Statistical Derivation of Wind Speeds From CYGNSS Data

Maria Paola Clarizia, *Senior Member, IEEE*, and Christopher S. Ruf, *Fellow, IEEE*

***Abstract*— In this article, a statistical methodology to estimate wind speed from CYGNSS observables is proposed and implemented. The approach uses the cumulative distribution function (cdf) of the observable and of the ground-truth reference winds. It depends only on the statistical distributions of the CYGNSS data and the wind speed, and therefore, is simpler to implement than alternative approaches requiring coincident matchups between the data and the ground truth. This cdf matching method produces retrieved winds with a probability density function that is very close to that of the ground-truth winds. When compared to the current CYGNSS baseline winds for fully developed seas, the cdf matching winds show better behavior and agreement with reference wind speeds over the low to medium wind speed range, which constitutes the majority of the wind population that drives the statistics used by the algorithm. The performance is robust with respect to measurement geometry and transmitter and receiver hardware parameters, with the exception of a dependence of the error on the GPS satellite identifier (ID), probably due to uncorrected variations in GPS equivalent isotropically radiated power (EIRP). Validation using modeled winds and winds measured by other satellites reveals that CYGNSS winds behave in a very similar manner as the winds modeled by the Global Data Assimilation System (GDAS).**

***Index Terms*— GNSS-reflectometry.**

I. Introduction

GNSS-REFLECTOMETRY exploits reflections from navigation satellites, to estimate geophysical properties of the surface of the Earth [1]. In recent years, GNSS-R has proven particularly successful at estimating wind speed over oceans [2]–[7]. The capability to provide wind speed estimations even through rain with L-Band signals and the suitability of GNSS-R receivers to form a constellation of satellites are two major strengths of spaceborne GNSS-R, and the two main drivers behind the selection and the success of the NASA CYGNSS Mission [8], [9]. CYGNSS was launched in December 2016, and it continues to provide wind speeds with high space–time sampling and sufficiently low uncertainty, proving very useful to the science community [10]–[12]. The baseline retrieval implemented in the CYGNSS Science Operation Center (SOC) is based on the exploitation of two GNSS-R observables, called Delay/Doppler Map Average (DDMA) and Leading Edge Slope (LES), and derived from Delay/Doppler Maps (DDMs). Two empirical geophysical model functions (GMFs)—one for DDMA and one for LES—are generated to provide a wind speed estimate for each value of the observable. The estimates from DDMA and LES are then combined to produce a minimum variance (MV) estimator, which offers improved performance compared to the use of DDMA or LES alone [5]. [6]. A parametric model to fit the empirically derived GMF is also proposed in [6], and the final root mean square difference (RMSD) is estimated to be 2.0 m/s for wind speeds below 20 and 6.5 m/s for winds above 20 m/s. One drawback of the baseline method is that the development of its GMF requires every individual CYGNSS observation to be matched up with a near-coincident reference wind speed. The full range of possible wind speed values and measurement geometries is generally not present in sufficient number, resulting in the need for manual averaging and interpolating of the data to fully populate the GMF. In addition, a final pdf matching step is applied by the baseline algorithm which forces the pdf of the retrieved winds to match that of the wind matchups [13].

Here we propose an alternate retrieval methodology based on the statistics of the observable and of the ground-truth wind speed rather than on matchups between them. The method is based on the 1994 study by Freilich and Challenor [14] and is hereafter called the cdf matching method. It is simple and computationally easy to implement, and it produces retrievals with a pdf that naturally matches that of the wind matchups, thus not requiring the extra pdf matching step described above. The cdf matching algorithm also provides improved performances over the baseline algorithm in the low to medium wind speed range.

This article is organized as follows: Section II explains the data and reference wind matchups used for the analysis, as well as the filters applied to the data. Section III illustrates the retrieval methodology and the results. Section IV presents the performance characterization of the proposed algorithm with respect to geometry and hardware parameters, and transmitters and receivers. Section V illustrates a validation study using wind estimates from other models and other satellites. Section VI provides an overall summary and future perspectives.

II. CYGNSS Data and Ground-Truth Information

The CYGNSS data used in this analysis are the Level 1 (L1) version 2.1 data, available from http://podaac.jpl.nasa.gov/.

Manuscript received February 26, 2019; revised September 19, 2019, October 24, 2019, and October 30, 2019; accepted November 13, 2019. This work was developed under Contract SUBK00009666 between the University of Michigan and Deimos Space U.K., Ltd., issued under NASA Contract NNL13AQ00C. *(Corresponding author: Maria Paola Clarizia.)*

M. P. Clarizia is with Deimos Space UK Ltd, Harwell Business Campus, Didcot OX11 0QR, U.K. (e-mail: maria-paola.clarizia@deimos-space.com).

C. S. Ruf is with the Department of Climate and Space Sciences and Engineering (CLASP), University of Michigan, Ann Arbor, MI 48109-2143 USA.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2959715

The CYGNSS data files include the DDMA and LES observables, as well as a series of metadata describing the geometry and the instrument parameters at the time of the acquisition. Here we use CYGNSS data acquired between May 1, 2017 and October 3, 2017 to demonstrate the concept and collocated modeled 10 m referenced ocean surface wind speeds from the European Center for Medium-range Weather Forecast (ECMWF) as reference matchups [15]. The matchups are produced via bilinear interpolation in space and linear interpolation in time.

The data are filtered according to the following criteria.

1) The observables need to have good quality, which is determined by the Quality Control (QC) flag in the data.
2) The observables, as well as the wind speed matchups, need to be positive.
3) The measurements taken when the star tracker is not tracking due to solar contamination are discarded.
4) The measurements from the GPS Block type II-F satellites are discarded, due to the lack of accurate information on the transmitter antenna gain pattern for this family of GPS satellites.
5) The range-corrected gain (RCG) of the measurements is higher than 3. The definition and description of RCG can be found in [5].

Some of these criteria are the same as those adopted in [6].

III. Retrieval Methodology

The methodology used here was first illustrated in [14], and it was originally developed to derive a GMF to estimate wind speed from radar cross sections measured by altimeters. The algorithm is based upon the assumption that a model function M relating the wind speed w to an observable o—in our case, either DDMA or LES—exists as follows:

$$w = M(o) \tag{1}$$

and the assumption that w varies monotonically with o, such that the inverse model function $o = M^{-1}(w)$ also exists.

Then, the cdf of the observable o can be written as follows:

$$F_o(o') = P(o \le o') = P[M^{-1}(w) \le o'] \tag{2}$$

$$= P[w \ge M(o')] = 1 - F_w[M(o')]. \tag{3}$$

Hence, if $F_o(o') = \beta$, then the corresponding wind speed $w' = M(o')$ is the wind speed for which $1 - F_w(w') = \beta$. The detailed derivation of the method is contained in [14].

The main steps to implement the method can be summarized as follows.

