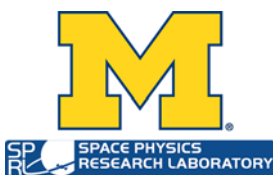# CYCLONE GLOBAL NAVIGATION SATELLITE SYSTEM (CYGNSS)

Algorithm Theoretical Basis Document **Level 2 Wind Speed Retrieval**	UM Doc. No.	148-0138
	SwRI Doc. No.	N/A
	Revision	Rev 5
	Date	17 August 2018
	Contract	NNL13AQ00C

CYCLONE GLOBAL NAVIGATION SATELLITE SYSTEM (CYGNSS)

Algorithm Theoretical Basis Document Level 2 Wind Speed Retrieval	UM Doc. No.	148-0138
	SwRI Doc. No.	N/A
	Revision	Rev 5
	Date	17 August 2018
	Contract	NNL13AQ00C

Prepared by: Maria Paola Clarizia, Valery Zavorotny, Chris Ruf   Date: 08/17/2018

Approved by:

_____   Date: _____
Chris Ruf, CYGNSS Principal Investigator

Approved by:

_____   Date: _____
Jillian Redfern, CYGNSS Project Manager

Approved by:

_____   Date: _____
Tim Butler, CYGNSS SOC Manager

Released by:

_____   Date: _____
Darren McKague, CYGNSS UM Project Manager
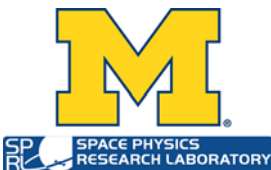
REVISION NOTICE

Document Revision History		
Revision	**Date**	**Changes**
PRE-RELEASE DRAFT	17 June 2013	n/a
INITIAL RELEASE	17 January 2014	L1 calibration and L2a correction added. Delay-Doppler range optimized. Time averaging added. Full 13-day nature run used for algorithm development and performance assessment.
REVISION 1	11 August 2014	L2 Retrieval algorithm modified to account for modifications in the GMF derivation
REVISION 2	03 November 2015	Section on L1b corrections removed, equations (8.8) and (8.13) revised, a few typos corrected. Equation numbering revised from Section 5 to Section 10.
REVISION 2 CHANGE 1	18 November 2015	Equation numbering revised from Section 5 to Section 10.
REVISION 2 CHANGE 2	19 November 2015	Typos corrected. Caption of figure 4 corrected. Documents 17790-ConOp-01 and 17790-SMDP-01 properly referenced;
REVISION 2 CHANGE 3	02 December 2015	Typos corrected. Caption of figure 4 corrected. Documents 17790-ConOp-01 and 17790-SMDP-01 properly referenced;
REVISION 3	20 August 2016	Section 9 updated to reflect new definition of L2 quality flags
REVISION 4	24 January 2017	a) Entire Description of L2 wind speed retrieval algorithm modified to be consistent with algorithm implemented by the SOC. b) Definition of modified effective scattering area introduced, and definition of L1b DDMA and L1b LES updated to be normalised by the modified effective scattering area; c) Description of the original L2 algorithm developed to meet the

		performance requirements has been moved to appendix A. d) Appendix B added to describe all the lookup tables used by the algorithm, and the content of the L1 and L2 NetCDF files e) Format and numbering of Sections, figures and equations, and references made consistent across the whole document
REVISION 5	17 Aug 2018	a) Generate geophysical model function (GMF), to map 10 m referenced ocean wind speed to CYGNSS L1 observables, empirically from matchups between measurements and near-coincident ground reference winds. b) Generate two different GMFs, one for fully developed seas (FDS) and the other for young seas with limited fetch (YSLF). FDS GMF is generated using ECMWF and GDAS numerical prediction model ground reference winds. YSLF GMF is generated using SFMR measurements on hurricane hunter aircraft. c) add debiasing adjustment to final minimum variance FDS wind speed to match global cumulative distribution function to NWP matchup database.

Table of Contents

1   Introduction and Summary

The CYGNSS Project implements a spaceborne earth observation mission designed to collect measurements of ocean surface winds through variations in the direct vs. reflected Global Positioning System (GPS) signals. The observatory portion of this mission consists of a constellation of eight satellites. The CYGNSS mission objective is to provide new information about ocean surface winds in Tropical Cyclones (TC), enabling advances in the knowledge of TC genesis and intensification.

The CYGNSS goal is to understand the coupling between ocean surface properties, moist atmospheric thermodynamics, radiation, and convective dynamics in the inner core of TCs. The goal of CYGNSS directly supports the NASA strategic objective to enable improved predictive capability for weather and extreme weather events. Near-surface winds are major contributors to and indicators of momentum and energy fluxes at the air/sea interface. Understanding the coupling between the surface winds and the moist atmosphere within the TC inner core is key to properly modeling and forecasting its genesis and intensification. Of particular interest is the lack of significant improvement in storm intensity forecasts over the past two decades, relative to forecasts of storm track. Advances in track forecast have resulted in large part from the improvements that have been made in observations and modeling of the mesoscale and synoptic environment surrounding a TC. The CYGNSS team hypothesizes that the lack of an accompanying improvement in intensity forecasting is in part due to a lack of observations and proper modeling of the TC inner core. The inadequacy in observations results from two causes:

1. Much of the inner core ocean surface is obscured from conventional remote sensing instruments by intense precipitation in the eye wall and inner rain bands.

2. The rapidly evolving genesis and intensification stages of the TC life cycle are poorly sampled by conventional polar-orbiting, wide-swath imagers.

CYGNSS addresses these two limitations by combining the all-weather performance of GPS based bistatic scatterometry with the spatial and temporal sampling properties of a constellation of observatories. The constellation consists of individual GPS bistatic radar receivers flown on 8 microsatellites. This provides the ability to measure the ocean surface winds with high temporal resolution and spatial coverage under all precipitating conditions, up to and including those experienced in the hurricane eyewall. The 8 microsatellites are launched on a Deployment Module (DM) that is attached to the NASA government furnished equipment (GFE) launch vehicle.

The baseline CYGNSS instrument is a Delay Doppler Mapping Instrument (DDMI) which resides on each observatory in the constellation. The DDMI is a Global Navigation Satellite System (GNSS) Receiver-Remote sensing Instrument. Each instrument will use two nadir pointing antennas for collecting reflected GPS signals and a zenith facing antenna to collect direct GPS signals. The GPS transmission frequency enables the instrument to make surface scattering observations during most precipitation conditions.

2   Overview and Background

For some years, GPS receivers have been used to provide position, velocity, and time measurements to satellite platforms in low Earth orbit. In a similar way, they are also used for ground-based navigation. Beyond navigation however, GPS signals have been increasingly used for remote sensing. Signals at L-band – with a bandwidth between 2 and 20 MHz – are broadcast globally from an altitude of 20,000 km (~12,427 mi) and are used to measure, amongst other things, tectonic plate motion and ionospheric and tropospheric parameters.

The United Kingdom Disaster Monitoring Constellation (UK-DMC-1) space-based demonstration mission showed that a microsatellite-compatible passive instrument potentially could make valuable geophysical measurements using GPS reflectometry. The left side of the figure below diagrams how the process works. The direct GPS signal is transmitted from the orbiting GPS satellite and received by a right-hand circular polarization (RHCP) receive antenna on the zenith (*i.e.* top) side of the spacecraft that provides a coherent reference for the coded GPS transmit signal. The signal that is scattered back from the ocean surface is received by a downward looking left-hand circular polarization (LHCP) antenna on the nadir side of the spacecraft. The scattered signal contains detailed information about the ocean surface roughness statistics, from which local wind speed can be retrieved.

The image on the right below shows scattering cross section as measured by UK-DMC-1 and demonstrates its ability to resolve the spatial distribution of ocean surface roughness. This type of scattering image is referred to as a Delay Doppler Map (DDM). A DDM exhibit a typical horseshoe-like shape, which is linked to the space-to-DD coordinate transformation and consequent reshape of the spatial scattered power. The delay-Doppler coordinate system is explained in detail in Section 4.3.
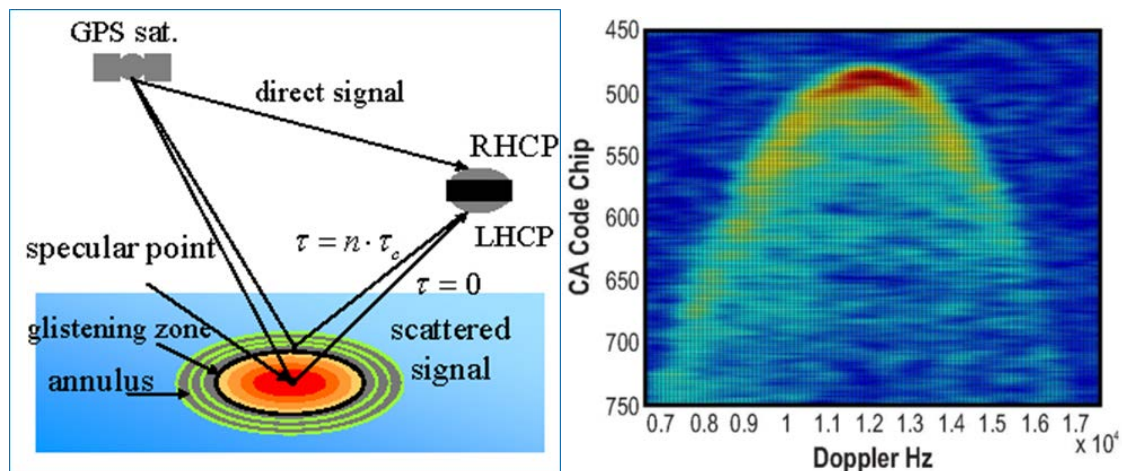
Figure 1. [left] GPS signal propagation and scattering geometries for *ocean surface bistatic quasispecular scatterometry*. The position of the spacecraft is determined from the direct GPS signal; the surface winds are determined by the indirect signal scattered off the ocean surface. Combining the position and scattering information allows for the creation of Delay Doppler Maps (DDM), from which ocean surface wind speeds can be inferred. [right] An example DDM measured

by the UK-DMC-1, showing the spatial distribution of the ocean surface scattering. Scattering cross section is plotted as a function of Doppler Shift (x-axis) and relative propagation time of flight (y-axis), which is measured in units of Coarse Acquisition GPS Code, or "Chips".

The forward scattering is dominated by specular reflection. The reflected signal comes mostly from a point where a specular reflection occurs, called Specular Point (SP), and a variable area around the SP called the glistening zone, where quasi-specular reflections are in place, and where scattered power is redirected towards the receiver. A very calm sea (flat surface) would therefore only produce a strong specular reflection from the SP, whereas a rougher sea causes less power to be scattered from the SP, and more power scattered from the glistening zone, which expands with increasing roughness. The wind and sea surface roughness have therefore an impact on both the distribution of the scattered power, and the effective scattering area.

Stronger wind and sea surface roughness induce two main types of changes in the DDM. One is a decrease in the peak power of the horseshoe shape in the DDM, corresponding to the power at the specular point; the other is the increase in the power along and between the horseshoe branches, which stretch towards larger delays and Doppler frequencies. The horseshoe shaped power pattern in the DDM essentially represents the power scattered in space by the whole glistening zone, and the increase of the power along and between the branches of the horseshoe shape is a sign of increasing size of the glistening zone.

2.1 *Science Objectives and Algorithm Requirements*

The CYGNSS science goals are enabled by meeting the following mission objectives:

- Measure ocean surface wind speed in most naturally occurring precipitating conditions, including those experienced in the tropical cyclone eyewall;
- Measure ocean surface wind speed in the tropical cyclone inner core with sufficient frequency to resolve genesis and rapid intensification.

The CYGNSS baseline science requirements are:

a) The baseline science mission shall provide estimates of ocean surface wind speed over a dynamic range of 3 to 70 m/s as determined by a spatially averaged wind field with resolution of 5x5 km.
b) The baseline science mission shall provide estimates of ocean surface wind speed during precipitation rates up through 100 millimeters per hour as determined by a spatially averaged rain field with resolution of 5x5 km.
c) The baseline science mission shall retrieve ocean surface wind speed with a retrieval uncertainty of 2 m/s or 10%, whichever is greater, with a spatial resolution of 25x25 km.
d) The baseline science mission shall collect space-based measurements of ocean surface wind speed at all times during the science mission with the following temporal and spatial sampling: 1) temporal sampling better than 12 hour mean revisit time; and 2) spatial sampling 70% of all storm tracks between 35 degrees north and 35 degrees south

latitude to be sampled within 24 hours.

e) The CYGNSS project shall conduct a calibration and validation program to verify data delivered meets the requirements in sections 4.1.1a, 4.1.1b, 4.1.1c and 4.1.1d within individual wind speed bins above and below 20 m/s.

f) Support the operational hurricane forecast community assessment of CYGNSS data in retrospective studies of new data sources.

2.2 Measurement Overview

GPS-Reflectometry (GPS-R) exploits pre-existing signals of opportunity from the Global Positioning System (GPS) constellation. It measures the direct GPS signal, received through a zenith antenna, to pin-point the position of the transmitting and receiving satellite, and the reflected GPS signal from the surface of the ocean, through a downward pointing antenna, for retrieval of sea surface wind and roughness. GPS-R is based upon scattering in a bistatic geometry, where the transmitter and receiver are not collocated on the same platform. An illustration of the GPS-R overview and measurement is shown in Figure 2. The transmitting GPS satellites are a constellation of up to 32 Medium- Earth orbit Satellites in operation at any given time. They are in six different orbital planes, and have a near circular orbit with an inclination angle of 55°, an orbital period of about 12 hours, and an altitude of about 20200 km.
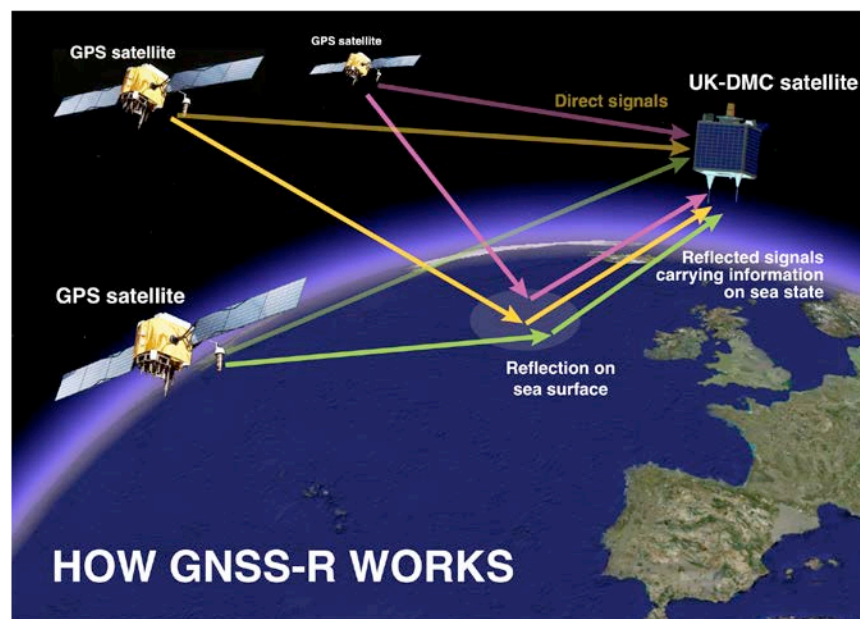
Figure 2. Measurement principle of GPS-Reflectometry over the ocean.

The CYGNSS mission is comprised of 8 microsatellites deployed into a common orbit plane of 35° inclination at ~525 km altitude. Each S/C is able to track up to four SP simultaneously, and generate a 1-second DDM for each SP. This results in 32 wind measurements (32 DDMs) per second across the globe, providing a wind field imagery of TC genesis, intensification and decay

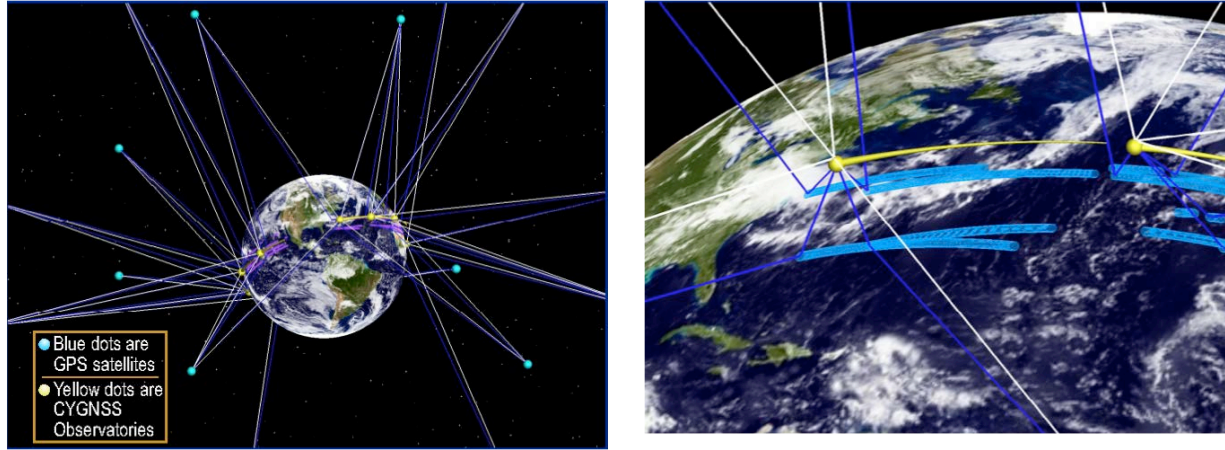with unprecedented spatial and temporal resolution (Figure 3).

