Data Science in Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/udss20

Comparison of CYGNSS and Jason-3 Wind Speed Measurements via Gaussian Processes

William Bekerman & Joseph Guinness

To cite this article: William Bekerman & Joseph Guinness (2023) Comparison of CYGNSS and Jason-3 Wind Speed Measurements via Gaussian Processes, Data Science in Science, 2:1, 2194349, DOI: 10.1080/26941899.2023.2194349

To link to this article: https://doi.org/10.1080/26941899.2023.2194349

Published online: 21 Apr 2023.

DATA SCIENCE IN SCIENCE
2023, VOL. 2, NO. 1, 2194349
https://doi.org/10.1080/26941899.2023.2194349

Comparison of CYGNSS and Jason-3 Wind Speed Measurements via Gaussian Processes

William Bekerman[a] and Joseph Guinness[b]

[a]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [b]Department of Statistics and Data Science, Cornell University, Ithaca, New York, USA

ABSTRACT

Wind is a critical component of the Earth system and has unmistakable impacts on everyday life. The CYGNSS satellite mission improves observational coverage of ocean winds *via* a fleet of eight micro-satellites that use reflected GNSS signals to infer surface wind speed. We present analyses characterizing variability in wind speed measurements among the eight CYGNSS satellites and between antennas, using a Gaussian process model that leverages comparisons between CYGNSS and Jason-3 during a one-year period from September 2019 to September 2020. The CYGNSS sensors exhibit a range of biases, mostly between -1.0 m/s and $+0.2$ m/s with respect to Jason-3, indicating that some CYGNSS sensors are biased with respect to one another and with respect to Jason-3. The biases between the starboard and port antennas within a CYGNSS satellite are smaller. Our results are consistent with, yet sharper than, a more traditional paired comparison analysis. We also explore the possibility that the bias depends on wind speed, finding some evidence that CYGNSS satellites have positive biases with respect to Jason-3 at low wind speeds. However, we argue that there are subtle issues associated with estimating wind speed-dependent biases, so additional careful statistical modeling and analysis is warranted.

ARTICLE HISTORY
Received 15 November 2022
Revised 2 February 2023
Accepted 19 March 2023

KEYWORDS
Remote sensing; satellite; measurement bias

1. Introduction

Wind is a crucial component of the atmosphere and climate, having significant implications in numerous areas of daily life, from safety and transportation to industry and science. Recording accurate wind measurements provides critical information for precisely defining weather hazards Adelekan (2000), building skyscrapers, and landing aircraft Smith and Chow (1998). Reliable wind speed observations are also consequential in allowing us to efficiently conduct crop spraying Endalew et al. (2010), monitor the global climate Eichelberger et al. (2008), and avoid the obstruction of essential global shipping routes Rusu et al. (2018).

Scientists and engineers have developed a diverse suite of tools for measuring wind speeds. Weather stations and buoys are often equipped with anemometers to measure wind speed directly. Many earth-observing satellites carry sensors capable of inferring wind speeds. A thorough review of satellite-based methods for measuring wind speeds is provided in Young et al. (2017), which includes radiometers, scatterometers, and altimeters, which are usually attached to low-earth-orbiting satellites. In addition, geostationary satellites are capable of inferring upper-air wind speeds by detecting movements in clouds *via* derived motion winds algorithms Daniels et al. (2019).

The Cyclone Global Navigation Satellite System (CYGNSS) is a fleet of eight micro-satellites that use the scattered signals from existing GNSS satellites to infer wind speeds at the ocean surface (Ruf et al. 2012, 2013). CYGNSS is a relatively new and low-cost system that, due to its ability to distribute its sensing effort over eight satellites, has the advantage of greater spatial-temporal coverage of the oceans relative to a single-satellite system, an important feature for its mission of monitoring tropical cyclones.

Our primary goal is to study the internal variability in wind speed measurements among the eight CYGNSS satellites and across each satellite's starboard and port antennas. As noted in Asharaf et al. (2021), this variability is still under study, and further calibrations are a possibility: "it is more likely the differences in bias, both between antennas and between spacecraft, are caused by residual errors in the engineering calibration, which is performed individually for each spacecraft and antenna. This is an ongoing area of investigation by the CYGNSS project team … " Our secondary goal is to study differences between CYGNSS and Jason-3 wind speed measurements. Jason-3 is a separate satellite that uses reflected signals from its radar altimeter to infer wind speed and other ocean surface parameters.

Several recent articles study statistical properties of CYGNSS wind speed measurements. CYGNSS was

CONTACT William Bekerman bekerman@wharton.upenn.edu Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA, USA

compared with weather model forecast winds and found to be positively biased at low wind speeds and negatively biased at high wind speeds, but no comparisons were made among individual CYGNSS satellites Pascual et al. (2021). In the tropics, a similar pattern of positive bias for low wind speeds and negative bias for high wind speeds, relative to hourly-averaged buoy data, was detected Asharaf et al. (2021). CYGNSS biases with respect to modeled and sensed wind speeds were assessed in the context of building a complex bias correction algorithm Saïd et al. (2022).

The spatial-temporal patterns of low-earth-orbiting satellite observations, coupled with the inherent spatial-temporal variability in wind speeds, present a data-analytic challenge for conducting the desired comparisons in our study. By design, the eight CYGNSS satellites do not measure winds concurrently at the same locations, so it is difficult to draw direct comparisons between observations from most pairs of these satellites. Thus, we rely on repeated approximate crossings between CYGNSS and Jason-3 for indirect comparisons. We believe that Jason-3 is a suitable comparison because, like CYGNSS, it retrieves a snapshot of surface wind speed, as opposed to an hourly average or a model forecast, both of which can be overly smooth. To maximize statistical power, it is important to make judicious use of the observations arising from the limited number of nearby CYGNSS-Jason-3 passes, while avoiding a subjective determination of which passes constitute a close-enough match.

To address these issues, we analyze the data using Gaussian process models. The models contain parameters for capturing biases among the satellites, and they have terms for modeling winds that vary continuously over space and time. Gaussian process models have become an indispensable tool for analyzing and interpolating scattered remote sensing data. Recent examples include the analysis of Argo float data Kuusela and Stein (2018), Orbiting Carbon Observatory-2 data Susiluoto et al. (2020); Katzfuss et al. (2020), surface temperatures Rayner et al. (2020), Microwave Atmospheric Satellite data Ruan et al. (2017), and Jason-3 wind speeds Guinness (2018).