1) Construct the empirical cdfs for both the DDMA observable $F_{\mathrm{DDMA}}(\mathrm{DDMA})$ and the wind speed $F_w(w)$.
2) For each observable DDMA', find its cdf value $F_{\mathrm{DDMA}}(\mathrm{DDMA}') = \beta$.
3) Retrieve the wind corresponding to DDMA' as the wind value w' such that $1 - F_w(w') = \beta$.

One advantage of the cdf matching method is that it only requires the generation of the cdf of the observable and wind speed. There is neither the need to assemble the wind speed matchups exactly collocated with the CYGNSS acquisitions—which is typically needed to construct GMF—nor the need to empirically tune a large number of parameters, which is done for baseline retrievals [5], [6]. In principle, it can be applied to any other GNSS-R observable, to estimate any geophysical parameter, provided that: 1) the parameter itself varies monotonically with the observable [14] and 2) a large population of samples is available to derive robust statistics, and such statistics are themselves stable. As such, the cdf matching method is not suitable to derive wind speed in statistically nonstationary conditions, such as near-major storms or tropical cyclones, as the number of overpass data in these cases is much lower and the statistics of winds in these regimes is neither well known nor repeatable from storm to storm.

Similar to [5], the population of data obtained after the filters described in Section II is randomly split into training and test data sets, each of which represents half of the entire data set. The cdfs are derived only from training data, while the algorithm performances are evaluated using the test data set.

The training data population is then split into bins of incidence angles and RCG, and cdfs of the DDMA and LES observables are derived for each bin.

This segmentation is performed as the results are improved when both the incidence angle variation and the RCG variation are considered in the retrieval. This is similar to the baseline approach, where a different GMF is obtained for different incidence angle bins, but the RCG dependence of the GMF is not considered in that case. The improvement in the results when RCG is also considered as an input variable in the retrieval algorithm is most likely due to residual calibration effects, not fully corrected by the current L1 calibration implemented in CYGNSS.

The bins are chosen of 5° width in incidence angle, and ten RCG bins within each incidence angle bin are identified so that each bin contains 10% of the population of the incidence angle bin.

A. Results for DDMA and LES Retrievals

The wind estimates obtained from the cdf matching method are first compared to the fully developed seas (FDS) retrievals, which represent the baseline retrievals [6], so they are hereafter called either FDS or baseline retrievals. CYGNSS also outputs the Young Sea Limited Fetch (YSLF) wind retrievals, which typically provide better estimates under tropical storm or cyclone conditions, but the cdf matching retrievals are not suited for comparison with YSLF retrievals, as explained in Section III. The FDS and cdf matching retrievals are shown, respectively, in Fig. 1(a) and (b) for DDMA, and in Fig. 2(a) and (b) for LES. Here the FDS DDMA retrievals have been computed using the latest version of the empirical GMF, developed for v.2.1 of data [16].

Fig. 1 illustrates that the cdf matching method produces a more symmetric wind distribution around the 1:1 line and less erroneous high retrieved winds (>10 m/s) for reference winds between 5 and 10 m/s.

A better-retrieved wind distribution is also shown for the cdf matching case in Fig. 2, although the LES performance

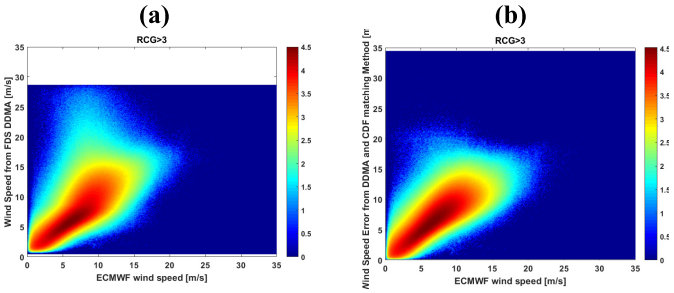

Fig. 1. Density-plot in log-scale of the reference ECMWF wind speeds (x-axis) versus (a) CYGNSS baseline wind speed and (b) cdf matching wind retrievals, obtained using the DDMA observable.

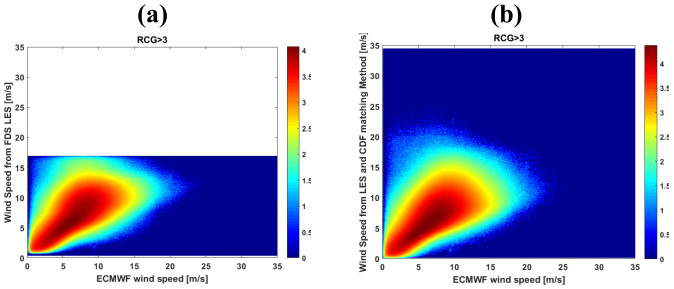

Fig. 2. Density-plot in log-scale of the reference ECMWF wind speeds (x-axis) versus (a) CYGNSS baseline wind speed and (b) cdf matching wind retrievals, obtained using the LES observable.

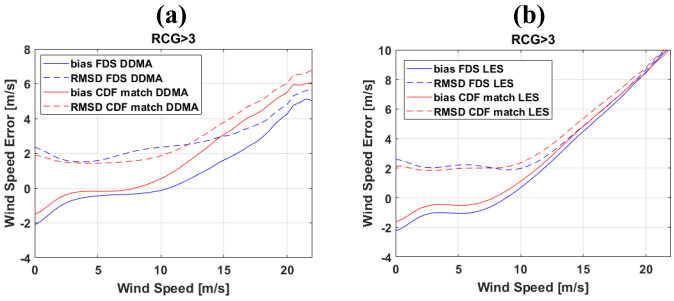

Fig. 3. Bias (reference minus retrieved) wind speed (cont. line) and wind speed RMSD (dashed line) as a function of reference wind speed, for baseline (blue) and cdf matching (black) method, and for (a) DDMA and (b) LES.

is worse than that of the DDMA. The baseline method cuts the wind speed at ∼17 m/s based on the behavior of the LES observable relative to the noise floor, but the symmetry around the 1:1 line is still improved when the cdf matching method is adopted.

Fig. 3 shows the wind retrieval bias and RMSD, for the baseline case (blue) and the cdf matching case (black), respectively, for DDMA [see Fig. 3(a)] and LES [Fig. 3(b)].

The cdf matching method provides a lower bias for winds up to ∼9 m/s, and a lower RMSD for winds up to ∼12 m/s. For wind speeds between ∼9 and ∼12 m/s, the bias of the cdf matching method is higher but the RMSD remains lower.

The improved performances are due to the fact that the wind speed range between 0 and 12 m/s contains the majority of the entire population; hence, the wind and observable statistics, which determine the cdfs, are determined by the winds in such range.