Figure 3. (left) overview of GPS Satellites and CYGNSS observatories. (right) illustration of the four simultaneous specular acquisitions for each observatory.

Each CYGNSS observatory is equipped with a Digital Doppler Mapping Instrument (DDMI). The DDMI generates DDMs continuously at a low data rate, which provides a source for ocean roughness measurements across the ocean. In special situations, such as when passing over an active tropical cyclone, the instrument is operated in Raw IF Mode, where ~60 seconds of raw sampled data is accumulated. This allows researchers to fully analyze and re-analyze the acquired data using different processing schemes to ensure that the nominal DDM mode of operation is not losing important geophysical data.

Each DDM pixel is obtained through a cross-correlation of the received scattered GPS signal with a locally generated replica of the C/A code of the transmitted signal, for the pair of delay-Doppler coordinate corresponding to that pixel. Such cross-correlation is done over 1 ms, and it is heavily affected by speckle noise, so that a number of incoherent accumulation of consecutive cross-correlation values is necessary to mitigate the noise. The incoherent accumulation time for DDM measured spaceborne is 1 second. A schematic of the blocks required to process the received signal to obtain a DDM is shown in Figure 4.
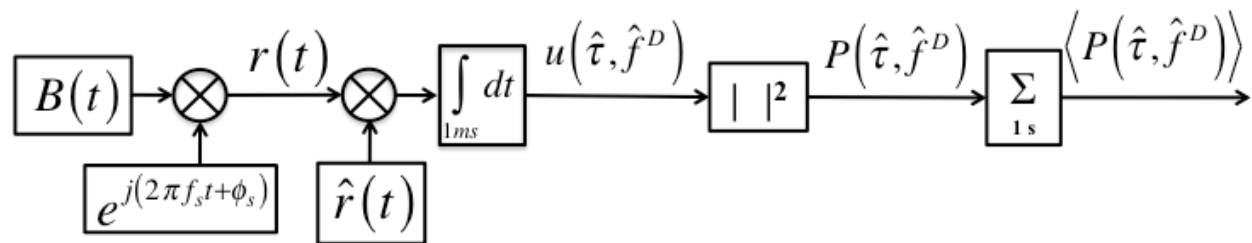

Figure 4. Schematic of a GPS receiver (modified from Gleason et al., 2005).

DDMs constitute the fundamental measurement for spaceborne GNSS-R, to which inversion

algorithms are then applied directly to estimate the sea surface wind and roughness.

2.3   Algorithm Approach

There are three popular approaches for extracting sea surface winds from GPS-R DDMs. The first approach relies on a forward scattering model to simulate DDMs. This model, developed by Zavorotny and Voronovich [2000], and described in details in Section 4, generates the delay-Doppler scattered power as a function of geometrical parameters, receiver/antenna parameters, and sea surface parameters. The latter are so-called Mean-Square Slope (MSS), which represent the sea surface roughness, and are related to the sea surface wind speed through MSS-to-wind relationships [Elfouhaily et al., 1997, Katzberg et al., 2006]. The MSS is therefore retrieved by fitting the simulated DDMs to the measured ones, usually in either a Least-Square (LS) or Maximum likelihood (ML) sense [Germain et al., 2004, Clarizia et al, 2009]. For airborne GPS-Reflectometry, the Doppler spreading is so small that it is usually sufficient to apply such fitting to the delay Waveforms, namely the scattered power as a function of delays, at the Doppler shift of the specular point [Garrison et al., 1998, 2002, Komjathy et al., 2004, Gleason et al., 2005]. In some cases, a matched filter approach has been applied to delay waveforms, rather than a full Least-Square fitting [Katzberg et al., 2000], but the overall results do not change. Generally speaking, this fitting approach can be applied to a full DDM, as well as to waveforms derived from a DDM (i.e. delay waveforms, integrated delay and Doppler waveforms, etc.). Note once again that this approach usually extracts the optimal MSSs, from which the winds are then derived.

The second type of algorithm is based on the use of a so-called DDM observable, and an empirical regression approach. A DDM direct observable or direct descriptor is a quantity derived from the DDM, which varies with respect to changing underlying wind and roughness in the DDM. Examples of DDM observables are the average or volume of DD pixels around the specular point (DDMA or DDMV, [Marchan-Hernandez et al, 2008, 2010]), different types of distances defined within the DDM ([Rodriguez-Alvarez et al., 2013]), the rising edge slope of the waveform obtained as DDM integration along the Doppler frequencies, known as Integrated Delay Waveform, or IDW [Clarizia et al., 2014], and the width of the waveform obtained as DDM integration along the delays, known as Integrated Doppler Waveform, or IDoW [Clarizia et al., 2013] The approach is usually to calculate these quantities from the measured DDMs, regress them against the match-up winds measured locally by some other sources or instruments, and extrapolate the relationship between the observable and the wind through a Look-Up Table (LUT) function. In this case, there is no intermediate step of MSS calculation, since the measured observable is directly converted into a wind estimate.

A third approach is a hybrid algorithm, which combines the two approaches described above [Gleason, 2006]. It extrapolates the Normalized Radar Cross Section (NRCS) at the specular point from the DDMs, through inversion of the theoretical model, and then regresses the NRCS values against the wind measured by match-ups, and converting the NRCS into a wind measurement through the definition of a LUT function.

3 Measurement Description

3.1 Satellite Constellation and Sampling

The satellite constellation configuration and its resulting spatial and temporal sampling characteristics are described in the following CYGNSS Project Engineering Memoranda:

- UM EM No. 148-0132, On the Dependence of Science Coverage on Launch Time
- UM EM No. 148-0133, Specular Point Algorithm for GPS Reflected Signals
- UM EM No. 148-0134, Effect of Satellite Clustering on Science Coverage

3.2 Observatory Platform

The observatory platform design and performance are described in a series of CYGNSS Project Engineering Memorandums and summarized in the following released document of Concept of Operations:

- SwRI EM No. 17790-CONOP-01, Cyclone Global Navigation Satellite System (CYGNSS) Concept of Operations Plan

The software associated with observatory operation and related ground operations is described in the CYGNSS released document:

- SwRI EM No. 17790-SMDP-01, Software Management & Development Plan

3.3 Delay Doppler Mapping Instrument

Each CYGNSS spacecraft carries a Delay Doppler Mapping Instrument (DDMI) capable of locating and tracking GPS signal reflections on the Earth's surface. A DDMI consists of two Earth pointing nadir antennas with low noise amplifiers and internal blackbody calibration target, a single upward (space) pointing zenith antenna with a low noise amplifier and internal blackbody calibration target, and a Delay Mapping Receiver (DMR) electronics unit. These components, as they are mounted on the CYGNSS platform, are illustrated in Figure 5.
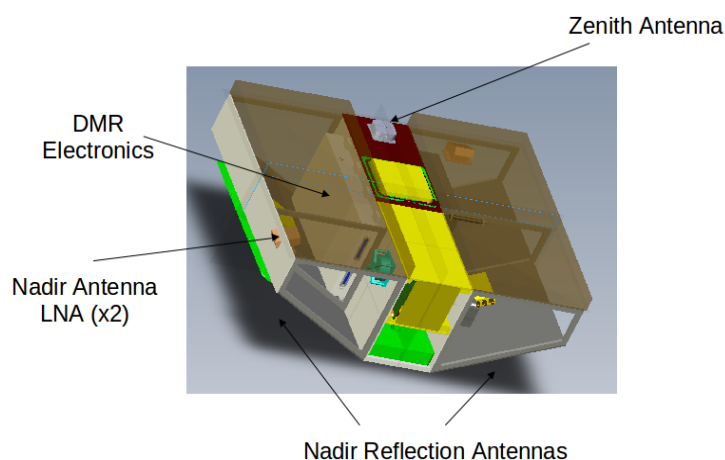

Figure 5. The Delay Doppler Mapping Instrument CYGNSS platform configuration.

The instrument is passive, with the signals being transmitted from the Global Positioning System (GPS) constellation. The instrument autonomously tracks and cross correlates the reflected signal power over a range of delay and Doppler bins. The DDMI outputs 4 DDMs every second to the S/C, which are compressed, sent to the ground, and calibrated into the Level 1 data products.

3.3.1   Zenith (Direct Signal) Antenna

The DDMI zenith antenna and processing channels serve two key functions. The first is to generate a navigation solution for the CYGNSS spacecraft, providing real-time position, velocity and time (PVT) information at 1 Hz. This is performed by tracking 4 or more direct GPS signal transmissions, making pseudorange measurements from each and generating an estimate of the receiver PVT data. The second function is to provide a direct signal power measurement for all of the tracked GPS satellites used in the navigation solution. This second function is an additional requirement on the navigation receiver, in that traditional satellite navigation usually does not place much emphasis on the strength of the tracked signal and rarely uses the received power for anything other than a general indicator of signal quality. However, in the case of GNSS-R the absolute power of the received signals is of critical importance. This presents an interesting problem when using GNSS constellations (including GPS): The transmit power of the GPS signal (and associated transmit antenna gain) is unpublished and only approximately known.

The solution to this problem is to use the received direct signal levels to map the effective isotropic radiated power (EIRP) of each GPS satellite. This is made possible because the direct signal does not undergo any surface scattering which makes the power relatively straightforward to estimate using the basic radar equation. With the positions of the CYGNSS receiver and transmitting GPS satellite known it is possible to accurately estimate the propagation path loss. Additionally, the zenith antenna gain pattern must be calibrated as a function of the azimuth and elevation of the incoming signal. Using these parameters and a calibrate signal to noise level of the tracked direct signals, the EIRP of the GPS satellite can be estimated as a function of geometry. This function, generated with numerous direct signal to noise measurements over a range of viewing angles, can then be applied to the reflected signal geometry to provide an estimate of the GPS EIRP in the reflected signal delay Doppler map. This value is needed as an input into the Level 1b calibration calculation of the bistatic radar cross section, which is one of the basic surface observable used for wind retrieval.

3.3.2   Nadir (Reflected Signal) Antennas

The DDMI is interfaced to two downward looking nadir antennas, which together result in the surface footprint shown in Figure 6.
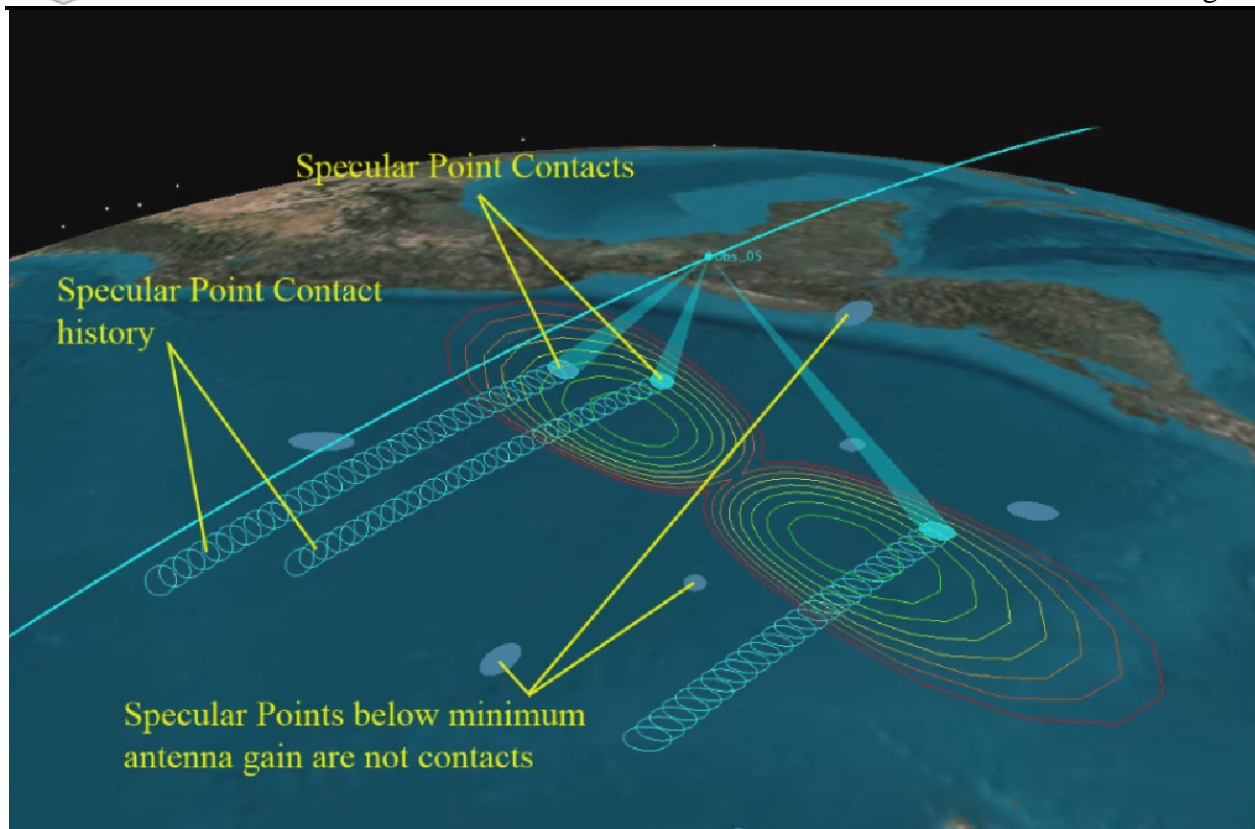
Figure 6. CYGNSS nadir antenna surface coverage footprint. The outermost (red) gain contour represents the threshold for acceptable received signal strength, inside of which scattered signals can be converted to near-surface wind speed with uncertainties that meet or exceed the Level 1 mission requirement.

Referring to Figs. 5-6, the nadir antennas are mounted on the spacecraft to project their patterns onto the surface. This high signal to noise coverage pattern in the cross track direction acts to capture signal reflections to the starboard (right) and port (left) sides of the satellite direction of motion.

3.3.3   Front End Receiver and Backend Digital Processor

The complete DDMI processing chain is shown in Figure 7, which includes both receiver RF front end components, on-board digital receiver back-end stages and ground based Level 1 data product processing steps.

Figure 7. DDMI signal processing chain. Front end receiver, consisting of RF to IF and digitization stages, and the Back-end Digital Processor constitute the stages of DDM generation.