Our models for CYGNSS and Jason-3 wind speeds contain bias and variance parameters that are directly related to our study goals. In particular, each model has parameters that are interpreted as the expected difference between CYGNSS starboard and port measurements and Jason-3 measurements if they had measured wind speed at the exact same location and time. These parameters are estimated *via* maximum likelihood and generalized least squares, which uses a variance-minimizing linear combination of observations to make efficient use of the available data. Computational challenges often associated with Gaussian process models are overcome by downsampling across time and using a state-of-the-art Gaussian process approximation implemented in the publicly available GpGp R package Guinness and Katzfuss (2018). The Gaussian process model and associated computational techniques are the main methodological novelties of this work.

We find that there are significant and persistent differences among some pairs of the CYGNSS sensors of a magnitude up to 1.11 m/s. There are smaller differences between the starboard and port sensors from the same satellite. Five of the eight CYGNSS satellites have a negative bias with respect to Jason-3 measurements. No substantial differences in variances among the eight satellites were detected. These results are successfully validated against a traditional empirical analysis, which shows similar trends but higher uncertainty. In addition, we investigate the possibility that bias depends on wind speed, finding some evidence that CYGNSS measurements are larger than Jason-3 at low wind speeds, though we argue that more careful analysis is needed. We conclude the paper with a discussion of the results and suggestions for how to modify the models to study variation in the biases. All of the code necessary for reproducing our results is available in a GitHub repository at https://github.com/WillBekerman/satellite-wind-speeds.

2. Datasets and Data Processing

We compile one year of measurements recorded by CYGNSS and Jason-3 between September 28, 2019 and September 25, 2020, specifically, CYGNSS Level 2 Science Data Record Version 3.0 and Jason-3 Level-2 X-GDR Data. CYGNSS data is obtained from the OPeNDAP 4 Data Server, also known as Hyrax, at https://podaac-opendap.jpl.nasa.gov/opendap/allData/cygnss/L2/v3.0/. Each CYGNSS satellite is capable of recording 240 wind measurements per minute and takes observations between roughly 38 degrees north and south latitude using its starboard and port antennas. CYGNSS wind speeds are the average surface wind speed of the 25×25 kilometer cell centered on the recorded latitude and longitude. Jason-3 data is obtained from the National Centers for Environmental Information at https://www.ncei.noaa.gov/data/oceans/jason3/gdr/gdr_ssha/. The Jason-3 dataset has about 40 observations per minute and records wind speeds between roughly 66 degrees north and south latitude. Due to several days of missing records in the Jason-3 data during the weeks of February 1, 2020 to February 14, 2020 and June 13, 2020 to June 19, 2020, we omit these weeks from our data collection, yielding 49 total weeks of satellite measurements for our analysis.

After acquiring the data in NetCDF format, we process the data in R, retaining information about spatial location, time of measurement, wind speed, CYGNSS satellite number, and CYGNSS sensor (port vs. starboard). We omit any observations with missing data and standardize the times to seconds since 2020-01-01 00:00 UTC. Since CYGNSS records wind speed measurements only over oceans, we keep only those Jason-3 observations with surface type "open oceans or semi-enclosed seas." The data are saved in standard R Data format.

In Figure 1, we compare the wind speed measurements taken over the same latitudes by CYGNSS 1, CYGNSS 4, and Jason-3 during the week of November 30–December 6, 2019. The value in each pixel is the average of all measurements taken within the pixel over the week. While the spatial patterns of wind speeds are similar among the three satellites, there are subtle differences. CYGNSS 4 appears to