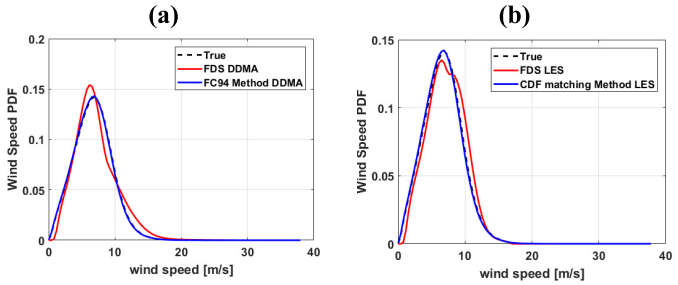

Fig. 4. PDF of reference (ECMWF) winds (dashed black) and CYGNSS winds retrieved using the FDS method (red) and the cdf matching method (solid blue). (a) DDMA retrievals. (b) LES retrievals.

Slightly higher errors are instead observed in the high wind speed range (>12 m/s) in the cdf matching case, compared to the baseline case. The high wind degradation is expected, as the cdf matching method gives less weight to the high wind speeds as they represent a very small part of the population.

Similar considerations apply for the LES results shown in Fig. 3(b), although the difference between baseline and cdf matching performances is smaller, and the performances is worse than that of the DDMA in both the baseline and cdf matching cases.

Fig. 4 shows the pdf of reference winds (black) and winds retrieved using the baseline method (red), and the cdf matching method (blue), respectively, for DDMA [see Fig. 4(a)] and LES [see Fig. 4(b)]. The cdf matching method produces by design pdf of retrieved winds virtually identical to that of ground truth winds, as opposed to the FDS pdf which shows stronger discrepancies with respect to the reference pdf.

B. Results for MV Retrievals

Here the DDMA and LES retrievals obtained using the cdf matching method are combined into a MV estimator as described in [5] and [6]. The MV estimator is essentially a best-weighted estimator, which exploits the degree of decorrelation present between the errors in the individual DDMA and LES estimates to minimize the error in the wind speed estimate. The MV estimator is a linear combination of the individual estimates, with weighting obtained by requiring that the estimator be unbiased (i.e., the expected value of its retrieval is equal to the true quantity to be estimated) and by minimizing its variance. The results of the MV estimator from the cdf-matching-based DDMA and LES retrievals are compared against the baseline v2.1 MV retrievals. The approach used to generate the final MV estimator is the same both in the baseline case and in the cdf matching case. Fig. 5 shows the log-density plots of MV retrievals against reference winds, for (a) baseline and (b) cdf matching retrievals. The improvement brought by the MV combination is more noticeable in the baseline case than that in the cdf matching case, but the relationship between reference and retrieved winds is also sharpened by the MV combination in the cdf matching case, where the dispersion of the winds around the 1:1 line is further reduced, compared to the behavior in Fig. 1(b). Figs. 6 and 7 show, respectively, the curves of bias and RMSD, and the pdf of reference versus retrieved winds. Similar considerations

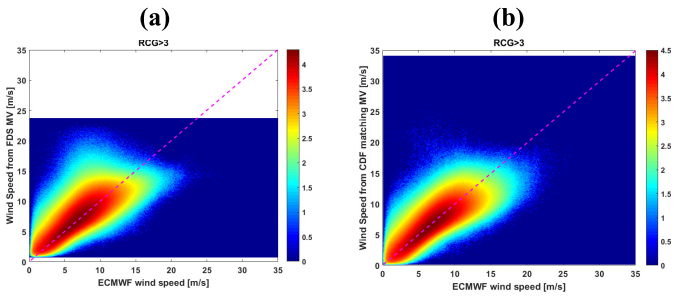

Fig. 5. Density-plot in log-scale of the reference ECMWF wind speeds (x-axis) versus (a) CYGNSS baseline wind speed and (b) cdf matching wind retrievals, obtained using the MV estimator.

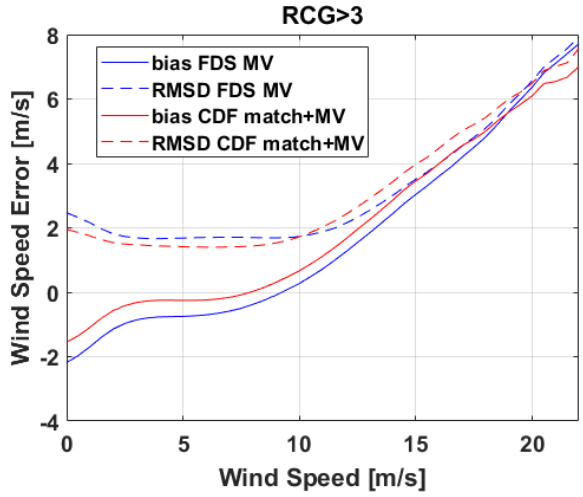

Fig. 6. Bias (reference minus retrieved) wind speed (cont. line) and wind speed RMSD (dashed line) as a function of reference wind speed, for baseline (blue) and cdf matching (black) method, and for the MV retrievals.

as those for Fig. 3 apply to Fig. 6, where the cdf matching performs slightly better in the low/medium wind speed regime and slightly worse in the high wind speed regime, compared to the baseline methodology. The overall performances of both methods in the high wind speed regime are lower than those provided by the CYGNSS estimations in a regime of YSLF, which should to date be considered as the optimal estimations at high winds from CYGNSS data.

In Fig. 7, the pdf of the retrievals obtained using the cdf matching method and the MV combination slightly deviates from that of the reference winds. This minor degradation of the pdf similarity is expected, as the linear combination of DDMA and LES retrievals—each of which produced a pdf coincident with the reference pdf—will inevitably impose slight changes to the pdf of the final retrieved winds, such that they no longer coincide with the reference pdf. However, the difference is small and the cdf matching pdf in Fig. 6 is still closer to the reference pdf than the baseline pdf. Fig. 8 illustrates the logarithmic pdf of the wind error (reference minus retrieved) for the baseline case (red) and for the cdf matching case (blue). In the cdf matching case, the tails of the pdf—i.e., the distribution for wind errors below -5 m/s and above 5 m/s—are more symmetric, which means there is a more balanced population of positive and negative errors; the errors are also

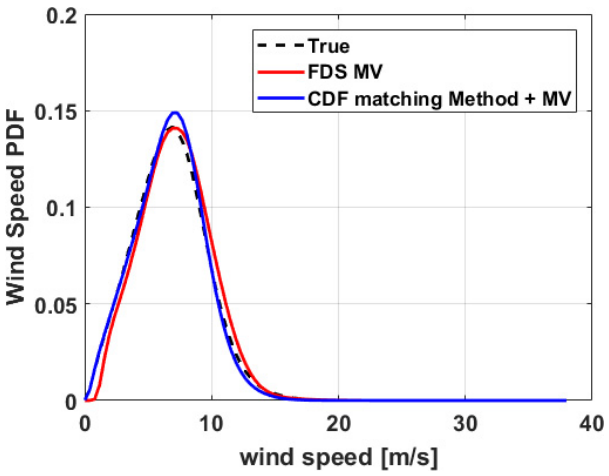

Fig. 7. PDF of reference (ECMWF) winds (dashed black) and CYGNSS winds retrieved using the FDS method (red) and the cdf matching method (solid blue), for the MV retrievals.