Upon acquisition by one of the nadir antennas, the signal goes through several stages of RF processing before it is sampled for the digital processor. Following, the Level 0 DDM product generated by the DMR is then calibrated in several stages on the ground at the science operations center. Referring to Figure 7, the hardware and processing stages of the DDMI processing chain are summarized below.
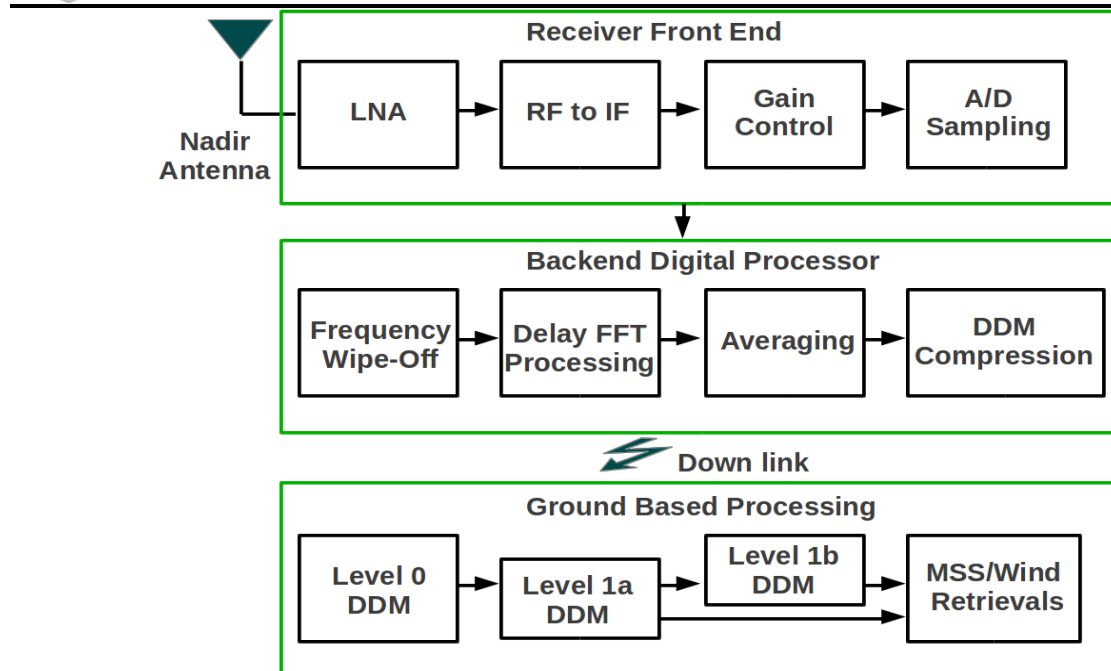
1. The signal is acquired by one of the two Nadir antennas.
2. The signal travels over a short cable and enters the low noise amplifier. This initial amplification stage largely determines the overall receiver noise figure.
3. The signal then travels to the DMR electronics where it is down converted and filtered in several stages by the RF front-end, transforming it from the L1 GPS frequency to an Intermediate Frequency (IF) suitable for digital processing.
4. The signal is then amplified by a gain control. The purpose of this last stage of amplification is to center the signal level into the optimal active range of the Analog to Digital (A2D) convertor.
5. The signal is then 2-bit sampled over 1 millisecond by the A2D convertor for digital processing by the DMR firmware.
6. The first stage of digital processing is the application of a digital carrier wipe-off technique to the incoming sample vector. This consists of generating sine and cosine sampling vectors at each Doppler frequency bin and mixing them with the incoming sampled signal.
7. The signal is then correlated with the unique GPS signal spreading code, isolating the surface reflection to the GPS satellite it originated from. The delay processing is

implemented across the entire range of time samples using FFT based techniques. A multiplication of the GPS code and the carrier wiped off signal in the frequency domain results in a 1ms correlation at every delay in the time domain after an Inverse FFT is performed. This process is repeated at every frequency bin in the DDM and results in a single look 1ms DDM.

8. 1000 single look DDMs are averaged over 1 second to produce the non-coherently summed DDM sent to the spacecraft. This averaging is performed to reduce the speckle and thermal noise present in the 1ms DDM looks.

9. The 1 second DDM's are then compressed on the spacecraft and sent to the ground. The DDM compression algorithm is documented in detail in [Ruf, 2013]. The compressed DDMs are then transmitted to the ground for processing at the Science Operations Center at the University of Michigan.

On the ground the raw Level 0 data is unpacked, and the Level 1a and Level 1b calibration algorithms are applied. The ocean mean square slope and wind retrievals can use either of the Level 1a or Level 1b data products for ocean roughness or wind speed estimation.

3.4   Delay Doppler Map Level 0 Data

The DDMI outputs (up to) 4 delay Doppler maps in uncalibrated counts to the spacecraft every second. These DDMs are be converted into the calibrated level 1 data products. An illustration of a raw Level 0 DDM and a Level 1a DDM calibrated to signal power units of Watts are shown in Figure 8. The algorithms for converting Level 0 DDMs to Level 1 DDMs is described in detail in the Level 1 Calibration ATBD.
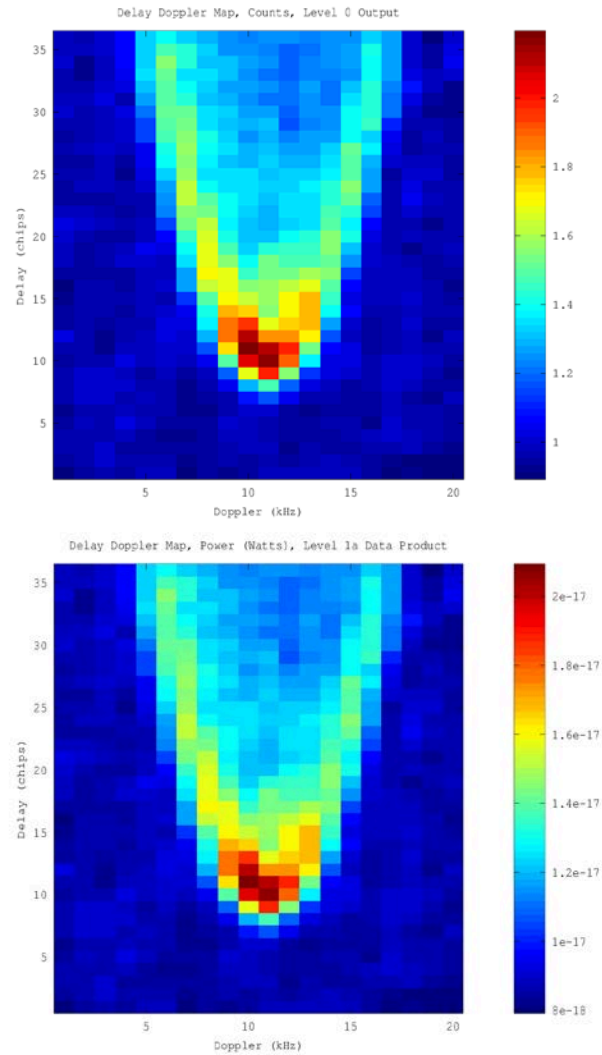
Figure 8. (top) raw Level 0 DDM in uncalibrated counts. (bottom) Level 1a calibrated DDM in units of Watts. DDMs generated by the CYGNSS End-to-End Simulator. Note the different magnitude scales before and after the Level 1a calibration.

The Level 0 raw data output by the DDMI to the spacecraft is a result of several levels of analog RF and digital processing steps. The resulting Level 0 DDM is in units of processed counts. This DDM is then compressed as detailed in (reference: DDM compression algorithm) and sent to the ground for calibration into the Level 1a and Level 1b data products.

3.5   Delay Doppler Map Level 1 Data Products

Both Level 1 data products are provided as a two dimensional DDM grid of floating point numbers. The content and units of each of the Level 1a and Level 1b products is listed below.

3.5.1 Level 1a Data Product and Calibration Approach

The Level 1a data product consists of a DDM calibrated to the units of Watts (in 1 second). Every DDM produced by the delay mapping receiver is calibrated to Level 1a product power values. There are three types of calibration environments and scenarios, each with specific input signal requirements. Below is a list of each scenario and the signals needed;

1. Pre-launch laboratory calibration: Black body load source. GNSS Signal simulator.
2. Regular on-orbit calibration: Black body load source. Estimate of antenna noise temperature over the open ocean.
3. Beacon on-orbit calibration: Ground based GNSS signal generator.

Pre-launch calibration in the lab was performed using a variety of known input noise temperatures. With two different input noise values it will be possible to perform initial estimates of the instrument calibration coefficients prior to launch.

The on-orbit approach is detailed in the Level 1 Calibration ATBD and involves making measurements with the instrument switched to the black body load source as well as over well modeled ocean noise temperatures.

The beacon calibration occurs occasionally and involves the reception of a beacon generated DDM signal, which can be used independently to re-validate the calibration coefficients used to generate the Level 1a product.

The Level 1a calibrated delay Doppler maps is provided as a data product to users as well as fed into the Level 1b algorithm for conversion to BRCS values.

3.5.2 Level 1b Data Product and Calibration Approach

The Level 1b data product consists of a calibrated DDM map of bistatic radar cross section values in units of dB.

The Level 1b calibration is performed after the Level 1a calibration and uses an extensive set of external meta-data to convert the Level 1a power in Watts to a DDM map of BRCS values. This conversion is done for every DDM and requires the following external information:

1. Absolute power, delay and Doppler of the direct signal. The GPS satellite processed in the DDM is tracked by the CYGNSS spacecraft navigation receiver, which provides an estimated of the GPS satellite signal to noise (which can then be converted to absolute power), the tracked delay code phase and Doppler.
2. The CYGNSS satellite GPS time, position and velocity in the Earth Centered Earth fixed (ECEF) reference frame and the receiver clock error terms. This is provided by the CYGNSS navigation receiver.
3. The GPS satellite position and velocity in the Earth Centered Earth fixed (ECEF) reference frame and clock error terms. This is downloaded from the International GNSS

service (IGS) daily.

4. Detailed knowledge of the gain pattern and orientation of the CYGNSS spacecraft nadir and zenith antennas.

Information which is not directly provided as part of the Spacecraft telemetry downlink or external public GNSS services is calculated on the ground by the Science Operations Center (SOC) at the University of Michigan and includes,

1. An accurate geolocation of the specular reflection point in the Earth Centered Earth fixed (ECEF) reference frame.
2. The GPS satellite transmit power.
3. The GPS satellite antenna gain at the observation reflection geometry.
4. The path distances between the GPS satellite and specular point and between the specular point and the CYGNSS receiving spacecraft.
5. The CYGNSS satellite antenna gain, calculated from the reflection geometry (as an azimuth and indecent angle) and the detailed knowledge of the antenna patterns.
6. The path distance between the GPS satellite and the CYGNSS spacecraft. For use in estimating the GPS satellite transmit power.
7. The effective scattering area on the surface of each delay/Doppler bin. Calculated as a function of the reflection geometry using the CYGNSS end-to-end simulation (E2ES).

The above parameters are then combined as described below to convert the Level 1a DDM power values into estimates of the bistatic radar cross section (BRCS) using the derived forward model. This algorithm is described in detail in the Level 1 Calibration ATBD.

4   Forward Model

4.1   Propagation

CYGNSS uses the GPS L1 frequency (1575 MHz) which exhibits negligible rain attenuation, even under heavy precipitating conditions. Nonetheless, the forward model accurately for rain attenuation, G_{rain}, using the formula

$$G_{rain} = \exp(-\alpha h(\csc\theta_t + \csc\theta_r)) \tag{4.1}$$

where h is the freezing height in km, α is the specific attenuation (dB/km), and θ_t and θ_r are the elevation angles to the transmitter and receiver, respectively. Note that all of these parameters will vary over the ocean surface, and this spatial variation is included in our modeling. For simplicity, the current rain attenuation model assumes that the rain rate is constant from the surface up to freezing height.

The specific attenuation α is obtained from the ITU R838-3 model

$$\alpha = aR^b \tag{4.2}$$

where R is the rain rate (mm/hr) and the coefficients a and b for circular polarization at the GPS L1 frequency are $a = 24.312 \times 10^{-5}$ and $b = 0.9567$. In the model, the values for the coefficients have been developed from curve-fitting to power-law coefficients derived from scattering calculations.

Figure 9 shows a plot of rain attenuation versus rain rate for a freezing height of 6 km. In the figure, each curve corresponds to a different elevation angle (the elevation angle to the receiver and transmitter are assumed equal, as would be the case at the specular point).
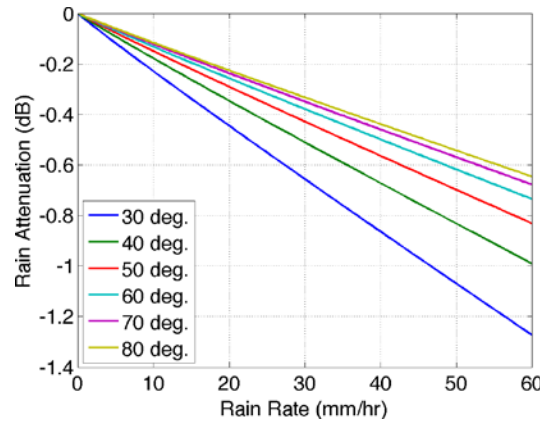
Figure 9. Rain attenuation versus rain rate for various elevation angles.

It should be noted that, in simulated hurricane wind fields, extremely large rain rates have been observed; however these convective cells are highly localized, move very rapidly, and evolve very rapidly. While they can have a very high rain rate, it may only be for a few minutes. So the peak rain rates are very high, but only occur for brief periods in fast-moving cells. Furthermore,

the nature of the GPS ambiguity function causes the rain field to be effectively smoothed over approximately a 20 km area (in the same way wind fields are smoothed), so that the effects of small regions of highest rain attenuation are effectively reduced.

4.2   Rough Surface Scattering

4.2.1   Introduction

In GNSS bistatic radar, the complex amplitude of the received signal (the voltage) is cross correlated with a replica of emitted signal over a coherent integration time, T_i. Frequently, this procedure is called a match-filter signal processing. The end result of this type of coherent signal processing is an ability to form a synthetic footprint which would ultimately determine the shape of the recorded 1D waveform, or 2D delay-Doppler map, and the spatial resolution of the GNSS bistatic radar. For every epoch t_0 the code cross-correlation relative to the received signal u taken at a variety of delays, τ, can be expressed as the integral [Parkinson et al, 1996, Zavorotny and Voronovich, 2000]

$$Y\left(t_0, \tau, f_c\right) = \int_0^{T_i} a\left(t_0 + t'\right) u\left(t_0 + t' + \tau\right) \exp\left(2\pi i f_c t'\right) dt' \qquad (4.3)$$

Here T_i is the coherent integration time, and $a(t)$ is the replica of the PRN code sequence taking values of $\{-1, +1\}$ on a time duration τ_c. The coherent integration time T_i should be comparable or smaller than the coherence time τ_{cor} of the scattered field at the receiver point in order to perform the convolution procedure (4.1) with linear phase shift between replica $a(t)$ and signal $u(t)$. The oscillating factor containing f_c is meant to compensate for a possible Doppler shift of the signal $u(t)$ associated with this phenomenon. For signals received from spacecraft, the signal coherence time τ_{cor}, has been observed to be on the order of 1ms [Gleason et al., 2005, Gleason, 2006], while signals received from aircraft can remain coherent for considerably longer durations, on the order of 5-10 ms depending on aircraft speed and altitude.

Only scattered waves with equal time delays and equal Doppler shifts could be successfully aligned with the code replica in order to produce maximum correlation according to (4.3), and it always happens within a so called glistening zone caused by a random distribution of the surface slopes. The size of the glistening zone is driven by the variance of surface slopes, where the larger the variance of surface slopes the larger the glistening zone extends across the surface.

The scattering toward the receiver is produced mostly by specular reflections from a statistical ensemble of large-scale (larger than several radio wavelengths) slopes of the surface. Therefore, the strongest scattered signal comes only from the center of the glistening zone near the nominal specular point on the mean sea surface. Away from the glistening zone, the contribution from the quasi-specular reflections diminishes, eventually to be replaced with significantly weaker diffraction scattering from a small-scale surface component. Here we neglect this type of scattering as it is too weak to make a significant contribution to the total received signal power. Using this understanding for the physical scattering mechanism, we can apply a Kirchhoff theoretical model to estimate the expected scattering behavior [Bass and Fuks, 1979]. In essence, we are combining a multitude of "smooth" reflection surfaces together to represent the signal

scattering from a rough ocean surface, where every point on the surface is approximated with a local tangent plane.

The scattered GNSS signal $u(t)$ arriving at the receiver position $\vec{R}_r$ can then be modeled by the integral taken over the mean sea surface [Zavorotny and Voronovich, 2000]:

$$u(\vec{R}_r, t) = \int D(\vec{\rho}) a[t - (R_0(t) + R(t))/c] g(\vec{\rho}, t) d^2\rho \tag{4.4}$$

Where $D(\vec{\rho})$ is the amplitude footprint of the receiver antenna; $a(t)$ is the GNSS signal PRN code; $R_0(t)$ and $R(t)$ are distances to the transmitter and the receiver, respectively, to some point $(\vec{\rho}, z = \zeta(\vec{\rho}, t))$ on the "smoothed" rough sea surface with an elevation of $\zeta(\vec{\rho}, t)$, fluctuating about the mean surface level. Over the individual local tangent planes the Earth's curvature is neglected; $\vec{\rho} = (x, y)$; the transmitter and receiver positions are in the $x = 0$ plane, and z is a vertical axis or local surface normal.