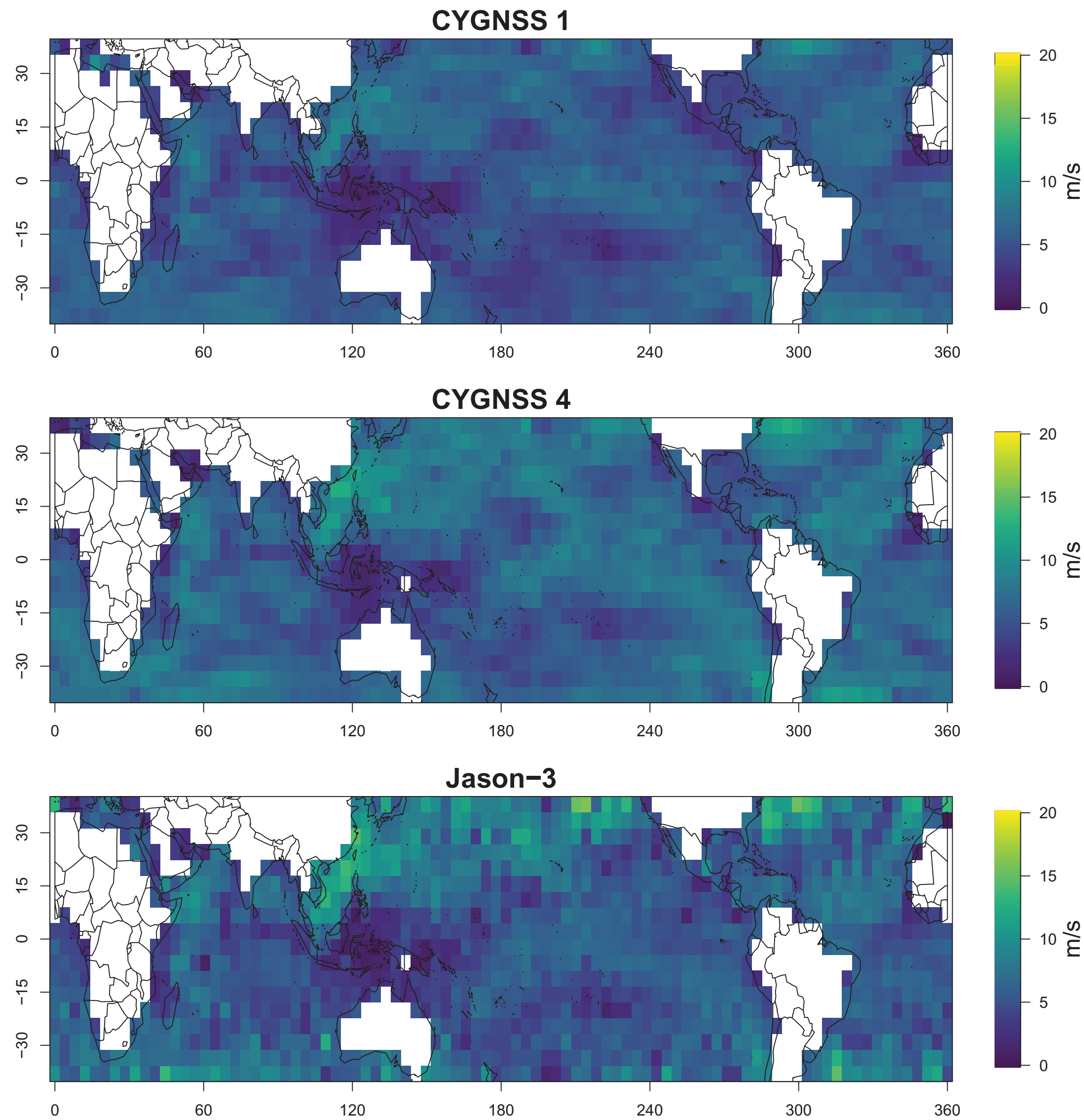

Figure 1. Wind speed (m/s) measurements recorded by CYGNSS 1, CYGNSS 4, and Jason-3 between −38 and 38 degrees longitude between November 30, 2019 and December 6, 2019. The value in each pixel is the sample average of all measurements falling within the pixel during the week. Pixel resolution is 4 degrees latitude and longitude. During this week, CYGNSS 1 has 819,298 observations; CYGNSS 4 has 811,677; and Jason-3 has 380,488, with 202,892 between −38 and 38 degrees latitude.

record larger wind speeds than CYGNSS 1, and Jason-3 tends to have less-smooth wind fields with more of the largest values.

The pixel-wise comparisons in Figure 1 can be misleading because even though the satellites have complete coverage in a one-week period, the timing of the measurements will differ among satellites, so differences could be attributable to the interaction between locally changing wind conditions and the observational times, rather than bias. In seeking to quantify these differences between satellite measurements, the simplest approach is to directly compare measurements recorded by CYGNSS and Jason-3 taken within small space-time windows. To explore the feasibility of the space-time matching analysis, we plot the time-varying distances between the eight CYGNSS satellites and Jason-3 over the course of one full day in Figure 2. All of the CYGNSS satellites come reasonably close to Jason-3 at one or more points during the day, with some variability in the number and proximity of such occurrences. By contrast, while some pairs of CYGNSS satellites nearly always measure winds at nearby locations, some pairs never do, like CYGNSS 1 and CYGNSS 4. In the next section, we propose a model designed to address the challenge of making efficient use of the CYGNSS and Jason-3 comparisons.

3. Analysis

3.1. Model Description

We first describe the statistical model used in our analysis in mathematical notation, and then provide interpretations for the model and its statistical parameters. Our analysis strategy is to fit separate models to datasets consisting of Jason-3 data and one CYGNSS satellite, and repeat the

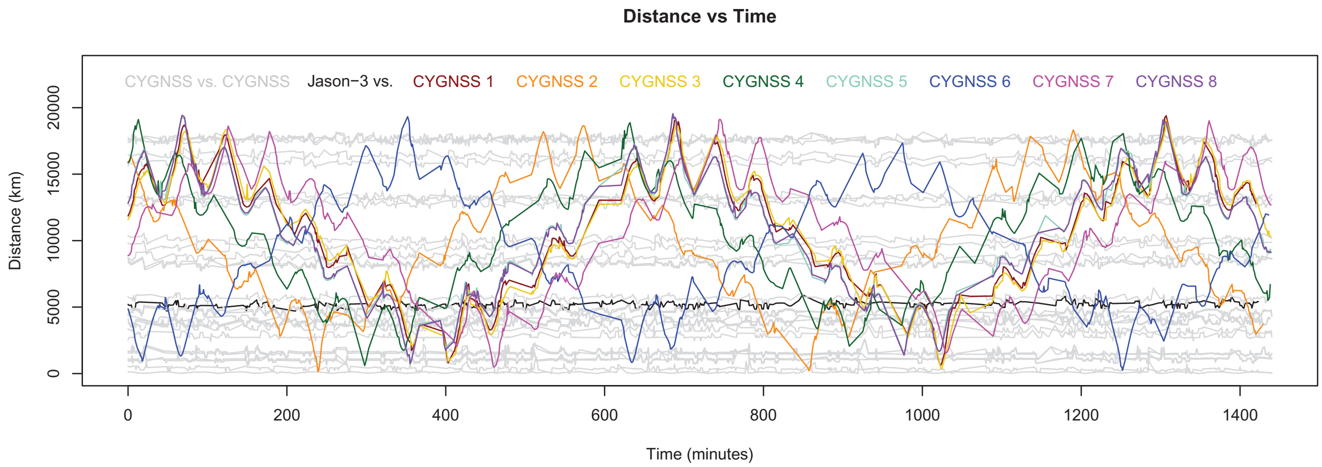

Figure 2. Distances (km) between each pair of CYGNSS micro-satellites and between Jason-3 and CYGNSS on September 29, 2019. Gray lines represent distances between pairs of CYGNSS micro-satellites while colored lines depict distances between Jason-3 and CYGNSS. One black line is shown to represent the distance between CYGNSS 1 and CYGNSS 4.

analysis for each week and each CYGNSS satellite, which means we will have many separate fits of the same general model formulation. This allows us to study whether biases are consistent over time. In the interest of keeping the number of symbols manageable, we do not provide notation for each individual model fit; the notation below is a model for an arbitrary CYGNSS satellite in an arbitrary week. Our results will contain a re-estimation of the parameters for each dataset.

We label the n combined observations from CYGNSS and Jason-3 with $i = 1, ..., n$. We use Y_i for the wind speed associated with observation i, recorded at spatial location (longitude and latitude) x_i and time t_i, and model it as follows:

$$Y_i = \mu_i + a_{k(i)} + Z(x_i, t_i) + \varepsilon_i \tag{1}$$

$$\mu_i = b_0 + b_1 t_i + b_2(\text{lat})_i + b_3(\text{lat})_i^2 + b_4(\text{lat})_i^3 \tag{2}$$

$$Z \sim GP(0, K) \tag{3}$$

$$\varepsilon_1, ..., \varepsilon_n \overset{ind}{\sim} N(0, \sigma^2). \tag{4}$$

The term μ_i is intended to capture broad-scale variation in wind speed over time t_i and latitude, independent of satellite. The mapping $k(i)$ indicates which sensor produced observation i, with $k = 1$ indicating Jason-3, $k = 2$ indicating the CYGNSS starboard sensor, and $k = 3$ indicating the CYGNSS port sensor. Therefore, a_1, a_2, and a_3 are the parameters of interest for determining biases among the sensors. We model space-time variation in wind speeds with the Gaussian process (GP) Z, with inputs spatial location x_i and time t_i.