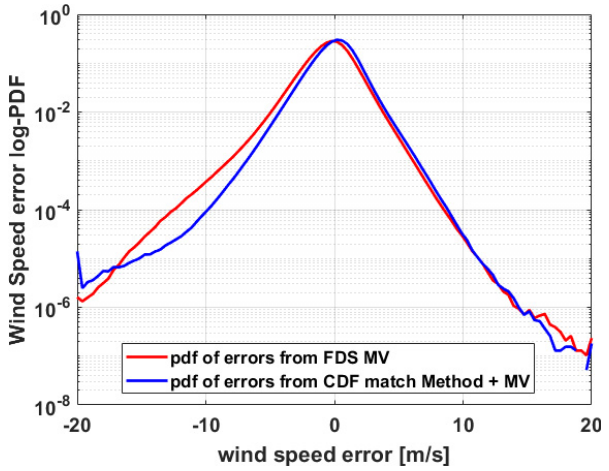

Fig. 8. Logarithmic pdf of wind speed error (ECMWF minus CYGNSS) for the FDS retrievals (red) and cdf matching retrievals (blue).

TABLE I

BIAS AND RMSD FOR BASELINE AND CDF MATCHING APPROACH, AFTER MV COMBINATION. THE STATISTICS ARE COMPUTED USING TEST DATA SET ONLY

	Bias [m/s]	RMSD [m/s]
Baseline MV	-0.38	1.78
CDF Match MV	0.07	1.62

much lower in the range of negative wind errors, compared to the baseline case. Table I summarizes the overall performances of the baseline and cdf matching algorithms with the MV combination applied. The cdf matching method is virtually unbiased and has a reduction in RMSD of 9% compared to the baseline methodology.

IV. PERFORMANCE CHARACTERIZATION WITH RESPECT TO GEOMETRY AND SATELLITE PARAMETERS

In this section, we analyze the performance of the algorithm with respect to geometry and satellite-related parameters to

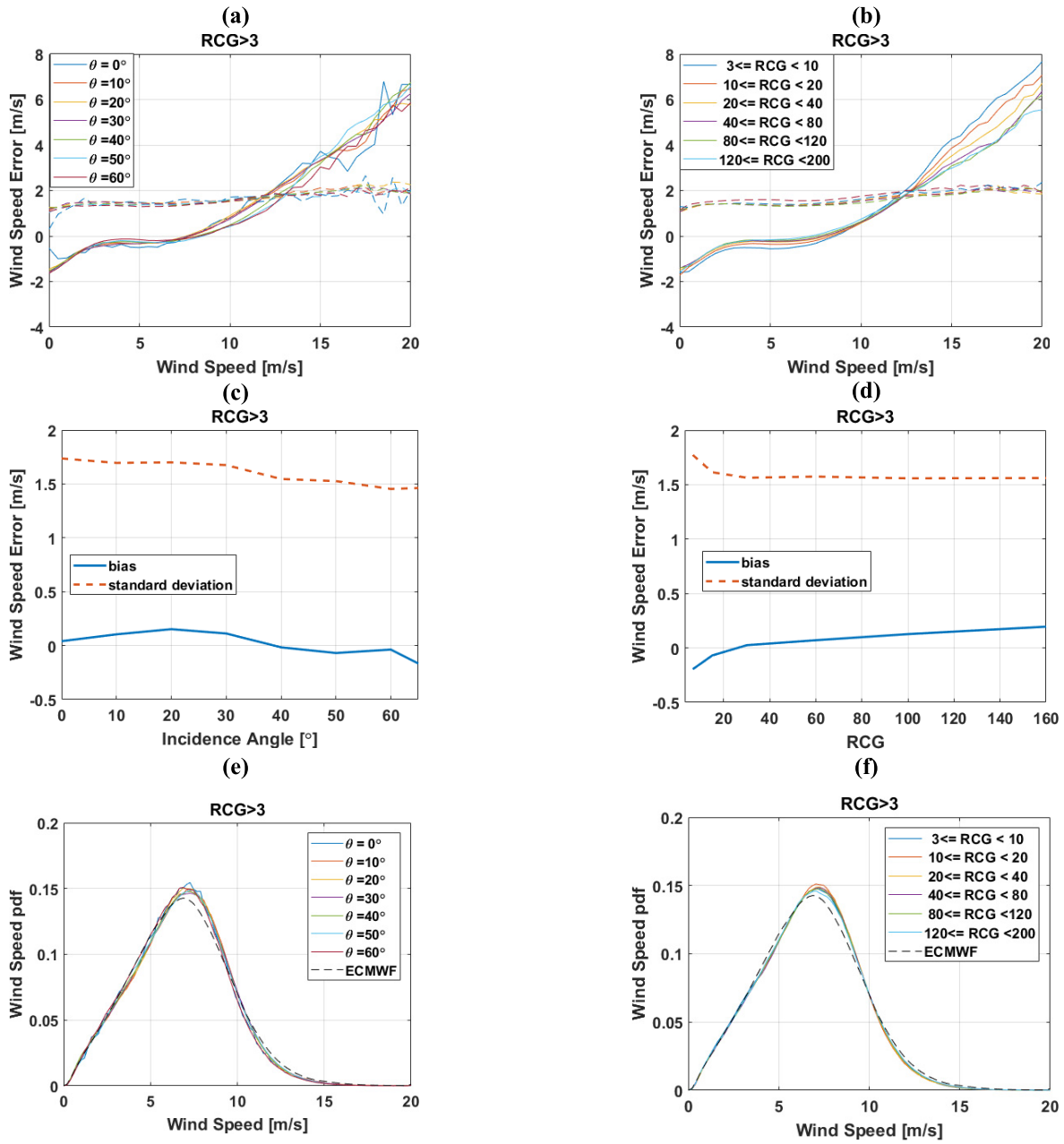

Fig. 9. (Top row) Wind error statistics (bias as cont. line, standard deviation of the wind speed difference as dashed line), as a function of reference winds, color coded by (a) incidence angle and (b) RCG. (Middle row) Total bias (blue) and standard deviation of the wind speed difference (red) as a function of (c) incidence angle and (d) RCG. (Bottom row) PDF of retrieved winds, color coded by (e) incidence angle and (f) RCG, with overlapped pdf of ECMWF winds (dashed black).

investigate potential dependencies of the retrieval error on these parameters.

We use data from August and September 2017, and we consider the following:

1) incidence angle;
2) RCG;
3) CYGNSS observatory;
4) GPS pseudo random noise (PRN).