The above analysis applies to the scattering of signals from surface components with spatial scales of several wavelengths greater than the incident carrier wavelength (i.e. the GPS L1 wavelength is ~19cm). Alternatively, a contribution to scattering from surface components with spatial scales smaller than several radio wavelengths can be calculated separately using the perturbation theory. Additionally, serious limitations occur for scattering at low grazing angles and from very rough surfaces. In this case, more sophisticated scattering models that take into account multiple scattering and diffraction effects due to sharp edges are required. On the other hand, when surfaces are very even and flat, such as lakes and seas under low wind conditions, or first-year, young ice, the coherent component rises in the scattered GNSS signal. Writing an expression for the received coherent component is a rather simple task.

In the Kirchhoff approximation, the function g describes propagation and scattering processes:

$$g(\vec{\rho}, t) = -\Re(\vec{\rho}) q^2 \exp\left[ik(R_0(t) + R(t))\right] / 4\pi i R_0 R q_z \tag{4.5}$$

where $\Re$ is the Fresnel reflection coefficient; $\vec{q} = k(\vec{n} - \vec{m})$ is the so-called scattering vector, where $k = 2\pi / \lambda$ is a radio wave number; $\vec{m}$ is the unit vector of the incident wave; and, $\vec{n}$ is the unit vector of the scattered wave. Upon substituting (4.5) into (4.4), and then into (4.3), and assuming that integration over the accumulation time T_a is equivalent to averaging over a statistical ensemble of surface elevations

$$\left\langle \left| Y(\tau, f) \right|^2 \right\rangle = \frac{1}{T_a} \int_0^{T_a} \left| Y(t_0, \tau, f) \right|^2 dt_0 \tag{4.6}$$

after making some additional assumptions, we arrive at the bistatic radar equation for the delay-Doppler map [Zavorotny and Voronovich, 2000]

$$\left\langle \left| Y_s(t_0, \tau, f) \right|^2 \right\rangle = T_i^2 \frac{P_T G_T \lambda^2 G_R}{(4\pi)^3} \iint F(\vec{\rho}) \Lambda^2(\tau, \vec{\rho}) \left| S(f, \vec{\rho}) \right|^2 R_0^{-2} R^{-2} \sigma_0(\vec{\rho}) d^2 \rho. \tag{4.7}$$

where P_T is the transmitter power; G_T is the transmit antenna gain; G_R is the receive antenna gain; $F(\vec{\rho})$ is the normalized directivity (beam) pattern for the receive antenna; $\Lambda^2(\tau, \vec{\rho})$ is the annulus function due to the cross-correlation with the replica; $\left| S(f, \vec{\rho}) \right|^2$ is the Doppler zone function due to the relative motions of both the transmitter and receiver with respect to the scattering surface; R_0, R are distances from a point $\vec{\rho}$ on the surface to the transmitter and receiver, respectively; $\sigma_0(\vec{\rho})$ is a bistatic radar cross section (BRCS) of the rough surface, and generally it is a function of two angles, the incidence angle and the scattering angle. Here, in (4.7) it is written as a function of surface coordinates. The scattered signal comes from the area formed by intersection of the equi-range zones, annular (function $\Lambda^2(\tau, \vec{\rho})$) and equi-Doppler, hyperbolic zones (function $\left| S(f, \vec{\rho}) \right|^2$). The width of the equi-range zone depends on the code length (different for C/A code and P code) and on all geometric parameters of the problem. The width of the Doppler zone depends on the receiver velocity and it is the inverse of the coherent integration time, $f_{Dop} = 2/T_i$.

The product of correlation function $\Lambda(\tau)$ and the Doppler zone function constitutes a Woodward Ambiguity Function (WAF) originally introduced in radar technique. The WAF that enters (4.7) is similar to the WAF used in the unfocused SAR technique [Elachi, 1988]. For fixed positions of the transmitter and the receiver both WAF and BRCS are functions of reference surface S coordinates. Looking at (4.7) it is straightforward to conclude that the delay-Doppler map emerges as a convolution of the WAF with BRCS function σ_0. The WAF is close to unity within an area formed by the annulus zone and the Doppler zone, and tends to zero outside this area. The geometry of these zones for two different elevation angles for a typical spacecraft receiver are shown in Figure 10a and b.

(a) **(b)**
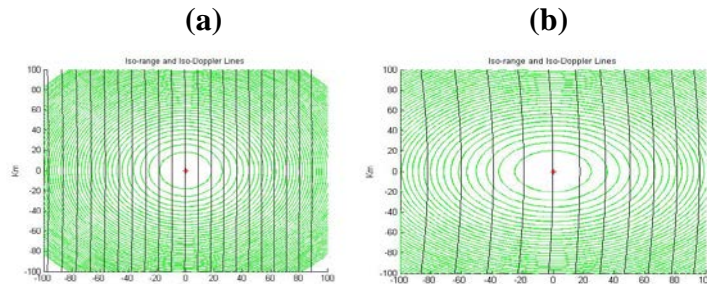
4.2.2 The Bistatic R

Figure 10. Geometries for two different elevation angles for a typical spacecraft receiver

The effect of surface roughness is described by σ_0, the normalized bistatic radar cross section

In the geometric-optics limit of the Kirchhoff approximation the BRCS function σ_0 is represented by the following expression [Bass and Fuks, 1979, Barrick, 1968]:

$$\sigma_0 = \pi |\Re|^2 \left(q / q_z \right)^4 P\left(-q_\perp / q_z \right) \tag{4.8}$$

Though this value is a function of the scattering vector, $\vec{q}$, for fixed positions of the transmitter and the receiver above a surface, this vector can be regarded as a function of the coordinate $\vec{\rho}$ in the mean surface plane. The value of σ_0 depends on a complex Fresnel coefficient $\Re$ which in turn depends on a signal polarization state, a complex dielectric constant of the reflecting medium, ε, and the local incidence angle. In the case of the GNSS the polarization state of the reflected signal is a left-hand circular polarization (LHCP). In this case, the Fresnel reflection coefficient $\Re$ for sea water is [Zavorotny and Voronovich, 2000]:

$$\Re = \frac{1}{2}\left[\frac{\varepsilon \cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon \cos\theta + \sqrt{\varepsilon - \sin^2\theta}} - \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} \right] \tag{4.9}$$

where ε is the complex dielectric permittivity of sea water, and θ is the local incidence angle.

According to Klein and Swift model [Kleain and Swift, 1977]:

at S = 35 ppt and T = 10 deg C ε = 74.62+i51.92 for L1 = 1.57542 GHz; ε = 75.02+i62.39 for L2 = 1.22760 GHz; at S = 30; T = 10 deg C ε = 76.16+i55.30 for L1; ε = 75.02+i62.39: for L2.

Factor $P(\vec{s})$ in (4.8) is the probability density function (PDF) of large-scale "smoothed" surface slopes $\vec{s} = \nabla_\perp \zeta(\vec{\rho})$. Usually, the most probable orientation of surface slopes is parallel to the mean plane, $z = 0$. Then, the PDF has a maximum at $s = 0$, and the bistatic cross-section σ_0 has a maximum at $\vec{q}_\perp = 0$, i.e., at the nominal specular direction with respect to the mean surface. Note that the width of σ_0 in terms of ρ describes a glistening zone produced by quasi-specular points on the surface.

Some GNSS reflection receivers have the capability to sample the waveform only with respect to time delay, τ, while the frequency offset f is fixed and intended to compensate the Doppler shift associated with the nominal specular point on the Earth's surface. In this case, we deal with 1-D delay waveforms, as shown in Figure 11 (a)-(d). The leading edge of such waveforms up to the peak value is produced by the central elliptic annulus zone (filtered by the S function) when it expands from zero to its maximal value. The 1-D waveform forms a decreasing trailing edge after the peak because of the WAF behavior over time lags, and/or of the BRCS recession along

radial directions according to the distribution of surface slopes. Because of the latter reason, the specific shape of the leading edge and an exact position of the correlation power peak is a function of surface roughness. For rougher surfaces, the leading edge is more stretched and the peak is more shifted toward later time lags.

Equation (4.7) deals with values obtained by averaging over a limited number of independent samples. Such values themselves contain residual noise, which might affect our ability to accurately measure the average waveform. The issue of noise in waveforms and their impact on the accuracy of remote sensing of ocean wind is addressed in Section 4.4. Equation (4.7) relies on the condition that $T_i < \tau_{cor}$. The coherence time can be estimated as $\tau_{cor} = \rho_{coh}/v_r$, where ρ_{coh} is the coherence length of the scattered field at the reception point, and v_r is the velocity of the receiver. According to Van-Cittert-Zernike theorem, ρ_{coh} in the far zone increases linearly with the distance from the instantaneous footprint patch on a scattering surface. The size of the footprint patch, or in our case, an annulus zone, depends on the current time delay between the replica and the reflected signal. Therefore, a computation of the coherence time becomes a non-trivial problem which was addressed in [Zuffada and Zavorotny, 2001, You et al., 2004, Soulat, 2004].

The strength of the bistatically-scattered signal from the ocean surface is mostly affected by the surface roughness since variations in salinity of the ocean is rather small. It is believed that for linear surface gravity waves the slope PDF $P(\vec{s})$ can be approximated by the anisotropic bivariate Gaussian distribution [Zavorotny and Voronovich, 2000, Elfouhaily et al, 2002, Soulat, 2004]:
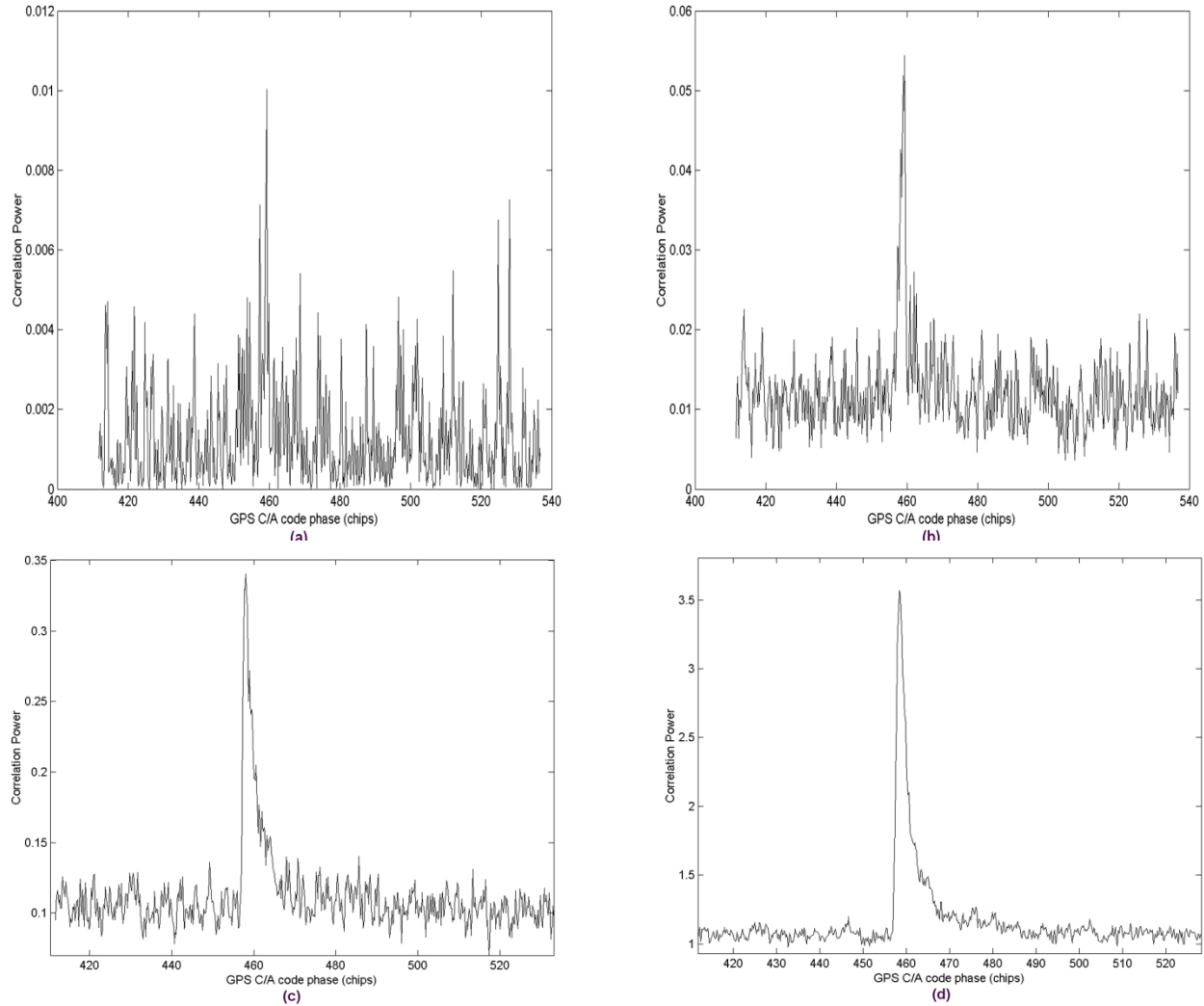
Figure 11. Examples of 1-D delay waveforms [from Gleason, 2006]

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{\det(M)}} \exp\left[-\frac{1}{2} \begin{pmatrix} s_x \\ s_y \end{pmatrix} M^{-1} \begin{pmatrix} s_x \\ s_y \end{pmatrix} \right] \tag{4.10}$$

where matrix M is

$$M = \begin{pmatrix} \cos\varphi_0 & -\sin\varphi_0 \\ \sin\varphi_0 & \cos\varphi_0 \end{pmatrix} \cdot \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix} \cdot \begin{pmatrix} \cos\varphi_0 & \sin\varphi_0 \\ -\sin\varphi_0 & \cos\varphi_0 \end{pmatrix} \tag{4.11}$$

where φ_0 is the angle between the up-down wind direction and x axis, which is chosen here to lie within the incidence plane; σ_u^2 is an upwind mean-square slope (upwind mss); σ_c^2 is a cross-

wind mean-square slope (cross-wind mss). $\sigma_{u,c}^2$ are wind-dependent and can be derived from a surface elevation spectrum $\Psi(\vec{\kappa})$ by integration over wave numbers κ smaller than a scale-dividing wave number κ_*. Sometimes, matrix M is called a directional mean-square slope in contrast to total mss which is defined as $2\sigma_u\sigma_c$.

When wind is directed along one of two axes (4.10) can be re-written in more common fashion:

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{mss_x mss_y(1-b_{x,y}^2)}}\exp\left[-\frac{1}{2(1-b_{x,y}^2)}\left(\frac{s_x^2}{mss_x} - 2b_{x,y}\frac{s_x s_y}{\sqrt{mss_x mss_y}} + \frac{s_y^2}{mss_y}\right)\right] \quad (4.12)$$

where mss_x and mss_y are mean-square slopes of the sea surface for two orthogonal components; $b_{x,y}$ is the correlation coefficient between two slope components:

$$mss_{x,y} = \left\langle s_{x,y}^2\right\rangle = \iint_{\kappa<\kappa_*} \kappa_{x,y}^2 \Psi(\vec{\kappa})d^2\kappa \quad (4.13)$$

$$b_{x,y} = \left\langle s_x s_y\right\rangle / \sqrt{mss_x mss_y} \quad (4.14)$$

$$\left\langle s_x s_y\right\rangle = \iint_{\kappa<\kappa_*} \kappa_x \kappa_y \Psi(\vec{\kappa})d^2\kappa \quad (4.15)$$

One advantage of a Gaussian distribution is that the variance of slopes in (4.13) can be derived solely from a wave spectrum $\Psi(\vec{\kappa})$, of full surface elevations by integrating it over wave numbers, κ, which are smaller than a dividing parameter, κ_*.

There are some indications that the actual PDF of slopes does not exactly follow a Gaussian shape at their tails [Cardellach and Rius, 2008]. In terms of the glistening zone, it implies that this departure affects a periphery of the zone. This would translate into some discrepancy for the value of the waveform, at relatively large time delays, τ, and large frequency offsets, f. An ability to discern the difference caused by the departure from the Gaussian PDF of slopes depends on residual noise of measurements for the peripheral area of the DDM.