Gaussian process models impose a normal assumption on all random terms, similar to normal random effects models. They are flexible in that they allow the random effects to be correlated; in this case Z is assumed to be correlated over space and time. Specifically, we assume it has mean zero and space-time Matérn covariances $\text{Cov}(Z(x_i, t_i), Z(x_j, t_j)) = K((x_i, t_i), (x_j, t_j))$, where

$$K((x_i, t_i), (x_j, t_j)) = \frac{\theta_1}{2^{\theta_2 - 1}\Gamma(\theta_2)} (d_{ij})^{\theta_2} \mathcal{K}_{\theta_2}(d_{ij}), \tag{5}$$

$$d_{ij} = \left(\frac{||x_i - x_j||^2}{\theta_3^2} + \frac{(t_i - t_j)^2}{\theta_4^2} \right)^{1/2}, \tag{6}$$

where θ_1 controls the variance of Z, θ_2 is the Matérn smoothness parameter, θ_3 controls the spatial-decay of the covariances, and θ_4 the temporal decay. In the formula for the Matérn function, $\mathcal{K}_{\theta_2}$ is the modified Bessel function of the second kind of order θ_2.

Since Z is a mean-zero process that depends only on space-time location, and not the satellite or random independent error, we interpret it as a wind speed anomaly, with the caveat that the anomaly is relative to a specific linear-in-time, cubic-in-latitude mean field. The mean field contains an intercept b_0, which means that a_1, a_2, and a_3 are not separately identifiable, but differences such as $a_2 - a_1$ and $a_3 - a_1$ are. As a consequence, without additional outside information, our analysis is not able to determine whether CYGNSS or Jason-3 is biased with respect to the true wind field, but it is capable of assessing whether CYGNSS and Jason-3 are biased with respect to one another *via* estimation of differences such as $a_2 - a_1$.

In terms of the model, we are principally interested in how CYGNSS starboard, CYGNSS port, and Jason-3 measurements would differ if they had measured wind speed at the same location at the same time. To see how this quantity relates to our model parameters, suppose that measurement i was taken by CYGNSS starboard and measurement j was from Jason-3, and those two measurements were recorded at the same time and location. Then

$$Y_i - Y_j = a_2 - a_1 + \varepsilon_i - \varepsilon_j \sim N(a_2 - a_1, 2\sigma^2), \tag{7}$$

meaning that $a_2 - a_1$ is the bias and $2\sigma^2$ is the variance of the difference. These are the parameters of interest in our study. By comparing the estimates of $a_2 - a_1$ and $a_3 - a_1$ across CYGNSS satellites, we can achieve our primary goal of understanding variability among the CYGNSS measurements.

3.2. Model Estimation

All of the parameters in the model in Section 3.1 must be estimated for each of the 49 weeks and each of the 8 CYGNSS satellites, requiring a total of 392 model fits. Since the Gaussian process model relies on multivariate normal distributions, which are computationally difficult to handle, we take steps to reduce the computational burden. Specifically, we use a computationally efficient approximation, fit separate models for each CYGNSS satellite for each week, and subset the data to allow us to fit the 392 models within a reasonable amount of time. A typical week has roughly 600, 000 to 1 million observations per CYGNSS satellite and 400, 000 Jason-3 observations. We subset the data to 20, 000 randomly selected observations from CYGNSS and 20, 000 from Jason-3 for each model fit. Retaining 20, 000 observations per week leaves an average of about 30 s between observations and about 180 observations per orbit, which is dense relative to the spatial and temporal scales on which wind speeds vary. We employ a popular computationally efficient Gaussian process approximation proposed by Vecchia (1988), implemented in the R package GpGp Guinness and Katzfuss (2018). This particular approximation relies on an ordering the observations and the conditional distributions of each observation given nearest neighbors from earlier in the ordering. We use the max-min ordering described in Guinness (2018) and 30 neighbors per observation. Each model fit delivers estimates of the model parameters *via* maximization of the approximate likelihood function, as well as standard errors for the mean parameters.

Gaussian process likelihood functions are typically not convex in their parameters, making optimization difficult. GpGp uses a Fisher scoring algorithm to maximize the likelihood function. See Guinness (2021) for details.

3.3. Model-Based Results

In Figure 3, for each week and each CYGNSS satellite, we plot the CYGNSS vs. Jason-3 bias estimates $a_2 - a_1$ (starboard) and $a_3 - a_1$ (port). Estimates for starboard and port from the same week and CYGNSS satellite are connected by a black line. The bias estimates vary by week, satellite, and antenna. Five of the eight CYGNSS satellites (1, 2, 3, 5, and 6) produce negative biases with respect to Jason-3 for every week and for both antennas. The three other satellites (4, 7, and 8) have a mix of negative and positive biases across weeks. CYGNSS 1 has the largest negative biases, on average approximately -0.83 m/s for starboard and -0.94 m/s for the port antenna. Nearly all of the CYGNSS 1 biases are more negative than the most negative CYGNSS 4 bias, suggesting that CYGNSS 1 and CYGNSS 4 were persistently biased with respect to one another during our study period. Within each satellite and antenna, the bias estimates vary by roughly 0.5 to 1.0 m/s from week to week. The differences across weeks in the biases could be due to uncertainty in the parameter estimates, rather than a bias that truly varies over time.

By inspecting the black lines in Figure 3, we observe differences between the estimates of the starboard and port biases from the same satellite within the same week. Figure 4 explores these differences in more detail by directly plotting estimates of $a_3 - a_2$, which measure the bias between the starboard and port antennas. Most of the starboard vs. port bias estimates are smaller than the CYGNSS vs. Jason-3