The left column of Fig. 9 summarizes the algorithm performances with respect to different incidence angle bins, and the right column shows the performances for different RCG intervals. In particular, Fig. 9(a) and (b) shows the error statistics as bias (cont. lines) and standard deviation of the difference between reference and estimated winds (dashed lines) as a function of reference ECMWF wind; Fig. 9(c) and (d) illustrates the overall bias and standard deviation of the difference, respectively, as a function of incidence angle and RCG, and finally, Fig. 9(e) and (f) shows the pdf of retrieved winds (cont. lines), color coded, respectively, for incidence angle and RCG, and how they are compared to the reference ECMWF winds pdf (dashed black line). The performance is overall stable with respect to both incidence angle and RCG. The wind error statistics show no significant variation with respect to incidence angle [see Fig. 9(a)], and variations with

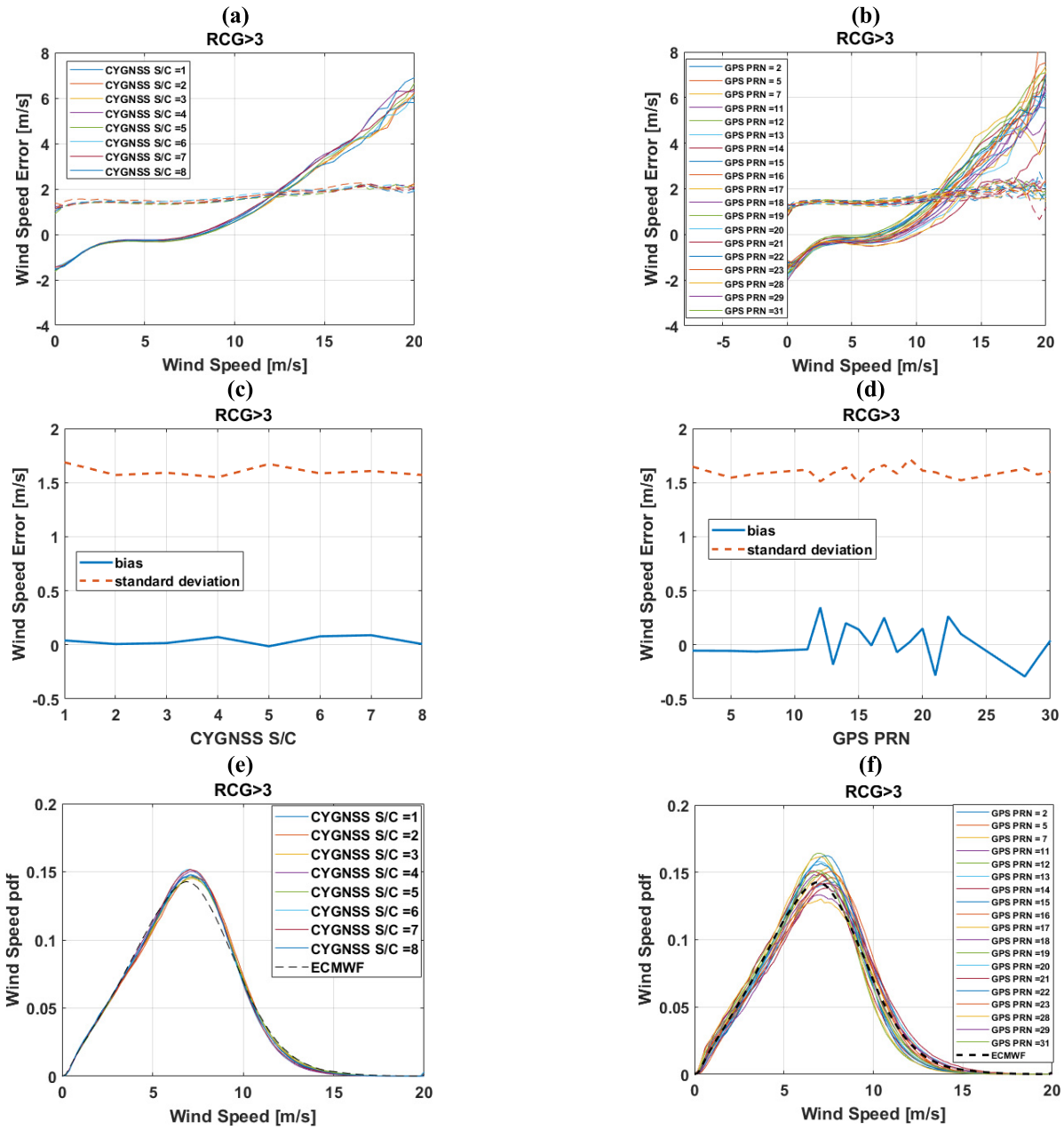

Fig. 10. (Top row) Wind error statistics (bias as cont. line, standard deviation of the wind speed difference as dashed line), as a function of reference winds, color coded by (a) CYGNSS S/C and (b) PRN. (Middle row) Total bias (blue) and standard deviation of the wind speed difference (red) as a function of (c) CYGNSS S/C and (d) PRN. (Bottom row) PDF of retrieved winds, color coded by (e) CYGNSS S/C and (f) PRN, with overlapped pdf of ECMWF winds (dashed black).

respect to the RCG only at the high wind speed end, where lower RCGs generate higher biases [see Fig. 9(b)]. The overall biases are fairly stable with respect to the incidence angle [see Fig. 9(c)]; however, the standard deviation tends to decrease for increasing the incidence angle. This is slightly unexpected and could be the result of calibration imperfections rather than a true reduction in the standard deviation of the retrievals for higher incidence angles. The overall standard deviation also decreases for increasing RCG until it reaches a plateau [see Fig. 9(d)], whereas the bias decreases and reaches zero around RCG $=$ 30, and then it starts to increase again for higher RCG values. This could be once again explained by residual calibration issues, not fully corrected by the L1 calibration. The pdfs of retrieved winds [see Fig. 9(e) and (f)] are sufficiently aligned and close to the reference pdf, so no significant dependence on the pdf is observed with respect to incidence angle or RCG.

Similar to Figs. 9 and 10 summarizes the performances with respect to different CYGNSS observatories (left column) and different GPS PRN (right column). No significant differences across the different CYGNSS spacecraft are detected in Fig. 10(a), 10(c), and 10(e), and it is evident that there is overall good intersatellite calibration. A dependence of the wind retrieval error on the different

antennas (starboard versus port) has been found recently [18]. This dependence is not discussed here, so the retrieval errors shown in Fig. 10(a) represent an average across both antennas. Significant differences can be instead detected for different PRNs, both in terms of error statistics (biases, standard deviations) and in terms of retrieved wind distributions. This is probably a result of the unaccounted fluctuations of the equivalent isotropically radiated power (EIRP) across the different GPS satellites, which ultimately affects the calibration and the retrieval performances. Strategies and plans to tackle the EIRP fluctuations have been proposed [17] and will be implemented in the future releases of CYGNSS data.