One of the most popular models for the spectrum $\Psi(\vec{\kappa})$ is the model proposed by [Elfouhaily et al., 1997]. The integrand in (4.13) is called a slope spectral density. An example of Elfouhaily et al. slope spectrum taken along the wind direction is shown in Figure 12.
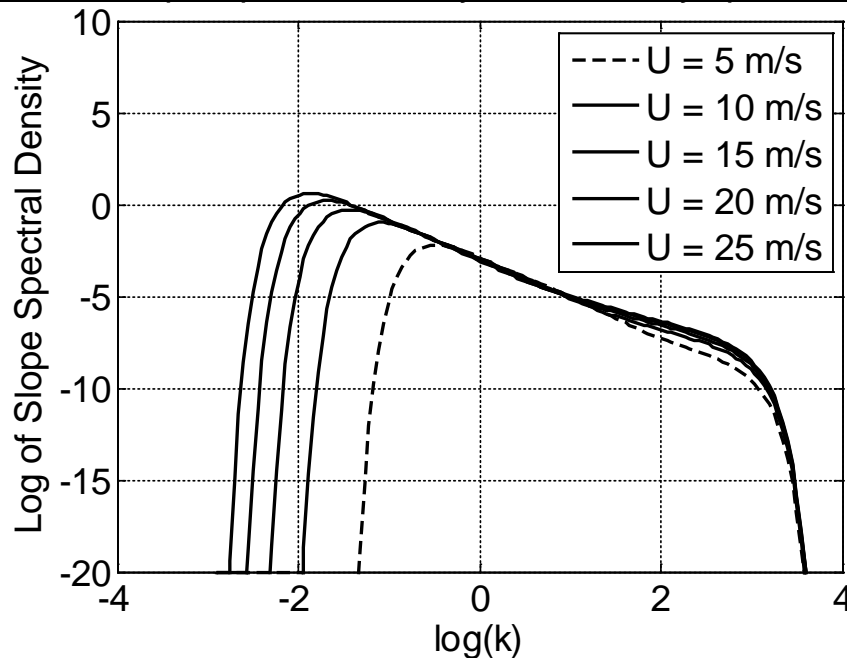
Figure 12. Example of Elfoihaily et al., slope spectrum

This model describes wind-driven waves in deep water under diverse wave age (often called 'fetch') conditions and agrees with the *in situ* observations of the first sun-glint derived wave slope measurements of [Cox and Munk, 1954], performed several decades ago. According to the Elfouhaily et al. model, an elevation spectrum of well-developed wind-driven sea surface can be represented as a product of the radial, or omnidirectional, part of the spectrum, and the azimuthal part of the spectrum. The azimuthal part of the spectrum reproduces two main features of the directional spectrum: its anisotropy, or directionality, and the wavenumber dependence of the angular spectral width. The azimuthal part of the spectrum is a two-sided function; it does not distinguish between up- and down-wind directions. There are other situations when wind direction does not coincide with the maximum of the spectrum, e.g., when gravity waves undergo refraction on currents or on bathymetry, or waves generated by a local wind are superimposed with a swell, or waves generated under the hurricane conditions. Such complicated scenarios are not described by Elfouhaily et al. spectrum.

As it was pointed out above, the mean-square slopes that determine the BRCS through the PDF of slopes are not full wave slopes. Even though, the sea surface contains wave harmonic components both larger and shorter than the L-band electromagnetic waves the short waves can be disregarded in a process of forward quasi-specular reflection under the geometric optics approximation adopted in (4.8). Therefore, the full surface spectrum should be cut off at high end of wave numbers. There are various choices of cutoff wave number κ_*. For example, there exists a "three-lambda" heuristic criterion for κ_* proposed by [Brown, 1978] based on fitting modeled curves for microwave back scattering cross sections with cross sections obtained in

experiments with satellite radar altimeters. The same criterion was initially applied for use of the Kirchhoff approximation for the two-scale calculations of the bistatic cross sections [Zavorotny and Voronovich, 2000]. Later on, a reasonable $\kappa_* = \kappa \cos\theta / 3$ on the incidence angle θ was assumed in [Garrison et al., 2002]. In [Thompson et al., 2005], an expression for κ_* is obtained which contains also a dependence on wind speed, $\kappa_* = k \cos\theta \left(1 + U_{10} / 20\right) / 7.5$. It was obtained by fitting modeled curves for GNSS bistatic scattering cross sections with cross sections obtained in that particular aircraft experiment.

An alternative approach is to obtain an empirical model for mss of slopes against wind speed by performing multiple measurements of GNSS waveforms under controlled wind conditions. The best fit between measured waveforms and modeled ones using (4.8) for various mss values will give the sought dependence mss vs wind speed. This approach was adopted in [Katzberg et al., 2006]. The empirical model from [Katzberg et al., 2006] gives the following expression:

$$
\begin{aligned}
mss_{\square} &= 0.45 \cdot (0.00 + 0.00316) f(U) \\
mss_{\perp} &= 0.45 \cdot (0.03 + 0.00192) f(U)
\end{aligned}
\tag{4.16}
$$

where

$$
f(U) = \begin{cases}
U & 0.00 < U < 3.49 \\
6 \cdot \ln(U) - 4.0 & 3.49 < U < 46 \\
0.411 \cdot U & 46.0 > U
\end{cases}
\tag{4.17}
$$

Wind speed U here is m/s and measured at 10-m height. The extension of $f(U)$ beyond $U = 46$ m/s proposed in [Katzberg et al., 2006] was rather arbitrary because GPS reflection data were not available for such high winds.

Below in Figure 13 we present comparisons between mss calculated using all three approaches: two based on the Elfouhaily et al spectrum with two different cutoff numbers κ_* from [Garrison et al., 2002] and [Thompson et al., 2005], and the empirical one from [Katzberg et al., 2006].
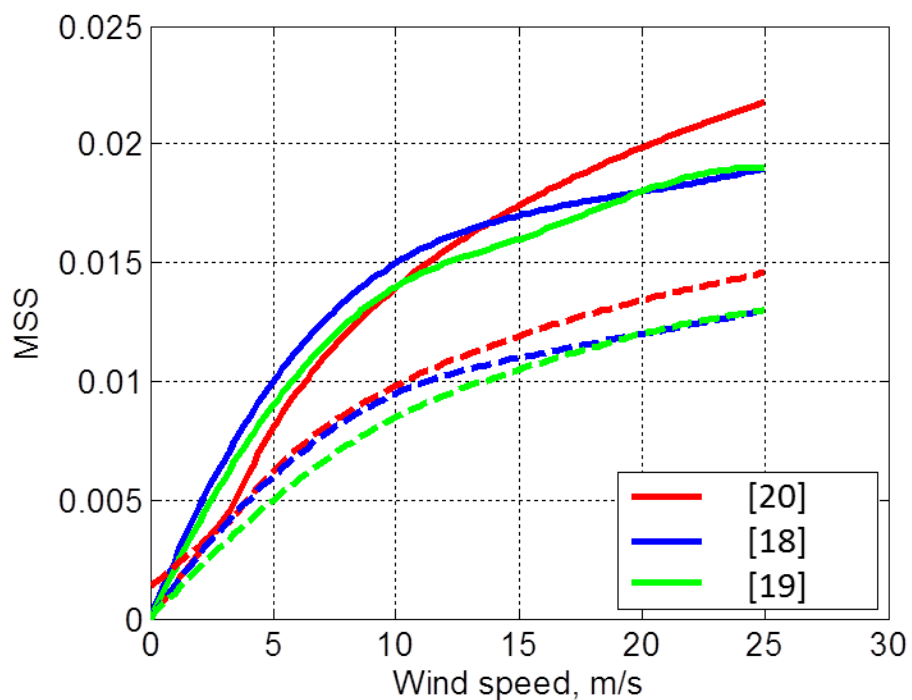
Figure 13. comparison between mss calculated using the approaches in [Garrison et al., 2002], [Thompson et al., 2005], and the empirical one from [Katzberg et al., 2006]. They are respectively indicated as [18], [19] and [20].
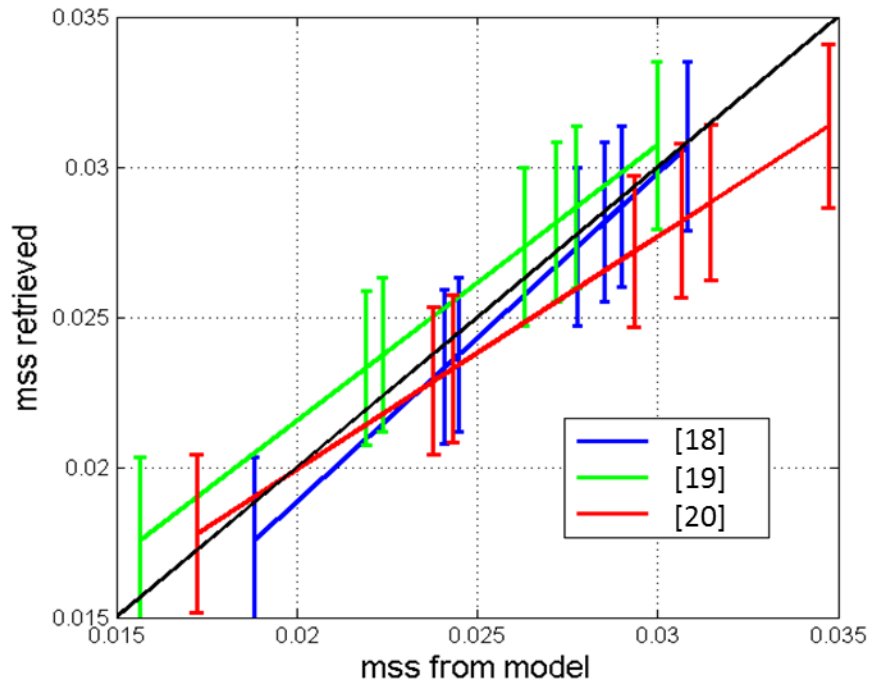
Figure 14. comparison between modeled mss and mss retrieved from DDM measurements during aircraft experiments, using the approaches in [Garrison et al., 2002], [Thompson et al., 2005], and the empirical one from [Katzberg et al., 2006]. They are respectively indicated as [18], [19] and [20].

Figure 13 demonstrates a comparison between three GO models for mss in the up-down wind direction (solid curves) and in the cross-wind direction (dashed curves) for the range of winds between 0 and 25 m/s. Some disagreement between them is seen but overall it is not significant. Figure 14 shows a comparison between three modeled mss and mss retrieved from DDM measurements during aircraft experiments [Rodriguez-Alvarez et al, 2013, Valencia et al., 2014].

In order to make a choice between these three models we performed calculations of σ_0 using a more accurate (than the GO) approximation, a so called small slope approximation which does not require use of spectral dividing parameter κ_*. This material is presented below.

4.2.2 Integrated Scattering Model: The Bistatic Radar Cross Section in Small Slope Approximation

The small slope approximation (SSA) was developed earlier in [Voronovich, 1994, Voronovich, 1999] and was used successfully for solving various scattering and radiometric problems (see, e.g., [Voronovich and Zavorotny, 2001, Elfouhaily and Guerin, 2004, Johnson, 2005, Johnson and Elfouhaily, 2007, Soriano and Guerin, 2008, Arnold-Bos, 2007, 2007b , Guerin et al., 2010, Voronovich and Zavorotny, 2014]).   The geometry of the scattering problem is shown in Figure 15. Three typical scenarios are depicted, although the model considered below allows any possible combination of incident, scattering, and azimuthal angles and arbitrary polarization

states. There are known two approximations of the SSA, the SSA of the 1ˢᵗ order and the SSA of the 2ⁿᵈ order. The latter is more accurate than the former, and is required for solving backscattering problems with shorter EM wavelengths such as the X- and K-band. Practice shows that for the L-band and for the forward scattering regime it is suffice to use the SSA of the 1ˢᵗ order, or SSA1.

Note that the expression for the scattering amplitude in SSA1 coincides with the expression for scattering amplitude in the Kirchhoff approximation (KA) to the accuracy of the pre-integral factor. The major difference, however, is that KA gives a correct answer only for the roughness
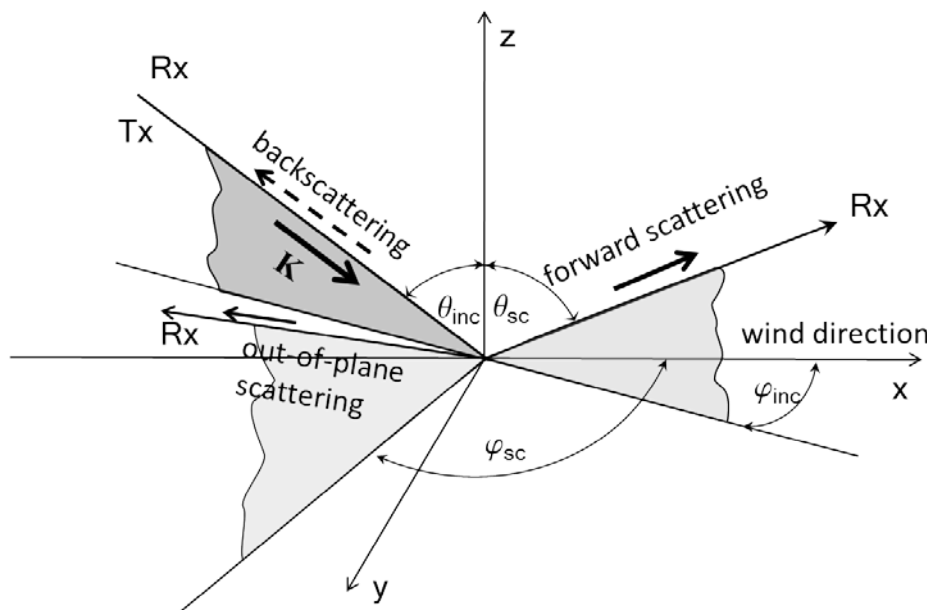
Figure 15. Geometry of the scattering problem for SSA.

$h(\vec{r})$ which is smooth on the wavelength scale; in this case, the corresponding integral can be evaluated by the stationary phase method, thus leading to the geometric optics (GO) approximation. The difference between the GO and KA approximations most likely exceeds the accuracy of the KA itself. In contrast to the KA, SSA1 allows $h(\vec{r})$ to contain a component with a horizontal scale comparable (or even less) than the wavelength, provided that the slope remains small. In this case the corresponding integral also describes the Bragg scattering process and cannot be calculated by the stationary phase method.

For the case of a large Rayleigh parameter, when the contribution from the average-field-related terms can be neglected, the SSA1 gives the following expression for the bistatic radar cross section [Johnson, 2005]:

$$\sigma_{\alpha\beta,\alpha'\beta'}\left(\vec{k},\vec{k}_0\right) = \frac{4q_k^2 q_0^2}{\pi\left(q_k+q_0\right)^2} B_{\alpha\beta}\left(\vec{k},\vec{k}_0\right) \bar{B}_{\alpha'\beta'}^*\left(\vec{k},\vec{k}_0\right) \int_{r<r_{max}} \exp\left[-i\left(\vec{k}-\vec{k}_0\right)\vec{r} - \left(q_k+q_0\right)^2\left(C(0)-C(\vec{r})\right)\right]d\vec{r} \tag{4.18}$$

where r_{max} determines the area significant for integration. $\alpha,\beta,=1,2$ and $\alpha',\beta'=1,2$ are linear polarization indices for incident and scattering waves, respectively. Function $B_{\alpha,\beta}\left(\vec{k},\vec{k}_0\right)$ in (4.18) is a 2x2 matrix representing polarizations (1 stands for vertical and 2 stands for horizontal linear polarization), respectively; they depend on the scattering geometry and dielectric constant of the medium. Expressions for them can be found in [Voronovich and Zavorotny, 2001]. Expressions for the LHCP bistatic scattering cross section can be expressed through corresponding cross sections for linear polarization as follows [Fung et al., 2001, Zuffada et al., 2004]:

$$\sigma_{RL} = \frac{1}{4}\left\{\sigma_{11,11} + \sigma_{22,22} + \sigma_{12,12} + \sigma_{21,21} + 2\left[-\operatorname{Re}\sigma_{11,22} + \operatorname{Re}\sigma_{12,21} - \operatorname{Im}\left(\sigma_{11,12} + \sigma_{11,21} + \sigma_{12,22} + \sigma_{21,22}\right)\right]\right\} \tag{4.19}$$

We performed calculations of BRCS using (4.18) and (4.19) and compared it with corresponding BRCS based on the above described GO models for a typical CYGNSS setting and for a range of incidence angles and winds. These results are discussed below.