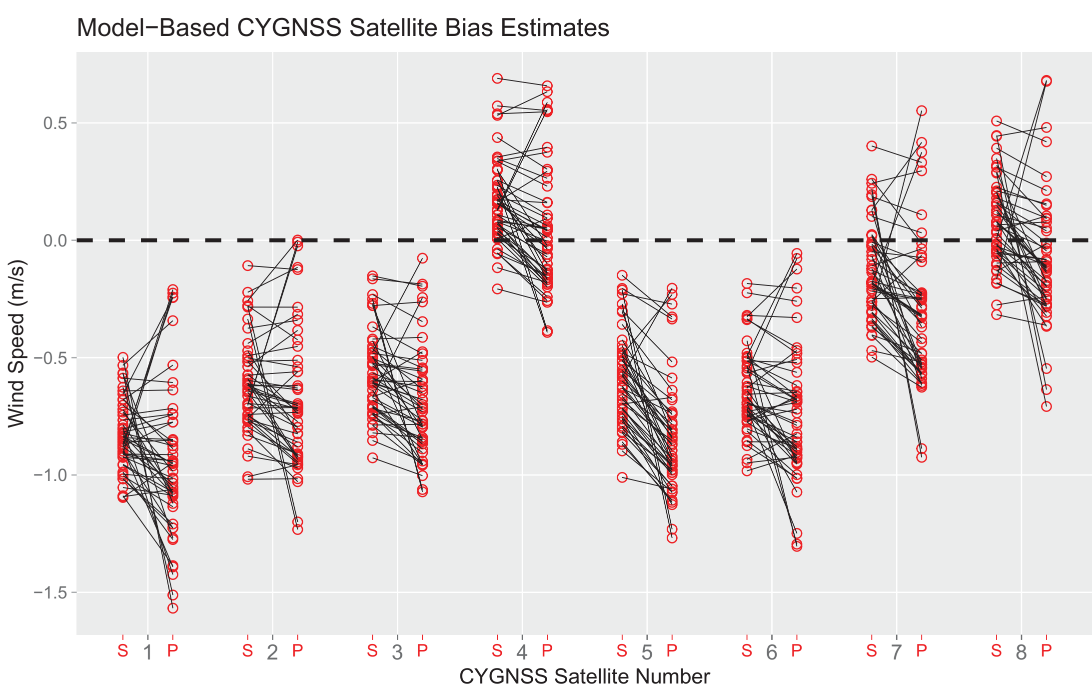

Figure 3. Model-based estimates of CYGNSS starboard (S) vs Jason-3 bias ($a_2 - a_1$) and CYGNSS port (P) vs Jason-3 bias ($a_3 - a_1$) for each CYGNSS satellite and each of the 49 weeks in our study. Starboard and port biases from the same week are connected with a black line. Discussed in more detail in Section 3.3.

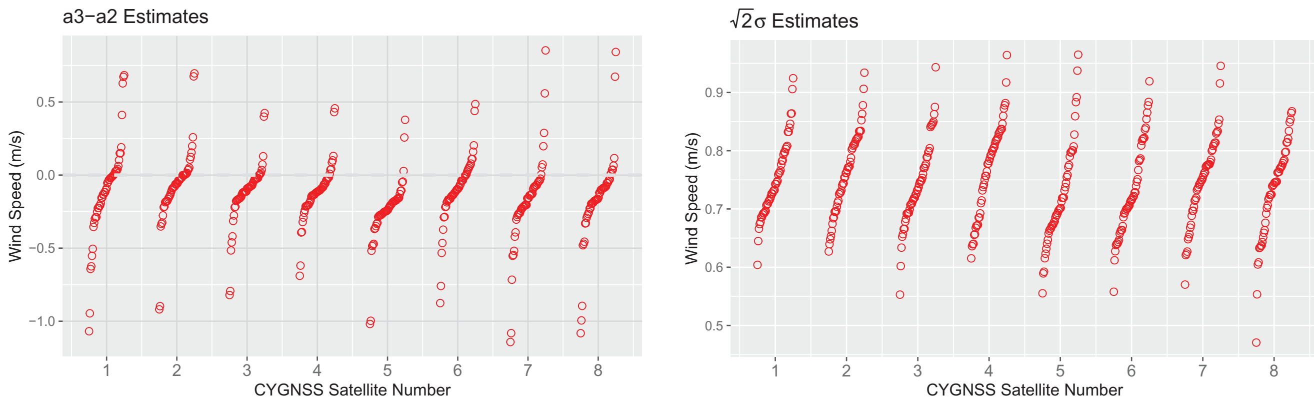

Figure 4. Left: Estimates of $a_3 - a_2$ for each week and each CYGNSS satellite. Right: Estimates of $\sqrt{2}\sigma$ for each week and each CYGNSS satellite. Within each satellite, estimates are sorted and spaced horizontally to visually depict the empirical quantile function of the estimates.

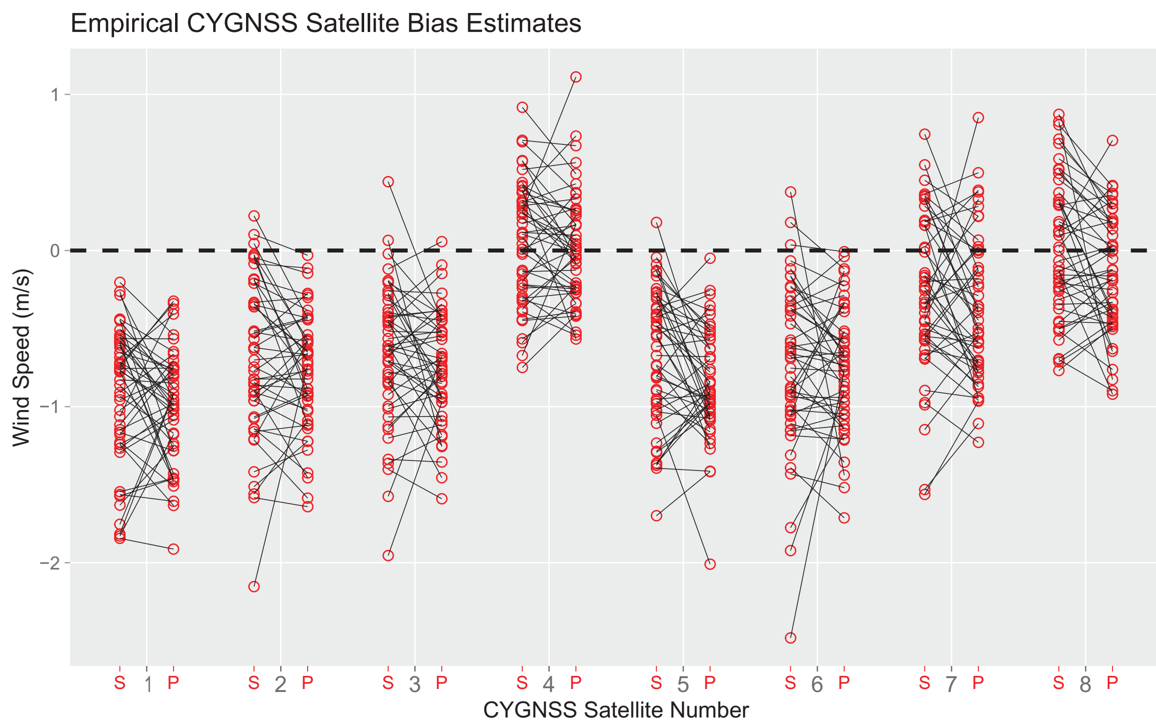

Figure 5. Empirical estimates of CYGNSS starboard (S) vs Jason-3 bias and CYGNSS port (P) vs Jason-3 bias for each CYGNSS satellite and each of the 49 weeks in our study. Starboard and port biases from the same week are connected with a black line. Discussed in more detail in Section 3.4.

biases, with most magnitudes less than 0.25 m/s, and show a mix of both negative and positive biases, though there are notably more negative than positive biases. CYGNSS 5 is the most lopsided with 46 of the 49 biases being negative.