Other calibration issues have been highlighted in CYGNSS data, such as a drift by the DDMA over time [18], [19]. This drift is not analyzed here, but it is expected to affect both the FDS winds and the cdf matching winds; hence, it should be addressed in the future, along with other efforts to improve the overall calibration of the data.

V. Validation Against Other Modeled and Measured Winds

In Section III, the CYGNSS retrieved winds obtained from cdf matching were compared to the ECMWF winds, which are the reference data also used to generate the statistics needed by the cdf matching algorithm. In order to consolidate the validation of CYGNSS winds, it is useful to compare more in detail the CYGNSS estimates from cdf matching to FDS winds, and to wind speeds output by other models or measured by other satellites, collocated with the CYGNSS acquisitions. The cdf matching winds are here validated against the following:

1) wind speeds output by the baseline approach, for FDS;
2) wind speeds output by the Global Data Assimilation System (GDAS) model;
3) wind speeds measured by the Advanced SCATterometer (ASCAT) satellites;
4) wind speed measured by the WindSat polarimetric radiometer;
5) wind speed measured by the Advanced Microwave Scanning Radiometer-2 (AMSR-2);
6) wind speeds measured by GPM Microwave Imager (GMI).

Details on the modeled GDAS winds can be found in [20]. For AMSR-2, GMI, and WindSat, the low-frequency (LF) wind speed products are used, as they have been empirically found to have a slightly better correlation with CYGNSS winds. The ASCAT, WindSat, AMSR-2, and GMI data sets are all provided by Remote Sensing Systems (www.remss.com). The collocation criteria are a ±90-min time window and a 25-km distance threshold [13]. Fig. 11 shows winds gridded with a 0.5° resolution in latitude and longitude and averaged over August and September 2017. Fig. 11 (from top to bottom) shows the CYGNSS winds estimated from the cdf matching approach, the CYGNSS FDS winds, the winds output by the ECMWF model, the wind output by the GDAS model, the winds measured by the ASCAT scatterometer, the winds measured by the WindSat radiometer, the winds measured by the AMSR-2 radiometer, and finally the winds measured by the GMI radiometer. CYGNSS successfully reproduces the large-scale wind patterns, although from Fig. 11 the underestimation of the higher winds is visible, when CYGNSS is compared to all the other cases. Despite the fact that the algorithm has been

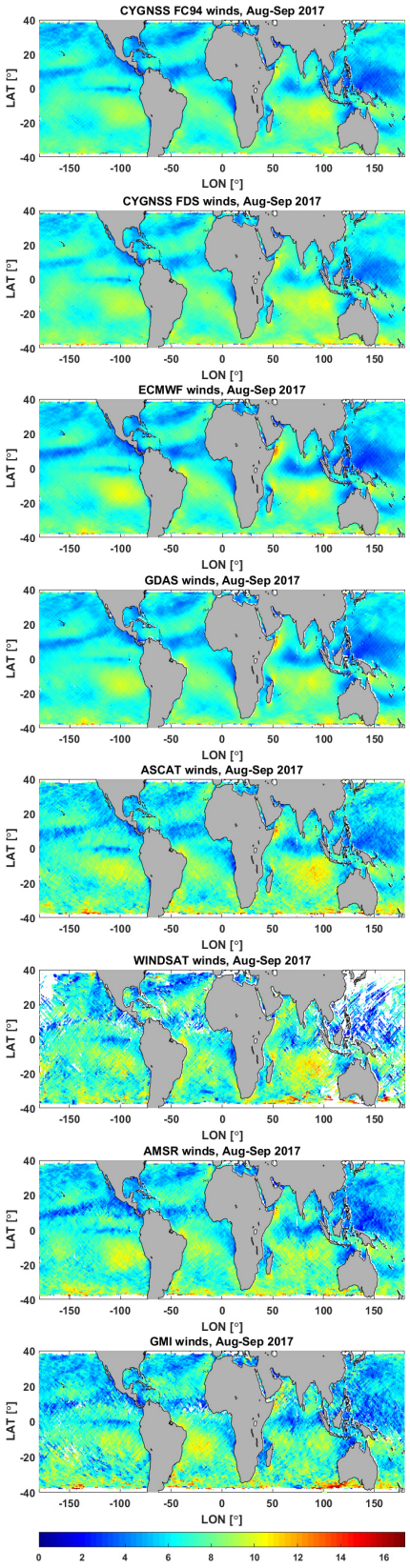

Fig. 11. Gridded winds generated with a resolution of 0.5° latitude and longitude and averaged over two months.

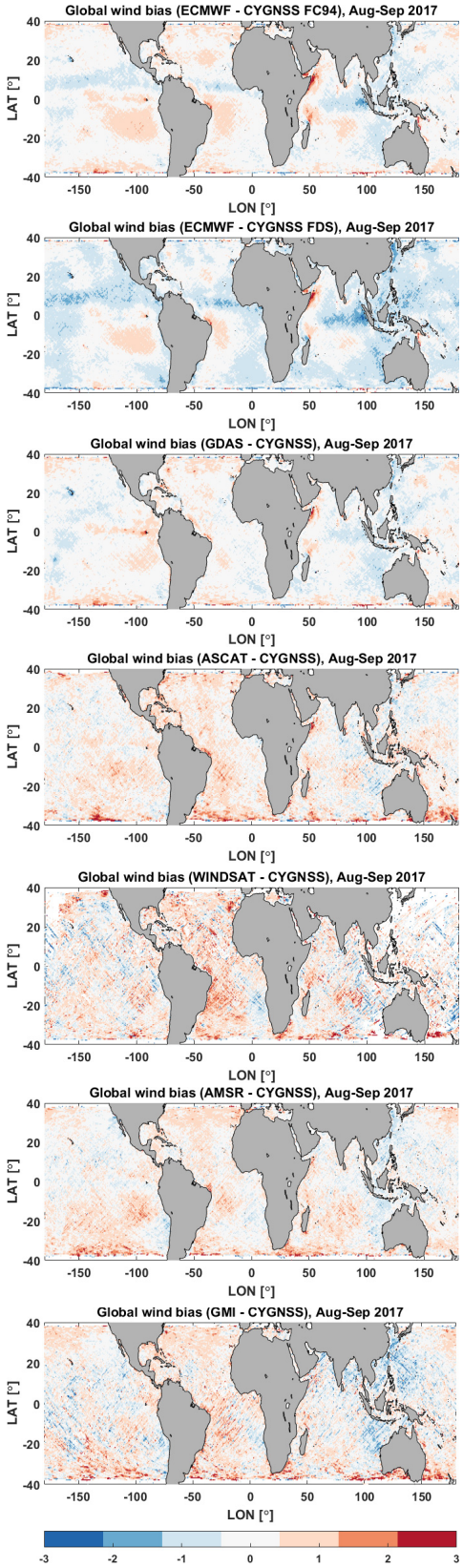

Fig. 12. Gridded winds bias, generated with a resolution of 0.5° latitude and longitude and averaged over two months.

