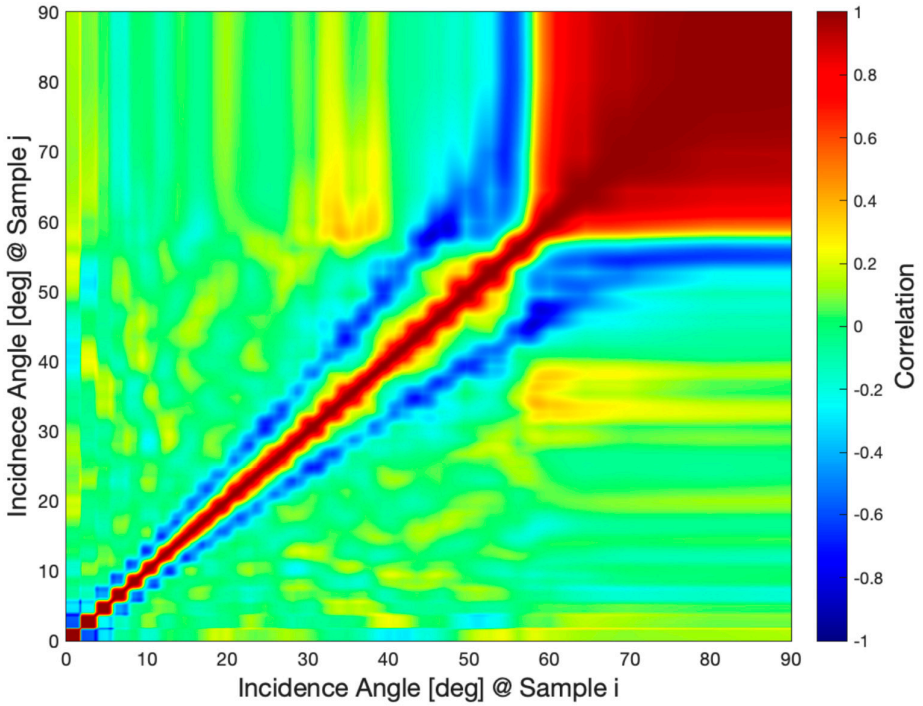


Figure A3. This figure depicts an error correlation matrix derived from Equation (A36) for GPS PRN 2, which is used to produce $Ecorr_{\zeta}$. The matrix is a function of the observation incidence angle for sample points i and j. Note that the mapping from coordinates in GPS elevation angle θ_z and θ_s to Earth scattering incidence angle θ_{inc} is coarse and produces the checkerboard-like pattern near the diagonal. A single error in the measurement of G_T in the GPS antenna pattern is will highly correlate within a range of incidence angles in θ_{inc}, as mapped. At high incidence angles, errors are strongly correlated as the power series fit is likely to be wrong in the same direction.

References

1. Ruf, C.S.; Chew, C.; Lang, T.; Morris, M.G.; Nave, K.; Ridley, A.; Balasubramaniam, R. A New Paradigm in Earth Environmental Monitoring with the CYGNSS Small Satellite Constellation. *Sci. Rep.* **2018**, *8*, 8782. [CrossRef] [PubMed]
2. Lorenc, A.C.; Ballard, S.P.; Bell, R.S.; Ingleby, N.B.; Andrews, P.L.F.; Barker, D.M.; Bray, J.R.; Clayton, A.M.; Dalby, T.; Li, D.; et al. The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **2000**, *126*, 2991–3012. [CrossRef]
3. Bannister, R.N. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Q. J. R. Meteorol. Soc.* **2008**, *134*, 1951–1970. [CrossRef]
4. Bannister, R.N. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **2008**, *134*, 1971–1996. [CrossRef]
5. Morss, R.E.; Emanuel, K.A. Influence of added observations on analysis and forecast errors: Results from idealized systems. *Q. J. R. Meteorol. Soc.* **2002**, *128*, 285–321. [CrossRef]
6. Stewart, L.M.; Dance, S.L.; Nichols, N.K. Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. *Tellus A Dyn. Meteorol. Oceanogr.* **2013**, *65*, 19546. [CrossRef]
7. Fowler, A.M.; Dance, S.L.; Waller, J.A. On the interaction of observation and prior error correlations in data assimilation. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 48–62. [CrossRef]
8. Liu, Z.Q.; Rabier, F. The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc.* **2002**, *128*, 1367–1386. [CrossRef]