Figure 4 displays estimates of $\sqrt{2}\sigma$, which we recall from Equation (7) is the model's standard deviation of the difference between two observations taken by different sensors at the same location and time. The estimates generally fall between 0.6 and 0.9 m/s, meaning that the size of the noise is roughly equal to the largest CYGNSS vs. Jason-3 biases. There is some variation of the estimates of $\sqrt{2}\sigma$ across weeks but no substantial differences among the eight CYGNSS satellites.

3.4. Empirical Explorations

To validate our model-based results, we conduct additional analyses based on simple averages of differences between CYGNSS and Jason-3 wind speeds that fall within small space-time windows, defined as follows. For each of the eight CYGNSS satellites, each antenna, and each week between September 28, 2019 and September 25, 2020, we divide the week into 2-h windows and find the pair of CYGNSS and Jason-3 observations that are closest in distance within the 2-h window, ignoring any windows that do not have a pair that fall within 25 km. We then take the average of the differences between the selected pairs of CYGNSS and Jason-3 wind speeds. We refer to the averages of these differences as our empirical bias estimates.

Analogously to Figure 3, we plot in Figure 5 the empirical bias estimates over all CYGNSS satellites, antennas, and weeks. The general patterns of the empirical and model-based bias estimates are quite similar. The same five CYGNSS satellites have largely negative biases with respect to Jason-3, while the other three have a mix of negative and positive biases. The average size of the biases are similar as

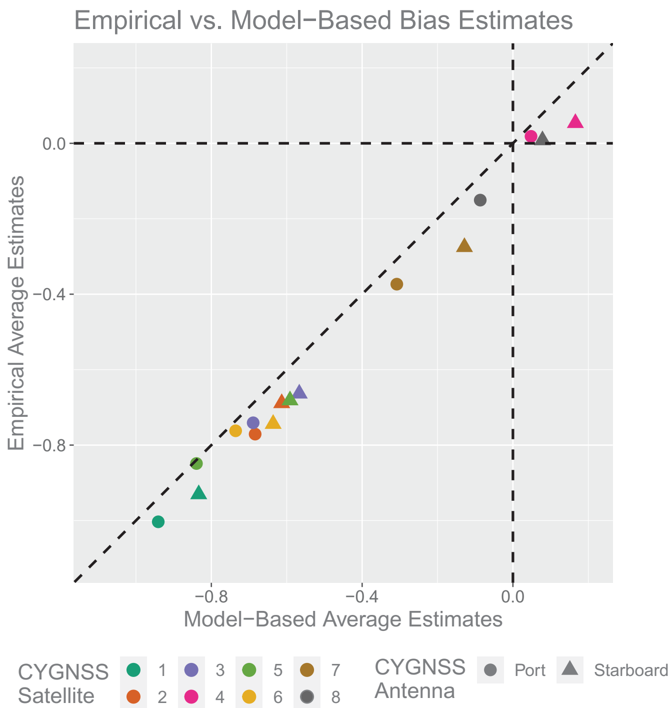

Figure 6. Average CYGNSS satellite bias estimates from our model-based approach and our empirical strategy.

well, ranging roughly from −1.0 to 0.05 m/s. As in the model-based analysis, the port biases in CYGNSS 5 are, on average, more negative than the starboard biases. The empirical biases differ in that the variation across weeks is larger than in the model-based biases, which produces more overlap among the eight CYGNSS satellites and two antennas. For instance, whereas there was essentially no overlap between the CYGNSS 1 and CYGNSS 4 model-based biases, the empirical biases show more substantial overlap. In addition, the difference between the CYGNSS 5 starboard and port model-based biases is more clear than the difference between the empirical biases.

To further compare the model-based and empirical bias estimates, we average the estimates over the 49 weeks and plot them in Figure 6. The estimates generally follow the 45 degree line, with the model-based estimates being slightly more positive than the empirical estimates. The points tend to cluster by satellite, suggesting that the difference between starboard and port within a CYGNSS satellite is generally smaller than the differences among the eight CYGNSS satellites. Interestingly, every starboard point is northeast of its corresponding port point, indicating that the average empirical and model-based starboard biases are more positive than the corresponding port biases for every CYGNSS satellite.

Previous studies have explored whether bias depends on the magnitude of the wind speed. These analyses are complicated by the fact that we never have access to the "true" wind speed. One could take the more accurate measurement as the "true" wind speed and estimate bias as a function of the more accurate measurement. This approach is not without its drawbacks; when one measurement is high, the other is likely to be lower, due to standard regression-to-the-mean. To partially circumvent this issue, we plot in Figure 7 the difference between CYGNSS and Jason-3 (CYGNSS minus Jason-3) against their average. As before, we break the year into 2-h intervals and within each interval, we extract the closest pair of observations, provided that the closest distance is less than 25 km. We see that among the port antennas, CYGNSS measurements are usually larger than Jason-3 for small average wind speed, but for larger average wind speeds, Jason-3 records tend to be larger. The pattern is similar for all satellites. The overall negative bias for satellites 1, 2, 3, 5, and 6 is also evident from the plots. The patterns for the starboard antennas are similar (not shown).

4. Conclusions

Our main finding is that during our study period of September 2019 to September 2020, persistent biases existed among the wind speed measurements recorded by the eight CYGNSS satellites and between some CYGNSS satellites and Jason-3. Considering the averages of the model-based parameter estimates over the study period, the largest bias between pairs of CYGNSS sensors was 1.11 m/s (CYGNSS 4 starboard minus CYGNSS 1 port), and the largest CYGNSS vs. Jason-3 bias was −0.94 m/s (CYGNSS 1 port minus Jason-3). We discovered smaller biases between the starboard and port antennas within a satellite, with the largest average bias being 0.25 m/s (CYGNSS 5 starboard – CYGNSS 5 port).