4.2.3 4.2.4. BRCS as a function of the incidence angle and wind speed: comparisons between three models

Here, we present comparisons between the SSA results and results obtained with the GO model, one using an MSS based on the Elfouhaily spectrum and the cutoff frequency from [Garrison et al., 2002], and another one using the empirical MSS model from [Katzberg et al., 2006]. For short we will call these two GO models "VZ model" and "SK model," respectively. First, we present plots showing the corresponding LHCP BRCS σ_0 in a forward, specular direction as a function of the zenith scattering angle (which in this case equals to the incidence angle) for a range of wind speeds from 4 to 30 m/s. The results for the VZ, KS and SSA models are shown in Figure 16a, 16b and 17, respectively. Each plot has twelve curves. The top curve on each plot corresponds to wind speed U = 4 m/s. The rest of the curves correspond to 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30 m/s consecutively.
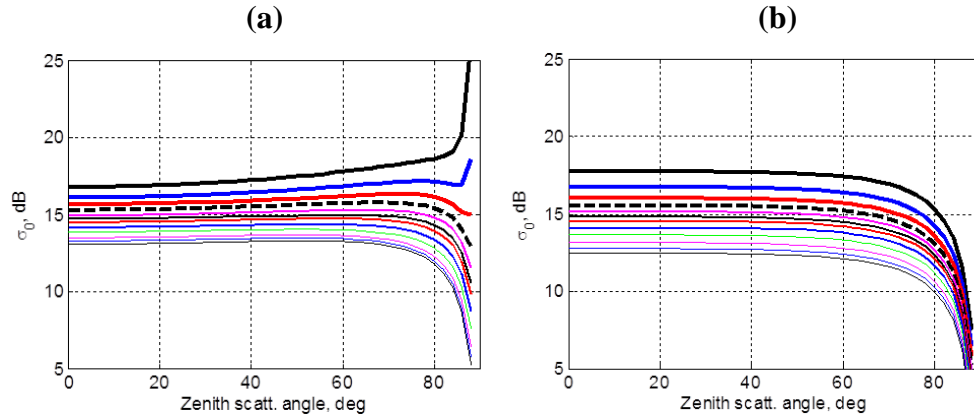
(a)   **(b)**

Figure 16. LHCP BRCS σ_0 in a forward, specular direction as a function of the zenith scattering angle for a range of wind speeds from 4 to 30 m/s, for the VZ (a) and KS (b) model.
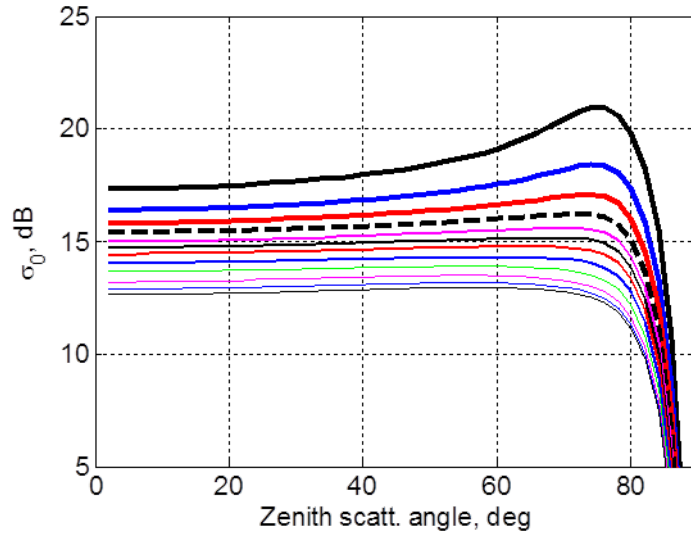

Figure 17. LHCP BRCS σ_0 in a forward, specular direction as a function of the zenith scattering angle for a range of wind speeds from 4 to 30 m/s, for the SSA model.

This dependence on wind speed reflects that fact that increased wind produces a stronger surface roughness which, in its turn, decreases scattering in a specular direction. One can see that σ_0 behaves differently for every of these models at scattering angles larger than 60°-70°. Remember, that any of those models are valid at large scattering angles so we can disregard this discrepancy. For the case of CYGNSS antenna pointing angle of about 30° this discrepancy is not relevant. Note only that the SK model was built on GPS reflection data obtained for low incidence/scattering angles, < 45°, therefore, it might not reflect the actual behavior of the scattering at larger angles. At the same time, all three models demonstrate a quite similar behavior over wind speeds for angles below 45°.

To investigate this behavior in more detail, we will plot the wind dependence of σ_0 for a set of small scattering angles and for a fixed moderate incidence angle below 45°. We would like to check how predictions for σ_0 from all three models correspond to each other for scattering originated from various point on the surface area limited to some number of delay zones that contributes to the Delay-Doppler Map. The corresponding scattering geometry is shown in Figure 18.
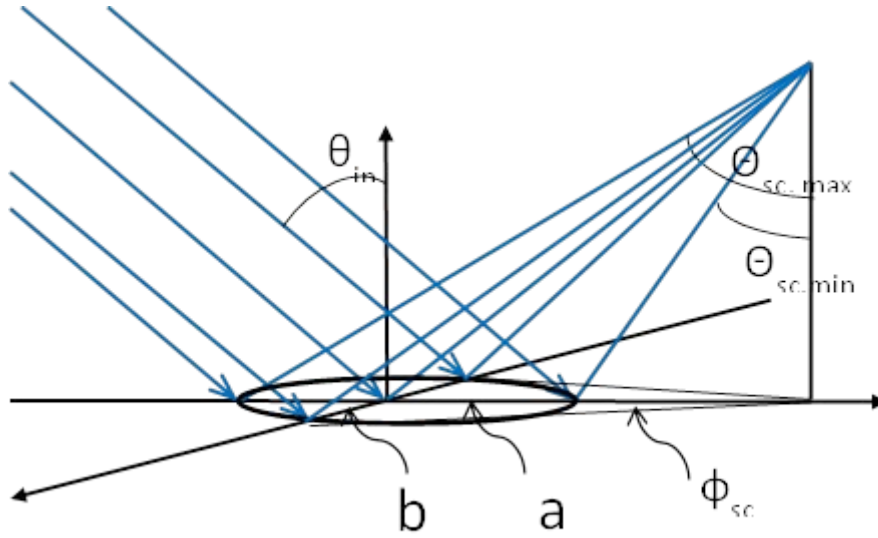
Figure 18. Scattering geometry for a fixed moderate incidence angle below 45°.

Notations in Figure 18 are as follows. a_n and b_n are major semi-axes of the elliptic delay zones where index n correspond to a . They can be expressed through the chip length l, receiver altitude H, and incidence angle θ_{in}, as: $a_n = b_n / \cos\theta_{in}, \quad b_n = \left(2nlH / \cos\theta_{in}\right)^{1/2}$.

Figure 19 shows how angles $\theta_{sc,\min}, \theta_{sc,\max}$ and ϕ_{sc} from Figure 18 can be related to the corresponding points on the delay-zone ellipse for a range of delay-zone index (from 0 to 10). The curves are plotted for $\theta_{in} = 30°$, $H = 600$ km, and $l = 300$ m (C/A code).

Figure 19. relationship between angles $\theta_{sc,\min}$, $\theta_{sc,\max}$ and ϕ_{sc} from Figure 18 and corresponding points on the delay-zone ellipse for a range of delay-zone index from 0 to 10.

The data tip in this figure shows that for the first delay zone the maximum azimuth scattering angle is equal to 2.56°.
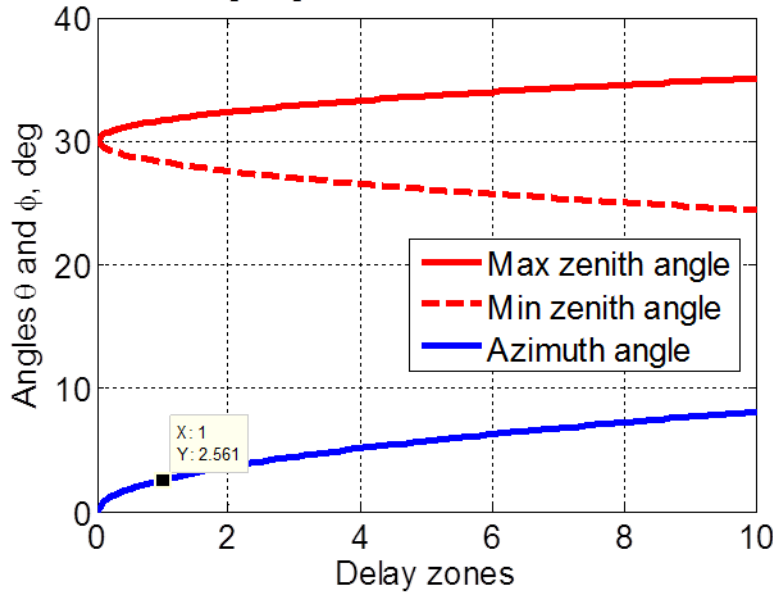
Below in Figure 20(a-d) we present plots of σ_0 obtained with the three models for $\theta_{in} = 30°$, $H = 600$ km, and for four directions of scattering vector described by following combinations of zenith and azimuth scattering angles: (a) $\theta_{sc} = 30°$, $\phi_{sc} = 0°$; (b) $\theta_{sc} = 30°$, $\phi_{sc} = 2.56°$; (c) $\theta_{sc} = 28°$, $\phi_{sc} = 0°$; and (d) $\theta_{sc} = 32°$, $\phi_{sc} = 0°$. This set of angles gives an angular extent for the first delay zone ($n = 1$). Case (a) describes a nominal specular direction originated from the center of the delay zone. Case (b) describes a scattering direction originated from both left and right most distant (in a cross direction) points of the first delay-zone ellipse. This is an example of out-of-plane scattering. Here, by "plane" we mean a specular plane which by definition passes through the specular point on the surface and both transmitter and the receiver points. Case (c) describes a scattering direction originated from the closest point on the first delay-zone ellipse. Correspondingly, case (d) is for the farthest point on the first delay-zone ellipse.
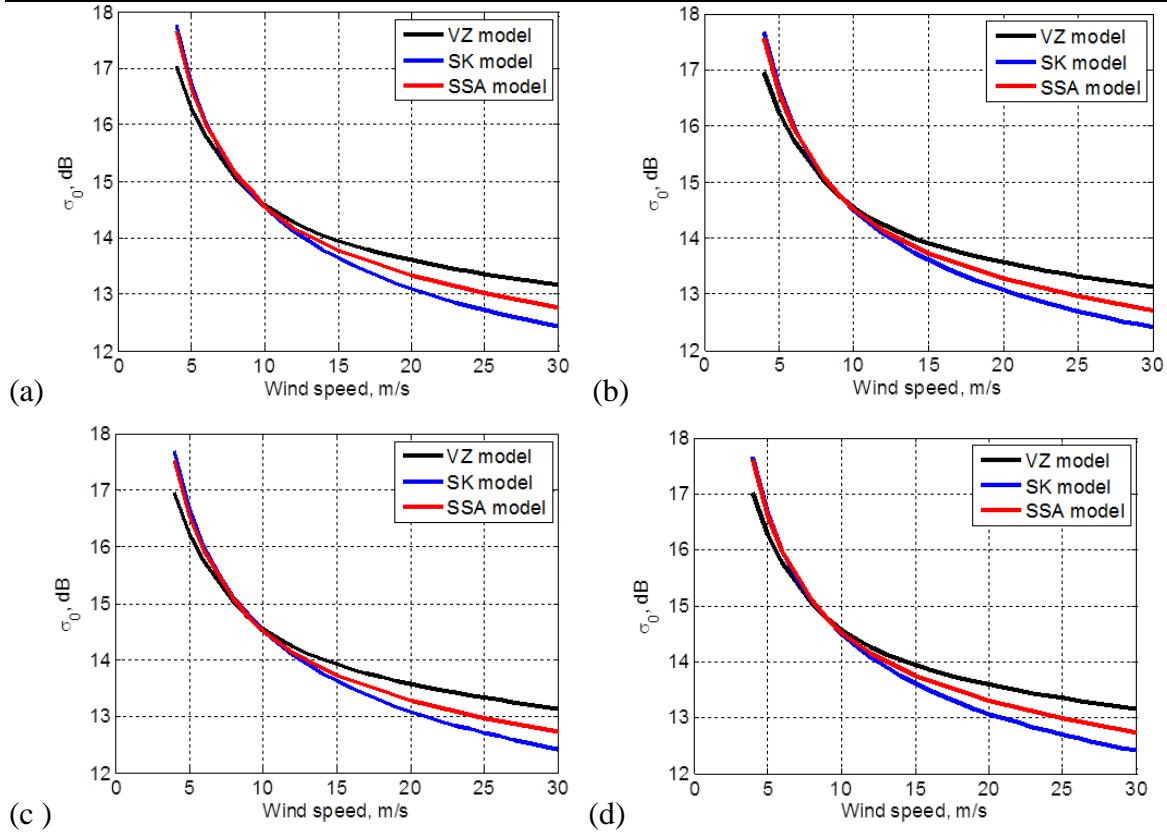
Figure 20. plots of σ_0 **obtained with the three models for** θ_{in} **= 30°,** *H* **= 600 km, and for four directions of scattering vector described by following combinations of zenith and azimuth scattering angles: (a)** $\theta_{sc} = 30°, \phi_{sc} = 0°$ **; (b)** $\theta_{sc} = 30°, \phi_{sc} = 2.56°$ **; (c)** $\theta_{sc} = 28°, \phi_{sc} = 0°$ **; and (d)** $\theta_{sc} = 32°, \phi_{sc} = 0°$ **.**

One can see that curves in Figure 20 practically repeat themselves at each panel. This means that while LHCP BRCS σ_0 is changing with the wind speed it does not appreciably change over the angles within the first delay zone. The discrepancy between curves for all three models are within 0.5 dB for wind speed below 15-17 m/s which is rather negligible given such adverse factors as speckle noise and natural wind speed variability that accompany real measurements. The discrepancy between the SSA curve and the SK curve (which we use in the end-to-end DDM simulator) is less than 0.5 dB for the entire range of wind speeds used for this simulation, i.e. below 30 m/s. More important, the steepness of these two curves is similar, which would result in a similar accuracy of the wind retrievals from the real GNSS-R data.

Below in Figure 21 (a-d) we present similar plots of σ_0 obtained with the three models for the same basic geometry but it gives an angular extent for the tenth delay zone (*n* = 10). The tenth delay zone covers the surface area which contributes to the DDM that will be routinely used during CYGNSS mission. Here, therefore: (a) $\theta_{sc} = 30°, \phi_{sc} = 0°$ (this plot repeats plot (a) from

the previous figure; it is given for comparison purposes); (b) $\theta_{sc} = 30°, \phi_{sc} = 8°$; (c) $\theta_{sc} = 25°, \phi_{sc} = 0°$; and (d) $\theta_{sc} = 35°, \phi_{sc} = 0°$.

(a)

(b)

(c)

(d)
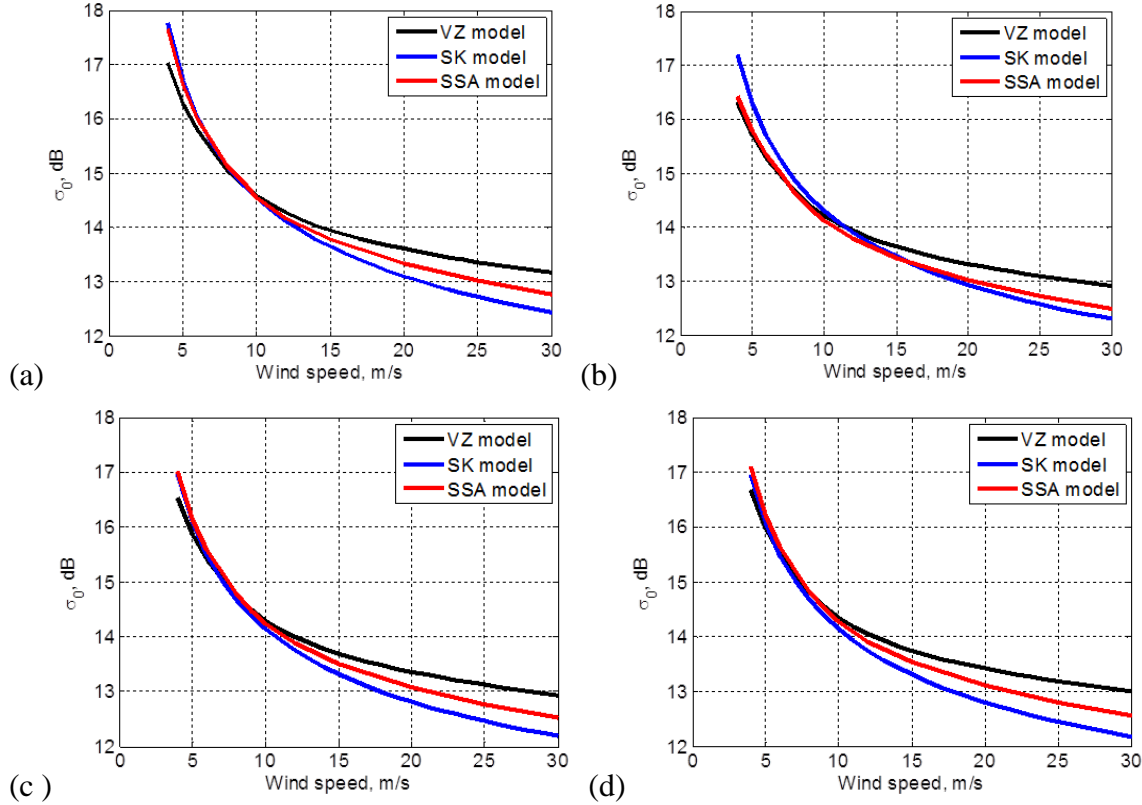
Figure 21. plots of σ_0 obtained with the three models for the same basic geometry and an angular extent for the tenth delay zone (n = 10). (a) $\theta_{sc} = 30°, \phi_{sc} = 0°$; (b) $\theta_{sc} = 30°, \phi_{sc} = 8°$; (c) $\theta_{sc} = 25°, \phi_{sc} = 0°$; and (d) $\theta_{sc} = 35°, \phi_{sc} = 0°$.