trained using ECMWF reference winds, the CYGNSS winds obtained from cdf matching appear closer to the GDAS winds, which represent the best match among all the modeled or measured winds considered. In Fig. 12, the biases (reference minus CYGNSS winds) computed over the same lat/long grid and the same temporal interval are illustrated. The CYGNSS FDS winds and cdf matching winds appear very similar in the two top panels of Fig. 11; however, a good difference can be visualized in the bias maps with respect to ECMWF, shown as the first two top panels of Fig. 12. CYGNSS winds estimated from the cdf matching approach yield an overall lower bias almost everywhere, compared to FDS winds. The bias for FDS winds is predominantly negative, meaning that the FDS approach tends to overestimate winds on average. Specific regional biases can be seen in both cases, and they tend to coincide with areas of high or low wind speeds. These biases are not mitigated, and in some cases even increased, in the case of the cdf matching approach. This is however expected from a statistical approach like the cdf matching which will work well on average winds, but might produce stronger biases in low or high wind speed regimes. Looking more specifically at the map of biases between ECMWF and CYGNSS with the cdf matching approach, a bias typically positive below the equatorial line and negative above it, in both the Western Atlantic and Eastern Pacific Ocean, can be observed, as well as clusters of positive and negative biases in the Indian Ocean. The subsequent panels of Fig. 12 always use CYGNSS winds from the cdf matching approach for the bias map. The regional biases observed with respect to ECMWF are considerably reduced when CYGNSS is instead compared to GDAS, especially in the Atlantic and Pacific Oceans. The difference between ASCAT and CYGNSS winds produces mostly positive biases, suggesting a stronger underestimation of the wind speeds by CYGNSS, compared to scatterometers. A prevalence of positive biases is also visible in the bias map with respect to WindSat, which is the case with the highest biases overall. The comparison with AMSR-2 yields similar results as for ASCAT, with similar bias patterns. GMI also shows similar positive biases as AMSR-2, but stronger negative biases as well. High positive biases tend to occur at latitudes of around -40, in almost all cases. The total bias, RMSD, and correlation coefficients are calculated for each case and shown in Table II. The lowest bias is achieved with GDAS winds, and the lowest RMSD and highest correlation coefficient is obtained for AMSR-2 winds.

TABLE II

BIAS, RMSD, AND CORRELATION COEFFICIENTS. AS FOR TABLE I, THE CALCULATIONS ARE PERFORMED USING LEVEL 2 WINDS ALONG THE SPECULAR POINT TRACKS, AND NOT GRIDDED WINDS

	GDAS	***ASCAT***	***WindSat***	***AMSR-2***	***GMI***
Bias [m/s]	-0.00	0.29	0.31	0.23	0.22
RMSD [m/s]	1.59	1.56	1.44	1.38	1.40
Correlation Coefficient	0.83	0.85	0.86	0.88	0.87

VI. CONCLUSION

In this article, a statistical methodology to derive wind speeds has been illustrated. The approach stems from the

work in [14] and provides estimations of wind speed from the DDMA and the LES observable using the cdf of each observable and of the wind speed. In principle, the methodology can be extended to any observable, and to derive geophysical parameters other than the wind speed. It has, for example, been considered for the retrieval of the significant wave height and mean square slope of the ocean surface [21], [22]. This cdf matching approach is separately applied to bins of incidence angle and RCG, as this accounts for different measurement geometries and calibration imperfections. The DDMA and LES retrievals obtained using the cdf matching method are then combined into the MV estimator. The MV performance is compared to the baseline MV estimator for FDS, and some important improvements are found in the low to medium wind speed regime. The density plots of reference versus retrieved winds reveal a more symmetric and clustered distribution around the 1:1 line with the cdf matching methodology. The wind overestimations when the reference winds are between 5 and 10 m/s are considerably reduced. The pdf of the winds retrieved using the cdf matching method is very close to that of the reference winds, whereas more discrepancies are observed between the reference pdf and the pdf of the baseline winds. Such discrepancies at the moment impose a further step at the end of the baseline algorithm chain, which is a bias correction to force the final pdf of the retrieved winds to match that of the reference winds. This step is not necessary with the cdf matching method. The other main advantages of the cdf matching method are: 1) the computational simplicity and elimination of the need to empirically tune many parameters, as is done with the baseline approach when deriving its GMF; 2) the elimination of the need to assemble matchups to generate the retrievals; and 3) the provision of unbiased retrievals, with a pdf matching that of reference data, and with slightly improved performances in the low to medium wind speed regime, compared to the baseline retrievals. The overall performances of both the cdf matching winds and the baseline FDS winds are similar in the high wind speed regime, and lower than those provided by the YSLF estimations, which are still the preferable choice of CYGNSS wind estimations for the high wind case.

An investigation into how the algorithm performances vary with respect to incidence angle, RCG, CYGNSS observatory, and GPS PRN has been conducted. The results highlight a dependence of the retrieval error on GPS PRN. This will need to be addressed in the future by measuring and accurately monitoring the GPS EIRP. A slight increase in the retrieval error with increasing RCG and decreasing incidence angle has also been found. Since incidence angle and RCG are in any case related, this could be a further indication of residual calibration errors. A comparison of CYGNSS MV winds retrieved using the cdf matching method has also been carried out with respect to baseline FDS winds, modeled GDAS winds, and winds measured by scatterometers and microwave radiometers. The results show that wind estimates obtained from the cdf matching approach are improved with respect to FDS winds, as global biases are reduced. Despite having been trained on ECMWF winds, the CYGNSS cdf matching winds are in good agreement with GDAS outputs, even more than with ECMWF winds. Some regional biases, which are noticeable in the map of wind bias between CYGNSS and ECMWF, are strongly reduced when the same bias is computed with respect to GDAS winds. There is overall a good agreement between CYGNSS and all the winds considered in the validation exercise, with good capability by CYGNSS to successfully reproduce the large-scale wind patterns. The biases appear reasonably low even when CYGNSS is compared to ASCAT and AMSR-2, whereas they tend to increase with respect to GMI and WindSat.

Future work will concentrate on further exploiting the cdf matching methodology for CYGNSS, which includes the following.

1) The implementation of the methodology on the forthcoming new release of CYGNSS data, with improved calibration and improved knowledge of the GPS EIRP, whose fluctuations currently cause a variation in the estimation performances across the different PRNs.
2) The implementation of the methodology using other observables. The LES observable is noisy, and it does not offer significant improvement when included in the MV combination. Other observables, such as the Trailing Edge Slope (TES) [23] or the Principal Component Analysis (PCA) observable proposed in [24] could be tested and included in the retrieval.
3) The implementation of the cdf matching methodology to estimate wave parameters (i.e., Significant Wave Height or Mean Square Slopes) from CYGNSS.