It is not surprising to us that the two sensors on the same CYGNSS satellite would be reasonably well-calibrated with respect to one another. Since they tend to measure wind speeds at fairly close locations, direct comparison across antennas is easier. However, direct comparisons between CYGNSS satellites are more difficult due to the fact that some pairs nearly always measure wind speeds at disparate locations. Similar to other studies that make indirect comparisons with forecast winds or buoy data, we use Jason-3 as an intermediary to achieve indirect comparisons among every pair of CYGNSS satellites. We believe that Jason-3 data is appropriate because both CYGNSS and Jason-3 attempt to measure snapshots of wind speed within small space-time windows, and they pass each other somewhat regularly.

These findings were facilitated by the use of a Gaussian process model that contained parameters directly related to the expected difference between measurements from different instruments taken at the same time and location. In addition to the bias parameters, the models contained a parameter related to the variance of the difference between two observations taken by different sensors at the same time and location. We did not find substantial differences in the estimates of noise parameters among the models for the eight CYGNSS satellites. The size of the noise averaged about 0.75 m/s, meaning that the noise was of roughly equal magnitude to the largest biases, implying that about half of the mean square errors between measurements from different sensors was due to bias, half due to noise. In other words,

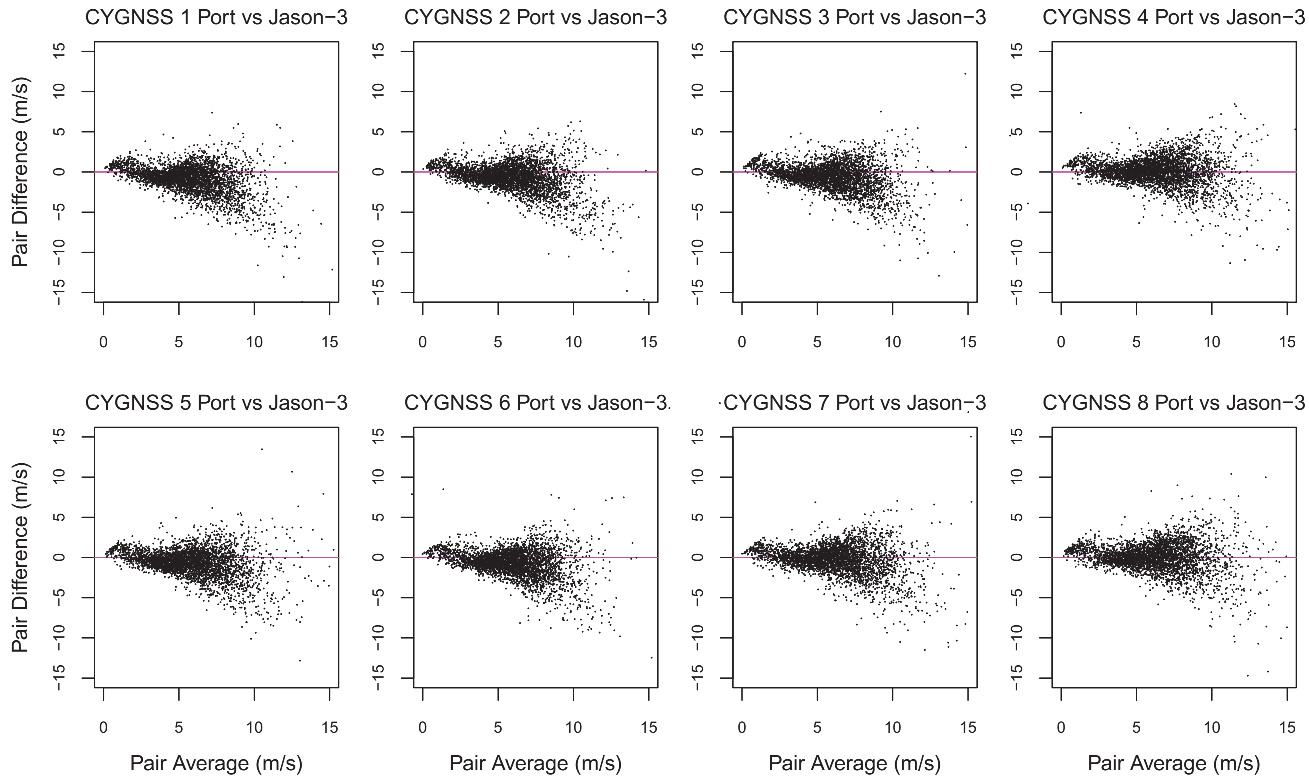

Figure 7. CYGNSS Port minus Jason-3 vs. the average of CYGNSS Port and Jason-3 for pairs of observations. Each pair represents the two closest observations from each 2 h window over the entire year, provided that the distance is less than 25 km.

eliminating the bias could potentially reduce the mean squared errors by a factor of two.

We validated the model-based findings with a more traditional empirical analysis that searched for matching pairs of observations from different sensors within small space-time windows. The general pattern of the estimates in the empirical analysis was similar to that in the model-based analysis. The analyses differed in that the week-to-week variation in the empirical estimates of the bias was larger. This is expected because the model-based estimates use generalized least squares, which provides variance-minimizing parameter estimates under the assumed model.

Some studies have suggested that the size and direction of the bias may depend on the magnitude of the true wind speed. We caution against over-interpreting these results because we do not have access to the true wind speed. If a particular sensor is chosen as the reference, we expect to see positive bias at low wind speeds and negative bias at high wind speeds due to regression-to-the-mean effects, even if the bias does not vary with wind speed. We attempted to mitigate this effect by comparing paired differences against paired averages, finding that generally the Jason-3 wind speeds increase relative to CYGNSS as the average of their two measurements increases. This aspect is certainly worthy of more exploration, with consideration of the aforementioned statistical issues. Due to the differing accuracies of CYGNSS and Jason-3, the average may not be the best estimate of the wind speed. We could also seek out a third wind speed measurement to serve as the baseline. One could also pursue model-based estimates of biases that depend on true wind speed. To this end, consider the following extension of our model:

$$Y_i = \mu_i + a_{k(i)} + b_{k(i)} Z(x_i, t_i) + \varepsilon_i, \tag{8}$$

which contains a sensor-dependent slope multiplying the wind speed anomaly $Z(x_i, t_i)$. Inferences about biases that depend on wind speed could be obtained *via* estimation of $a_{k(i)}$ and $b_{k(i)}$. This is still a Gaussian process model, so we could use the same methodology to fit the model, though one would have to be careful about non-identifiability of parameters; for example, the variance of $Z(x_i, t_i)$ is not identifiable separately from the $b_{k(i)}$ parameters.