9. Stewart, L.M.; Dance, S.L.; Nichols, N.K.; Eyre, J.R.; Cameron, J. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1236–1244. [CrossRef]
10. Bormann, N.; Bonavita, M.; Dragani, R.; Eresmaa, R.; Matricardi, M.; McNally, A. Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 1767–1780. [CrossRef]
11. Bormann, N.; Bauer, P. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 1036–1050. [CrossRef]
12. Smith, D.; Hunt, S.E.; Etxaluze, M.; Peters, D.; Nightingale, T.; Mittaz, J.; Woolliams, E.R.; Polehampton, E. Traceability of the Sentinel-3 SLSTR Level-1 Infrared Radiometric Processing. *Remote Sens.* **2021**, *13*, 374. [CrossRef]
13. Yang, J.X.; You, Y.; Blackwell, W.; Da, C.; Kalnay, E.; Grassotti, C.; Liu, Q.; Ferraro, R.; Meng, H.; Zou, C.-Z.; et al. SatERR: A Community Error Inventory for Satellite Microwave Observation Error Representation and Uncertainty Quantification. *Bull. Am. Meteorol. Soc.* **2023**, *105*, E1–E20. [CrossRef]
14. Simonin, D.; Waller, J.A.; Ballard, S.P.; Dance, S.L.; Nichols, N.K. A pragmatic strategy for implementing spatially correlated observation errors in an operational system: An application to Doppler radial winds. *Q. J. R. Meteorol. Soc.* **2019**, *145*, 2772–2790. [CrossRef]
15. Waller, J.A.; Dance, S.L.; Nichols, N.K. On diagnosing observation-error statistics with local ensemble data assimilation. *Q. J. R. Meteorol. Soc.* **2017**, *143*, 2677–2686. [CrossRef]
16. Bauer, P.; Buizza, R.; Cardinali, C.; Noël Thépaut, J. Impact of singular-vector-based satellite data thinning on NWP. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 286–302. [CrossRef]
17. Gao, Y.; Xiao, H.; Jiang, D.; Wan, Q.; Chan, P.W.; Hon, K.K.; Deng, G. Impacts of thinning aircraft observations on data assimilation and its prediction during typhoon nida (2016). *Atmosphere* **2019**, *10*, 754. [CrossRef]
18. Hoffman, R.N. The Effect of Thinning and Superobservations in a Simple One-Dimensional Data Analysis with Mischaracterized Error. *Mon. Weather Rev.* **2018**, *146*, 1181–1195. [CrossRef]
19. Rainwater, S.; Bishop, C.H.; Campbell, W.F. The benefits of correlated observation errors for small scales. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 3439–3445. [CrossRef]
20. Stewart, L.M.; Dance, S.L.; Nichols, N.K. Correlated observation errors in data assimilation. *Numer. Methods Fluids* **2008**, *56*, 1521–1527. [CrossRef]
21. Daley, R. The Effect of Serially Correlated Observation and Model Error on Atmospheric Data Assimilation. *Mon. Weather Rev.* **1992**, *120*, 164–177. [CrossRef]
22. Waller, J.A.; Dance, S.L.; Nichols, N.K. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quart. J. R. Meteoro. Soc.* **2016**, *142*, 418–431. [CrossRef]
23. Dee, D.P.; Da Silva, A.M. Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology. *Mon. Weather Rev.* **1999**, *127*, 1822–1834. [CrossRef]
24. Desroziers, G.; Berre, L.; Chapnik, B.; Poli, P. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 3385–3396. [CrossRef]
25. Ruf, C.; McKague, D.; Posselt, D.J.; Gleason, S.; Clarizia, M.P.; Zavorotny, V.U.; Butler, T.; Redfern, J.; Wells, W.; Morris, M.; et al. *CYGNSS Handbook*, 2nd ed.; Michigan Publishing Services: Ann Arbor, MI, USA, 2022. ISBN 978-1-60785-787-7.
26. Clarizia, M.P.; Ruf, C.S. Wind Speed Retrieval Algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) Mission. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4419–4432. [CrossRef]
27. Ruf, C.S.; Balasubramaniam, R. Development of the CYGNSS Geophysical Model Function for Wind Speed. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 66–77. [CrossRef]
28. Gleason, S.; Ruf, C.S.; Orbrien, A.J.; McKague, D.S. The CYGNSS Level 1 Calibration Algorithm and Error Analysis Based on On-Orbit Measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 37–49. [CrossRef]
29. Wang, T.; Ruf, C.; McKague, D.; Russel, A.; O'Brien, A.; Gleason, S. The Important Role of Antenna Pattern Characterization in the Absolute Calibration of GNSS-R Measurements. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 144–146.
30. Wang, T.; Ruf, C.; Gleason, S.; McKague, D.; O'Brien, A.; Block, B. Monitoring GPS Eirp for Cygnss Level 1 Calibration. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 17 February 2020; pp. 6293–6296.
31. Steigenberger, P.; Thölert, S.; Montenbruck, O. Flex power on GPS Block IIR-M and IIF. *GPS Solut.* **2019**, *23*, 8. [CrossRef]
32. Marquis, W.A.; Reigh, D.L. The GPS Block IIR and IIR-M Broadcast L-band Antenna Panel: Its Pattern and Performance: GPS Block IIR and IIR-M L-band Antenna Panel Pattern. *J. Inst. Navig.* **2015**, *62*, 329–347. [CrossRef]
33. Wang, T.; Ruf, C.S.; Gleason, S.; O'Brien, A.J.; McKague, D.S.; Block, B.P.; Russel, A. Dynamic Calibration of GPS Effective Isotropic Radiated Power for GNSS-Reflectometry Earth Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–12. [CrossRef]
34. Gleason, S. *Level 1B DDM Calibration Algorithm Theoritical Basis Document (Rev. 3)*; University of Michigan: Ann Arbor, MI, USA, 2020.
35. Powell, C.E.; Ruf, C.S.; Russel, A. An Improved Blackbody Calibration Cadence for CYGNSS. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–7. [CrossRef]
36. Gleason, S. *Level 1A DDM Calibration Algorithm Theoretical Basis Document (Rev. 2)*; University of Michigan: Ann Arbor, MI, USA, 2018.
37. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]