In addition to the above, a two-parameter implementation of the cdf matching method, where the two algorithm inputs are the observables derived from CYGNSS data as well as ancillary wave information sufficiently decorrelated from the local wind speed, will be explored. The current wind retrieval error is correlated with wave parameters; hence, such an approach should reduce the wave effect and improve the final wind speed retrievals.

ACKNOWLEDGMENT

The authors would like to thank National Oceanic and Atmospheric Administration (NOAA) for providing the wind speed matchups.

REFERENCES

[1] V. U. Zavorotny, S. Gleason, E. Cardellach, and A. Camps, "Tutorial on remote sensing using GNSS bistatic radar of opportunity," *IEEE Geosci. Remote Sens. Mag.*, vol. 2, no. 4, pp. 8–45, Dec. 2014, doi: 10.1109/MGRS.2014.2374220.

[2] J. Garrison, A. Komjathy, V. U. Zavorotny, and S. J. Katzberg, "Wind speed measurement using forward scattered GPS signals," *IEEE Trans. Geosci. Remote Sens.vol.*, 40, no. 1, pp. 50–65, Jan. 2002.

[3] S. J. Katzberg, O. Torres, and G. Ganoe, "Calibration of reflected GPS for tropical storm wind speed retrievals," *Geophys. Res. Lett.*, vol. 33, no. 18, 2006, Art. no. L18602, doi: 10.1029/2006GL026825.

[4] S. Gleason, "Space-based GNSS scatterometry: Ocean wind sensing using an empirically calibrated model," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4853–4863, Sep. 2013, doi: 10.1109/TGRS.2012.2230401.

[5] M. Clarizia and C. S. Ruf, "Wind speed retrieval algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4419–4432, Aug. 2016, doi: 10.1109/TGRS.2016.2541343.

[6] C. S. Ruf and R. Balasubramaniam, "Development of the CYGNSS geophysical model function for wind speed," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 12, no. 1, pp. 66–77, Jan. 2019, doi: 10.1109/JSTARS.2018.2833075.

[7] W. Lin, M. Portabella, G. Foti, A. Stoffelen, C. Gommenginger, and Y. He, "Toward the generation of a wind geophysical model function for spaceborne GNSS-R," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 655–666, Feb. 2019, doi: 10.1109/TGRS.2018.2859191.

[8] C. Ruf *et al.*, "New ocean winds satellite mission to probe hurricanes and tropical convection," *Bull. Amer. Meteorol. Soc.*, vol. 97, no. 3, pp. 385–395, 2016, doi: 10.1175/BAMS-D-14-00218.1.

[9] C. S. Ruf *et al.*, "A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation," *Sci. Rep.*, vol. 8, Jun. 2018, Art. no. 8782, doi: 10.1038/s41598-018-27127-4.

[10] S. Zhang, Z. Pu, D. J. Posselt, and R. Atlas, "Impact of CYGNSS ocean surface wind speeds on numerical simulations of a hurricane in observing system simulation experiments," *J. Atmos. Ocean. Technol.*, to be published, doi: 10.1175/JTECH-D-16-0144.1.

[11] S. M. Leidner, B. Annane, B. McNoldy, R. Hoffman, and R. Atlas, "Variational analysis of simulated ocean surface winds from the Cyclone Global Navigation Satellite System (CYGNSS) and evaluation using a regional OSSE," *J. Atmos. Ocean. Technol.*, to be published, doi: 10.1175/JTECH-D-17-0136.1.

[12] J. Park, J. T. Johnson, Y. Yi, and A. J. O'Brien, "Using 'Rapid Revisit' CYGNSS wind speed measurements to detect convective activity," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 12, no. 1, pp. 98–106, Jan. 2019, doi: 10.1109/JSTARS.2018.2848267.

[13] F. Saïd, S. Soisuvarn, Z. Jelenak, and P. S. Chang, "Performance assessment of simulated CYGNSS measurements in the tropical cyclone environment," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 9, no. 10, pp. 4709–4719, Oct. 2016, doi: 10.1109/JSTARS.2016.2559782.

[14] M. H. Freilich and P. G. Challenor, "A new approach for determining fully empirical altimeter wind speed model functions," *J. Geophys. Res.*, vol. 99, no. C12, pp. 25051–25062, Dec. 1994.

[15] E. Andersson, A. Persson, and I. Tsonevsky, "User guide to ECMWF forecast products," ECMWF, Reading, U.K., Tech. Rep., v1.2, 2015, p. 121. [Online]. Available: https://www.uio.no/studier/emner/matnat/geofag/nedlagte-emner/GEF4220/v18/undervisningsmateriale/ecmwf_user_guide_update_v1.2_20151123.pdf

[16] M. P. Clarizia and V. C. Z. Ruf, *CYGNSS Algorithm Theoretical Basis Document Level 2 Wind Speed Retrieval*, document CYGNSS Project 148-0138, Rev. 5, Aug. 2018.

[17] T. Wang, C. S. Ruf, B. Block, D. S. McKague, and S. Gleason, "Design and performance of a GPS constellation power monitor system for improved CYGNSS L1B calibration," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 12, no. 1, pp. 26–36, Jan. 2019, doi: 10.1109/JSTARS.2018.2867773.

[18] F. Saïd, Z. Jelenak, P. S. Chang, and S. Soisuvarn, "An assessment of CYGNSS normalized bistatic radar cross section calibration," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 12, no. 1, pp. 50–65, Jan. 2019, doi: 10.1109/JSTARS.2018.2849323.

[19] P. S. Chang, Z. Jelenak, F. Said, and S. Soisuvarn, "CYGNSS observations of ocean winds and waves," *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 4285–4288, doi: 10.1109/IGARSS.2018.8517846.

[20] *Global Data As-similation System*, National Centers for Environmental Information, NOAA, Silver Spring, MD, USA, 2018. [Online]. Available: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas

[21] M. P. Clarizia and J. C. Reynolds Ruf, "Retrieval of significant wave height from CYGNSS," in *Proc. ESA Living Planet Symp.*, Milan, Italy, May 2019. [Online]. Available: https://www.lps19.esa.int

[22] M. P. Clarizia, J. Reynolds, and P. Hwang, "Retrieval of mean square slopes from CYGNSS data," in *Proc. GNSS+R workshop Reflectometry GNSS Signals Opportunity*, Benevento, May 2019.

[23] M. P. Clarizia, C. S. Ruf, P. Jales, and C. Gommenginger, "Spaceborne GNSS-R minimum variance wind speed estimator," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6829–6843, Nov. 2014, doi: 10.1109/TGRS.2014.2303831.

[24] N. Rodriguez-Alvarez and J. L. Garrison, "Generalized linear observables for ocean wind retrieval from calibrated GNSS-R delay–Doppler maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 1142–1155, Feb. 2016, doi: 10.1109/TGRS.2015.2475317.