One could imagine that the bias depends on various other factors, such as latitude, time, or GNSS satellite. This sort of variation could be handled within our model framework by adding interactions between the bias and the desired factor. To capture biases that vary in space, we could extend our model as

$$Y_i = \mu_i + a_{k(i)} + Z(x_i, t_i) + W_{k(i)}(x_i) + \varepsilon_i, \tag{9}$$

where W_1, W_2, and W_3 are independent spatial Gaussian processes. Then we can interpret $a_2 - a_1 + W_2(x_i) - W_1(x_i)$ to be the spatially-varying starboard bias, and $a_3 - a_1 + W_3(x_i) - W_1(x_i)$ as the spatially-varying port bias. We suspect that we would need to use more than one week of data at a time to accurately estimate a spatially-varying bias.

5. Supplementary Materials

We run our analysis using R 4.0.5 R Core Team (2013) on platform x86_64-w64-mingw32/x64 running under Windows 8 x64. We make frequent usage of R packages" fields"

Douglas Nychka et al. (2017)," maps" Becker et al. (2018), and" ggplot2" Wickham (2016) for creating visualizations, and" GpGp" Guinness and Katzfuss (2018) for modelling.

We maintain a Github repository at https://github.com/WillBekerman/satellite-wind-speeds which contains all data, as well as R scripts to replicate our analysis.

Acknowledgments

The authors would like to thank David Moroni for technical support downloading CYGNSS data.

Funding

This work was supported in part by the National Science Foundation, Division of Mathematical Statistics under grant numbers 1916208 and 1953088.

ORCID

William Bekerman http://orcid.org/0000-0002-6957-6863
Joseph Guinness http://orcid.org/0000-0003-0564-6639

References

Adelekan IO. 2000. A survey of rainstorms as weather hazards in southern Nigeria. Environmentalist. 20(1):33–39.

Asharaf S, Waliser DE, Posselt DJ, Ruf CS, Zhang C, Putra AW. 2021. CYGNSS ocean surface wind validation in the tropics. J Atmos Oceanic Technol. 38(4):711–724.

Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A. 2018. maps: draw geographical maps. R package version 3.3.0. Available from: https://CRAN.R-project.org/package=maps.

Daniels J, Bresky W, Wanzong S, Velden C, Berger H. 2019. GOES-R advanced baseline imager (ABI) algorithm theoretical basis document for derived motion winds. Available from: https://www.star.nesdis.noaa.gov/goesr/docs/ATBD/DMW.pdf

Eichelberger S, McCaa J, Nijssen B, Wood A. 2008. Climate change effects on wind speed. North American Windpower. 7:68–72.

Endalew AM, Debaer C, Rutten N, Vercammen J, Delele MA, Ramon H, Nicolaï B, Verboven P. 2010. A new integrated CFD modelling approach towards air-assisted orchard spraying. Part I. Model development and effect of wind speed and direction on sprayer airflow. Comput Electron Agric. 71(2):128–136.

Guinness J. 2018. Permutation and grouping methods for sharpening Gaussian process approximations. Technometrics. 60(4):415–429.

Guinness J. 2021. Gaussian process learning via fisher scoring of Vecchia's approximation. Stat Comput. 31(3):1–8.

Guinness J, Katzfuss M. 2018. GpGp: fast Gaussian process computation using Vecchia's approximation. R package version 040. https://cran.r-project.org/web/packages/GpGp/GpGp.pdf

Katzfuss M, Guinness J, Gong W, Zilber D. 2020. Vecchia approximations of Gaussian-process predictions. JABES. 25(3):383–414.

Kuusela M, Stein ML. 2018. Locally stationary spatio-temporal interpolation of argo profiling float data. Proc R Soc A. 474(2220):20180400.

Nychka D, Furrer R, Paige J, Sain S. 2017. fields: tools for spatial data. R package version 11.6; Available from: https://github.com/NCAR/Fields.

Pascual D, Clarizia MP, Ruf CS. 2021. Improved CYGNSS wind speed retrieval using significant wave height correction. Remote Sens. 13(21):4313.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org/.

Rayner NA, Auchmann R, Bessembinder J, Brönnimann S, Brugnara Y, Capponi F, Carrea L, Dodd EM, Ghent D, Good E, et al. 2020. The EUSTACE project: delivering global, daily information on surface air temperature. Bull Am Meteorol Soc. 101(11):E1924–E1947.

Ruan W, Milstein AB, Blackwell W, Miller EL. 2017. Multiple output Gaussian process regression algorithm for multi-frequency scattered data interpolation. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS); p. 3992–3995.

Ruf CS, Gleason S, Jelenak Z, Katzberg S, Ridley A, Rose R, Scherrer J, Zavorotny V. 2012. The CYGNSS nanosatellite constellation hurricane mission. 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE; p. 214–216.

Ruf C, Unwin M, Dickinson J, Rose R, Rose D, Vincent M, Lyons A. 2013. CYGNSS: enabling the future of hurricane prediction [remote sensing satellites]. IEEE Geosci Remote Sens Mag. 1(2):52–67.

Rusu L, Raileanu AB, Onea F. 2018. A comparative analysis of the wind and wave climate in the black sea along the shipping routes. Water. 10(7):924.

Saïd F, Jelenak Z, Park J, Chang PS. 2022. The NOAA track-wise wind retrieval algorithm and product assessment for CYGNSS. IEEE Trans Geosci Remote Sens. 60:1–24.

Smith M, Chow L. 1998. Prediction method for aerodynamic noise from aircraft landing gear. 4th AIAA/CEAS Aeroacoustics Conference; p. 2228.

Susiluoto J, Spantini A, Haario H, Härkönen T, Marzouk Y. 2020. Efficient multi-scale Gaussian process regression for massive remote sensing data with satgp v0. 1.2. Geosci Model Dev. 13(7):3439–3463.

Vecchia AV. 1988. Estimation and model identification for continuous spatial processes. J R Stat Soc Ser B (Methodol). 50(2):297–312.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag. Available from: https://ggplot2.tidyverse.org.

Young I, Sanina E, Babanin A. 2017. Calibration and cross validation of a global wind and wave database of altimeter, radiometer, and scatterometer measurements. J Atmos Oceanic Technol. 34(6):1285–1306.