As one can see there is no significant difference between this set of plots and the one from the previous figure. Therefore, same statement about σ_0 behavior can be presented here for the case of the tenth delay zone.

4.3   Conclusions on Rough Surface Scattering

We described how the bistatic radar cross section σ_0 of the ocean, wind-driven rough surface emerges within the framework of the bistatic radar equation which governs the average GNSS-R signal in the delay-Doppler domain. We presented two alternative approaches to simulate σ_0. One of them is based on the geometric optics (GO) limit of the Kirchhoff approximation, and another one is the Voronovich small slope approximation (SSA) of the 1st order. The latter approach is superior to the former one because it combines two scattering mechanisms: quasi-

specular reflections at steep incidence and the Bragg resonant scattering at the shallower incidence, whereas the GO approximation relies only on the first mechanism for the whole range of incidence angles. Both of these approaches require knowledge of the ocean wave spectrum, or, as in the case of the GO approach the model of the mean-square-slopes (MSS) will be suffice. To this end, the theoretical model based on Elfouhaily's ocean wave spectrum with two different frequency cutoffs was tested, as well as Dr. Katzberg's empirical MSS model. All these models demonstrate a good agreement for weak and intermediate winds. They depart from each other only for strong winds, and this departure is rather tolerable given such adverse factors as speckle noise and natural wind speed variability that accompany real measurements.

Originally, Katzberg's empirical MSS model has been chosen for the end-to-end simulator of the Delay-Doppler Map. This choice was made because this model is based on a collection of aircraft GPS reflection measurements obtained for a large variety of wind speeds including for hurricane conditions. Comparisons between σ_0 modeled with both the GO and the SSA approaches show that for the geometry of CYGNSS orbital observatories and for the range of winds up to 30 m/s the GO approximation with Katzberg's empirical MSS model works very well. The advantage of the GO approximation is its simplicity and high speed of calculations, whereas the SSA approximation is more time consuming. All this makes our choice for the σ_0 computational algorithm even more substantiated.

Previously, some concerns have been expressed (see, e.g., [Thompson et al., 2000]) that the GO approximation might not work well for the GNSS reflectometry because it cannot properly account for out-of-plane scattering. Generally, the GO approximation has its own limitations, especially for calculations of the RHCP σ_0, and particularly for the out-plane configuration. However, as it was demonstrated here, for small deviations from the specular plane, the LHCP σ_0 is quite close to that one predicted by the more accurate SSA approximation.

The more fundamental limitation of all above models lies in the fact that they either have been proven only for global winds below 25-30 m/s (such as for those based on the Elfouhaily spectrum), or their accuracy is not high for strong hurricane winds (such as in the case of Katzberg's MSS model). For hurricane conditions, a feasible wave-spectral model should include, apart from a local wind speed, also several other parameters such as a distance from the hurricane center, azimuthal angle (a quadrant), hurricane velocity and other hurricane parameters.

Plans are to use an existing WAVEWATCH III wave model [Tolman et al, 1998] developed at the NOAA National Centers for Environmental Prediction (NCEP) which is now widely used for hurricane long-wave predictions (see, e.g., [Fan et al., 2009]). As a first step, it is expected that this model will be able to provide us with the long-wave portion of the sea state spectrum (so called "fresh swell") in the hurricane eye specifically and everywhere in general. More challenging would be a task to extend this model toward much shorter waves up to the cutoff frequency introduce above. Also, plans are to verify the DDM output of such a model with already available radiometric, scatterometric and GNSS-R data obtained in hurricanes. For this, archives of past aircraft GNSS-R overpasses of hurricane eyes can be used in order to assemble time series records of the MSS and to infer surface wind speed in the eye. Other surface wind

estimates (e.g. from flight level winds or SFMR) are also available. Raw DDM-grade GNSS-R data from Hurricanes Ike (2008), Rafael (2012) and Sandy (2012) obtained with CU bistatic GNSS bistatic radar can also be used for validation purposes. The work on processing of these data for CYGNSS purposes is currently performed.

4.4 Delay and Doppler Coordinate System

The scattered signal can be thought of as a superposition of components scattered from various points on the sea surface. Each component will have a shift in both the time at which the signal arrives at the receiver (delay shift) and the frequency of the signal (Doppler shift). The diversity in delay is due to different paths followed by each scattered signal, while different frequency shifts are caused by the relative motion between transmitter, scattering point on the surface and receiver. Each point of the GZ is therefore characterized by its own delay and Doppler shift [Clarizia, 2012].
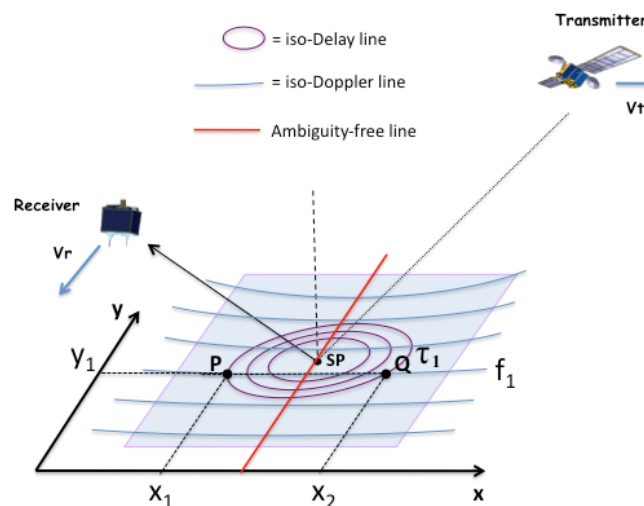

Figure 22. Delay-Doppler coordinates in GNSS-R and their relation to the space coordinate system (from [Clarizia, 2012]).

The pair of delay-Doppler values to which each point in space can be associated represents indeed a new domain in which the GZ can be mapped, and it is known as delay-Doppler domain. Such domain is fundamental for GNSS-R processing since it is the domain in which GNSS-R data are commonly presented and mapped, in the form of the so called delay-Doppler Maps (DDMs). However, different points on the sea surface will correspond to the same pair of delay-Doppler values. Lines corresponding to constant delays (iso-range) and constant Doppler shifts (iso-Doppler) can be identified on the sea surface, and they have respectively an elliptical and parabolic shape. Lines of constant delays, also called iso-range lines, are given by concentric ellipses around the SP, and they correspond to increasing delays for increasing distance from the SP, which is the point at minimum delay. Rigorously speaking, the iso-range lines are the intersections of spheroids (equi-range surfaces) having receiver and transmitter as foci, with the sea surface, which causes the ellipses to be not exactly concentric as their centers move towards the transmitter [Zuffada et al., 2004]. The iso-Doppler lines are parabolic shaped lines cutting through the GZ. They are also asymmetric and characterized by complicated equations, and lines

of lower and higher Doppler frequency shifts cannot be predicted, since they strictly depend on the relative velocities among the transmitter, the scattering point and the receiver. From Figure 22, we can notice that a generic point P on the GZ can be described by a delay and Doppler coordinate. Such a correspondence is however not biunivocal, since there is an ambiguity since the intersection between an iso-range and iso-Doppler line is made of two points in space, which will have the same Delay and Doppler frequency, like points P and Q in Figure 22. Despite that, it is interesting to note that there exists a line free of ambiguity, which can be thought as the transverse axis of the hyperbolic iso-Doppler lines, shown in red in Figure 22.

The space-to-DD transformation of coordinates is also what gives the DDM a characteristic horseshoe shape. Such transformation operates on the spatial domain by "folding" the glistening zone along the free ambiguity line, and by "bending" it at the specular point, or the peak power in the DDM. The scattered power at the specular point corresponds therefore to the central point of the horseshoe shape, and the horseshoe branches correspond to the scattered power from the glistening zone, with areas farther from the SP spanning larger delays and Doppler shifts.

One other important aspect of the delay-Doppler coordinate system is its dependence on the geometry, and in particular on the incidence angle, which strongly influences the configuration of the iso-delay and Iso-Doppler lines. Figure 23 shows the change in the iso-delay contours over a footprint of 100 x 100 km^2, for different incidence angles, where the incidence angle is the angle between the transmitter or receiver range and the normal to the surface. The iso-delay ellipses tend to stretch out and become wider for higher incidence angles. Here the transmitter and receiver altitudes have been assumed to be respectively equal to the GPS (~20200 km, and 475 km).

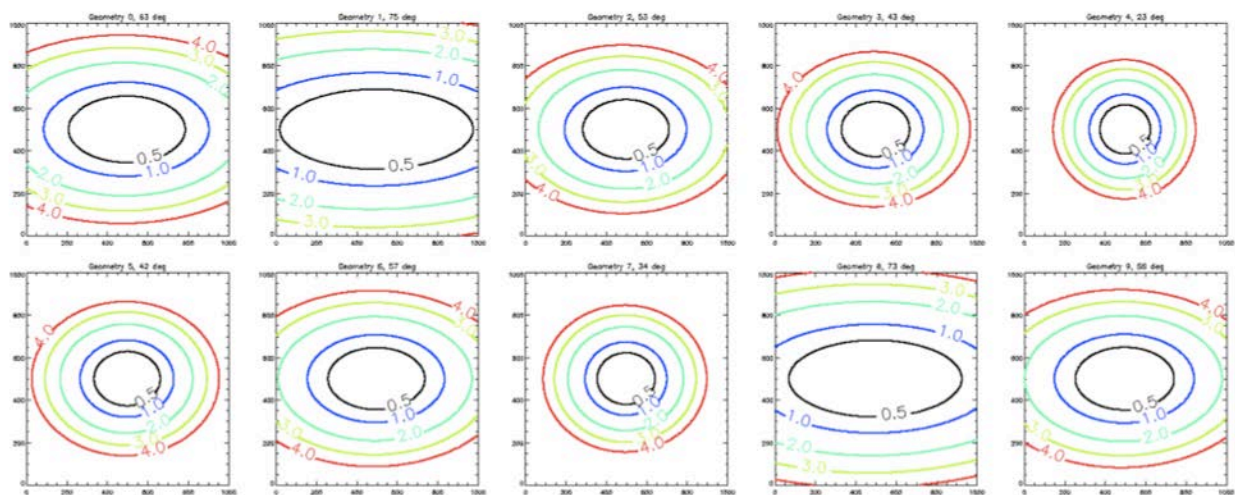**Figure 23. Iso-delay contours over a footprint of 100 x 100 km, for the following incidence angles (top, left to right): 63°, 75°,53°,43°,23°; (bottom, left to right): 42°, 57°, 34°, 73°, 58°.**

Figure 24 shows the range of maximum delays (a) and maximum Doppler frequencies (b) within a 50 x 50 km footprint, as a function of incidence angle. In Figure 24(b), a specific velocity vector has been assumed for the GPS and the receiver satellite. In principle, once the whole geometry is known (i.e. transmitter and receiver altitudes, incidence angles and velocity vectors), the range of delays and Dopplers spanned by a footprint of given size can be calculated numerically.

(a)   (b)
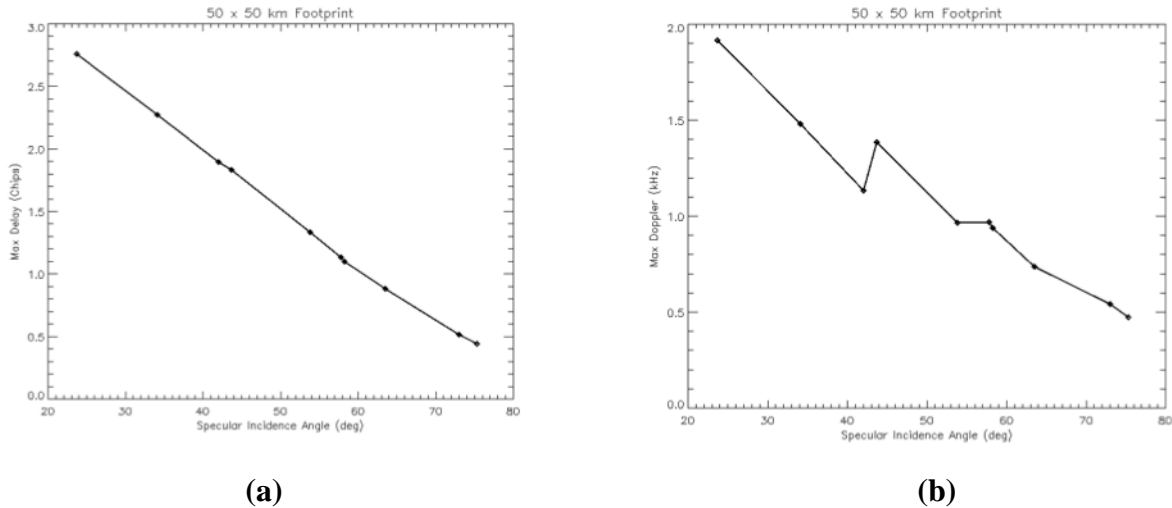
Figure 24. (a) maximum delay as a function of incidence angle, within a 50 x 50 km² footprint; (b) maximum Doppler as a function of incidence angle, within the same footprint.

4.5   *Mean Power and Signal-to-Noise Ratio for the GPS Reflected Signal*

Let us represent the instantaneous complex signal (the voltage) u, which is acquired directly by the receiver from the antenna output, or as a result of some coherent processing, in the form:

$$u(t) = s(t) + n(t), \tag{4.20}$$

where $s(t)$ is the complex amplitude of the scattered signal, and $n(t)$ is the complex amplitude of the additive noise. We assume that $s(t)$ and $n(t)$ are two uncorrelated, stationary random processes, both obey circular Gaussian statistics and have different time scales, and different variances $\sigma_1^2 \equiv \sigma_{\mathrm{Re}\,s}^2 \equiv \sigma_{\mathrm{Im}\,s}^2$ and $\sigma_2^2 \equiv \sigma_{\mathrm{Re}\,n}^2 \equiv \sigma_{\mathrm{Im}\,n}^2$, both with zero means. The Gaussian statistics for $s(t)$ can be justified if the signal at the antenna is formed by contributions from a large number of independent surface scatterers. Here, we exclude from a consideration fluctuations of the signal caused by propagation through ionospheric and tropospheric irregularities. Fluctuations of $s(t)$ generate multiplicative, self-noise, (other names: interference noise, Rayleigh fading, speckle noise), which are proportional to the signal, whereas fluctuations of $n(t)$ produce additive, background noise (i.e., thermal noise or shot noise). In a more complex situation, the additive noise could include extraneous emitted signals. So, in what follows we limit that background noise to thermal noise.