38. Wang, T.; Zavorotny, V.U.; Johnson, J.; Yi, Y.; Ruf, C. Integration of Cygnss Wind and Wave Observations with the Wavewatch III Numerical Model. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 8350–8353.
39. Berg, W.; Brown, S.T.; Lim, B.H.; Reising, S.C.; Goncharenko, Y.; Kummerow, C.D.; Gaier, T.C.; Padmanabhan, S. Calibration and Validation of the TEMPEST-D CubeSat Radiometer. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4904–4914. [CrossRef]
40. Kroodsma, R.A.; McKague, D.S.; Ruf, C.S. Inter-Calibration of Microwave Radiometers Using the Vicarious Cold Calibration Double Difference Method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1006–1013. [CrossRef]
41. Zec, J.; Jones, W.; Alsabah, R.; Al-Sabbagh, A. RapidScat Cross-Calibration Using the Double Difference Technique. *Remote Sens.* **2017**, *9*, 1160. [CrossRef]
42. Powell, C.E.; Ruf, C.S.; Gleason, S.; Rafkin, S.C.R. Sampled Together: Assessing the Value of Simultaneous Collocated Measurements for Optimal Satellite Configurations. *Bull. Am. Meteorol. Soc.* **2024**, *105*, E285–E296. [CrossRef]
43. *CYGNSS CYGNSS Level 1 Science Data Record Version 3.1*; NASA: Washington, DC, USA, 2021. [CrossRef]
44. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Razum, I.; et al. ERA5 Hourly Data on Single Levels from 1940 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available online: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview (accessed on 12 March 2022).
45. Taylor, J.R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed.; University Science Books: Sausalito, CA, USA, 1997; ISBN 978-0-935702-42-2.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.