A coherent processing of the scattered GPS signal by the correlator channel of the CYGNSS receiver consists of the convolution (correlation) of voltage $u(t)$ with the replica a of the GPS broadcast signal over a relatively short (milliseconds) coherent integration time T_i:

$$Y(t_0, \tau) = \int_0^{T_i} a(t_0 + t') u(t_0 + t' + \tau) dt'. \tag{4.21}$$

Taking into account (4.20), we obtain from (4.21) that

$$Y(t_0, \tau) = Y_s(t_0, \tau) + Y_n(t_0, \tau), \tag{4.22}$$

where

$$Y_s(t_0, \tau) = \int_0^{T_i} a(t_0 + t') s(t_0 + t' + \tau) dt', \tag{4.23}$$

$$Y_n(t_0, \tau) = \int_0^{T_i} a(t_0 + t') n(t_0 + t' + \tau) dt'. \tag{4.24}$$

Therefore, the quantities in (4.23) and (4.24) are short-integrated (practically, instantaneous) correlation voltages, respectively, for the signal and noise. The next step of the signal processing is obtaining the mean power of the correlator output. It is obtained by an additional averaging of $|Y(t_0, \tau)|^2$ over a long enough observation time, so both thermal and surface-induced fluctuations are substantially averaged out. The result is

$$\left\langle |Y(t_0, \tau)|^2 \right\rangle = \left\langle |Y_s(t_0, \tau)|^2 \right\rangle + \left\langle |Y_n(t_0, \tau)|^2 \right\rangle. \tag{4.25}$$

4.5.1   The Signal Term

The first term in Eq. (4.25) is known in the literature [Zavorotny and Voronovich, 2000] as the GPS radar bistatic equation:

$$\left\langle |Y_s(t_0, \tau)|^2 \right\rangle = T_i^2 \frac{P_T G_T \lambda^2 G_R}{(4\pi)^3}$$
$$\times \iint F(\vec{\rho}) \Lambda^2(\tau, \vec{\rho}) |S(f_{dop}, \vec{\rho})|^2 R_0^{-2} R^{-2} \sigma_0(\vec{\rho}) d^2\rho. \tag{4.26}$$

where P_T is the transmitter power; G_T is the transmit antenna gain; G_R is the receive antenna gain; $F(\vec{\rho})$ is the normalized directivity (beam) pattern for the receive antenna; $\Lambda^2(\tau, \vec{\rho})$ is the annulus function due to the cross-correlation with the replica; $|S(f_{dop}, \vec{\rho})|^2$ is the Doppler zone function due to the relative motions of both the transmitter and receiver with respect to the

scattering surface; R_0, R are distances from a point $\vec{\rho}$ on the surface to the transmitter and receiver, respectively; $\sigma_0(\vec{\rho})$ is a bistatic cross section of the rough surface, and generally it is a function of two angles, the incidence angle and the scattering angle. Here, in (4.26) it is written as a function of surface coordinates. The scattered signal comes from the area formed by intersection of the equi-range zones, annular (function $\Lambda^2(\tau, \vec{\rho})$) and equi-Doppler zones (function $|S(f_{dop}, \vec{\rho})|^2$). The width of the annulus depends on the code length (different for C/A code and P code) and on all geometric parameters of the problem. The width of the Doppler zone depends on the receiver velocity and a coherent integration time, T_i.

The bistatic radar cross section used in this model is based on geometric optics approximation (see, e.g., [Barrick, 1968]). With regard to our problem it has the following form [Zavorotny and Voronovich, 2000]

$$\sigma_0(\vec{\rho}) = \frac{\pi |\Re|^2 q^4}{q_z^4} P\left(-\frac{\vec{q}_\perp}{q_z}\right). \tag{4.27}$$

The cross section is maximal at $\vec{q}_\perp = 0$ because the probability density function (PDF) of slopes $P(\vec{s})$ has a maximum at $\vec{s} = 0$, i.e., for the most probable orientation of slopes, parallel to z = 0.

Here, we use the Gaussian bivariate statistics of anisotropic slopes:

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{mss_x mss_y (1 - b_{x,y}^2)}} \exp\left[-\frac{1}{2(1 - b_{x,y}^2)}\left(\frac{s_x^2}{mss_x} - 2b_{x,y}\frac{s_x s_y}{mss_x mss_y} + \frac{s_y^2}{mss_y}\right)\right] \tag{4.28}$$

where mss_x and mss_y are mean-square slopes of the sea surface for two orthogonal components; $b_{x,y}$ is the correlation coefficient between two slope components:

$$mss_{x,y} = \left\langle s_{x,y}^2 \right\rangle = \iint_{\kappa < \kappa_*} \kappa_{x,y}^2 W(\vec{\kappa}) d^2\kappa \tag{4.29}$$

$$b_{x,y} = \left\langle s_x s_y \right\rangle / \sqrt{mss_x mss_y} \tag{4.30}$$

$$\left\langle s_x s_y \right\rangle = \iint_{\kappa < \kappa_*} \kappa_x \kappa_y W(\vec{\kappa}) d^2\kappa \tag{4.31}$$

Here, $W(\vec{\kappa})$ is the surface elevation spectrum. For calculations we adopt the Elfouhaily spectrum [Elfouhaiyl et al., 1997]. The sea surface contains wave harmonic components both larger and shorter than the L-band electromagnetic waves. The short waves can be disregarded in a process of forward quasi-specular reflection. Because of this, the geometric optics approximation is applicable here but then the full surface spectrum should be cut off at high end of wave numbers. There are various choices of cutoff wave number κ_*. Here we choose one from [Thompson et al., 2005], since the author claimed that it gave the best fit with measurements:

$$\kappa_* = k \cos\theta \left(1 + U_{10}/20\right)/7.5 \qquad (4.32)$$

In (4.27) also enters $\Re$, the LHCP Fresnel reflection coefficient for sea water at L-band [Zavorotny and Voronovich, 2000]:

$$\Re = \frac{1}{2}\left[\frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} - \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}}\right] \qquad (4.33)$$

where ε is the complex dielectric permittivity of sea water, and θ is the incidence angle.

According to Klein and Swift model:

at S = 35 ppt and T = 10 deg C $\varepsilon = 74.62 + i51.92$ for L1 = 1.57542 GHz; $\varepsilon = 75.02 + i62.39$ for

L2 = 1.22760 GHz; at S = 30; T = 10 deg C $\varepsilon = 76.16 + i55.30$ for L1; $\varepsilon = 75.02 + i62.39$: for L2.

Let us explore the behavior of the σ_0 as a function of the incidence angle with the scattering angle which corresponds to the nominal specular direction, i.e. $\theta_{inc} = -\theta_{sc}$. In this case

$$\sigma_0 = \frac{|\Re|^2}{2\sqrt{mss_x mss_y (1 - b_{x,y}^2)}} \qquad (4.34)$$

Below is the plot of σ_0 for two winds (along the wind direction, $b_{x,y} = 0$) calculated according to above formulas and numbers. The roll-off at large incidence angle is due to the angular behavior of the LHCP reflection coefficient $\Re$.
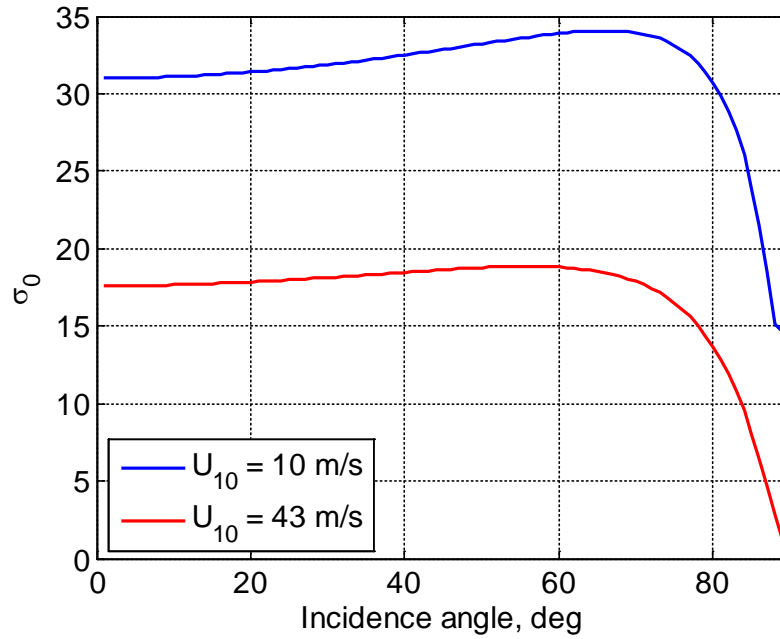
Figure 25. The bistatic radar cross section at the specular direction.

4.5.2 Thermal Noise

The second term in Eq. (4.25) is the background noise term. It can be written as double integral over the coherent integration time:

$$\left\langle \left| Y_n(t_0, \tau, f) \right|^2 \right\rangle = \int_0^{T_i} dt' \int_0^{T_i} dt'' a(t_0 + t', f) a(t_0 + t'', f).$$
$$\times \left\langle n(t_0 + \tau + t') n^*(t_0 + \tau + t'') \right\rangle. \tag{4.35}$$

Assume that the thermal noise is the "white" (delta-correlated) noise, i.e.:

$$\left\langle n(t') n^*(t'') \right\rangle = kT^\circ B_n b_n (t' - t''), \tag{4.36}$$

where k is Boltzmann constant; T° is the receiver noise equivalent temperature in Kelvin; $B_n = 1/T_{cor}$ is the receiver-front-end bandwidth, and T_{cor}, is a temporal correlation scale of the noise filtered by the front end;

$$b_n(t) = \int W_n(f)\exp(2i\pi ft)df; \qquad \int_0^{T_i} b_n(t)dt = T_{cor}. \qquad (4.36)$$

Here $W_n(f)$ is the normalized temporal spectrum of the noise. Usually, background noise has a much smaller temporal correlation scale, T_{cor}, than the C/A chip-length, $\tau_{chip} = 1\,\mu s$. Or, in other words, that the noise bandwidth is much greater then the bandwidth of the C/A pseudo-random phase modulation of the GPS signals. Then, we can regard function $b_n(t'-t'')$ as a delta-function, so two integrations over time can be performed trivially. The result is:

$$\left\langle \left| Y_n(t_0, \tau, f) \right|^2 \right\rangle = T_i^2 kT^\circ B_D, \qquad (4.37)$$

where $B_D = 1/T_i$ is the Doppler bandwidth of the signal.

The thermal noise is correlated between delay-Dopplers bins. The cross-correlation function of the noise in different bins is given by

$$\left\langle Y_n(t_0, \tau, f) Y_n^*(t_0, \tau', f') \right\rangle = \int_0^{T_i} dt' \int_0^{T_i} dt'' a(t_0 + t', f) a^*(t_0 + t'', f').$$
$$\times \left\langle n(t_0 + \tau + t') n^*(t_0 + \tau' + t'') \right\rangle. \qquad (4.38)$$

Since the noise is uncorrelated with the reference signal, the reference signal acts as a filter for the noise, causing it to be correlated with respect to delay and Doppler according to the GPS C/A code ambiguity function. (4.38) can be reduced to

$$\left\langle Y_n(t_0, \tau, f) Y_n^*(t_0, \tau', f') \right\rangle = T_i^2 kT^\circ B_D \Lambda^2(d\tau) \left| S(df) \right|^2 \qquad (4.39)$$

In the forward model, zero mean white Gaussian noise is generated with respect to delay and Doppler using the power level in (4.37) and then convolved with the ambiguity function to produce the correct bin-to-bin correlations.

Now we can construct the signal-to-noise ratio (SNR). There are various definitions of SNR. We use here the simplest one, which shows how much the mean power of the signal exceeds the mean noise level:

$$SNR = \left\langle \left| Y_s(t_0, \tau) \right|^2 \right\rangle \Big/ \left\langle \left| Y_n(t_0, \tau) \right|^2 \right\rangle. \qquad (4.40)$$

Remind that the SNR is the function of parameters τ and f_{dop}, i.e., the SNR is different for different portions of waveforms taken at different time delays and Doppler frequency offsets.

4.5.3 Statistics of the partially averaged GPS reflected signal affected by both thermal and speckle noise

Before, we considered an effect of additive thermal noise on the average SNR. It exists due to the physical temperature of both the receiver and the scene even in absence of the GPS reflected signal. Another type of noise, the multiplicative one, is a result of distractive and constructive interference of coherent signals arriving to the antenna upon scattering from a rough ocean surface. It is called Rayleigh fading, or speckle noise, and it is proportional to the signal itself. Below we consider statistics of the partially averaged signal affected by both thermal and speckle noise. In a real situation we deal with values averaged over a finite time interval. It happens because of, at least, two reasons. First, any measuring device has a finite time response. Second, often signals need to be accumulated over some time in order to improve signal-to-noise ratio. Since the integration, or averaging, time is finite the procedure doesn't lead to constant time-independent values. These partially averaged values are still random quantities and need to be described in statistical terms. Note that an instantaneous power U of the signal + noise does not comprise only of the sum of the instantaneous powers S and N for the signal and the noise, respectively. It contains also cross terms of s and n. Indeed

$$U(t) \equiv |u(t)|^2 = [s(t) + n(t)][s^*(t) + n^*(t)] = S(t) + N(t) + C(t) + C^*(t), \tag{4.41}$$

where

$$C(t) = s(t)n^*(t). \tag{4.42}$$

During the measurement we obtain an estimate of the signal + noise from the power of the received signal + noise averaged over an arbitrary time interval T (a bar above refers to that type of averaging):

$$\overline{U}(t) \equiv \overline{S}(t) + \overline{N}(t) + \overline{C}(t) + \overline{C}^*(t) = \frac{1}{T} \int_{-T/2}^{T/2} [S(t+t') + N(t+t') + C(t+t') + C^*(t+t')] dt'. \tag{4.43}$$

An estimate of the signal can be done by obtaining an estimate of the signal + noise, then obtaining an estimate of noise from an independent measurement, and then subtracting one from another:

$$\tilde{S}(t) = \overline{U}(t) - \overline{N}(t_0). \tag{4.44}$$

Since these estimates are obtained from an averaging over the finite period of time, the estimate of the signal, $\tilde{S}(t)$, is a fluctuating quantity. The accuracy of the estimate is governed by the variance of estimate $\tilde{S}(t)$. Since $\overline{U}(t)$ and $\overline{N}(t_0)$ are statistically independent the following equality holds

$$\sigma_{\tilde{S}}^2 = \sigma_{\overline{U}}^2 + \sigma_{\overline{N}}^2. \tag{4.45}$$

Observe that the mean value of the estimated power of the signal + noise is simply

$$\langle \overline{U} \rangle = \langle S \rangle + \langle N \rangle. \tag{4.46}$$

The variance of the total power of signal + noise is:

$$\sigma_{\overline{U}}^2 = \frac{1}{T^2} \left\langle \left| \int_{-T/2}^{T/2} \left[S(t') + N(t') + C(t') + C^*(t') \right] dt' \right|^2 \right\rangle - \langle \overline{U} \rangle^2. \tag{4.47}$$

The variance of the noise power is:

$$\sigma_{\overline{N}}^2 = \frac{1}{T^2} \left\langle \left| \int_{-T/2}^{T/2} N(t') dt' \right|^2 \right\rangle - \langle N \rangle^2. \tag{4.48}$$

Assuming that both the signal and the noise are stationary, and making several additional simplified assumptions without a loss of generality the standard deviation of the estimated signal power can be obtained in the following form:

$$\frac{\sigma_{\tilde{S}}}{\langle S \rangle} = \left[\left(1 + 2 \langle S \rangle + 2 T_{\text{int}} / \tau_{cor} \langle S \rangle^2 \right) / N \right]^{1/2}. \tag{4.49}$$

For $T \gg \tau_{cor}$ the parameter $N = T / \tau_{cor}$ is the number of correlation intervals contained within the measurement time, T. Or, it could be interpreted as a number N of independent samples.

In order to proceed further we need to choose the value of the correlation time of the signal, τ_{cor}. The approach for calculation of the correlation time based on the power spectrum of the scattered signal, or equivalently, through the coherence function of the signal was developed in [Zuffada and Zavorotny, 2001, You et al., 2004, 2006].

Actually, the time correlation can be estimated using the Van Cittert-Zernike theorem. From it follows that the size of the field correlation zone at the wavelength λ is:

$$r_{cor} = \lambda R / D \tag{4.50}$$

where R is a distance from the surface to the receiver, and D is a size of the illuminated area. At the peak correlation power the illuminated area is the first annulus zone modified by the smaller, Doppler zone. The smallest size matters because it created the biggest r_{cor} which translates into largest correlation time $\tau_{cor} = r_{cor} / v_{sat}$ (See, e.g., Fig. 1. in [5]). The size of the Doppler zone is dictated by the coherent integration time. The analysis shows that $\tau_{cor} \approx 2T_i$, therefore, if T_i is 1 ms, $\tau_{cor} = 2$ ms. Taking this into account, (4.49) simplifies:

$$\frac{\sigma_{\tilde{S}}}{\langle S \rangle} = \left(1 + 2\langle S \rangle^{-1} + 2T_{int}\tau_{cor}^{-1}\langle S \rangle^{-2}\right)^{1/2} N^{-1/2} \approx \frac{1 + 1/\langle S \rangle}{\sqrt{N}} \quad . \tag{4.51}$$

From here we can produce an expression for the standard deviation of S (the SNR) after incoherent averaging over N statistically independent samples:

$$\sigma_{SNR} == \frac{\langle S \rangle + 1}{\sqrt{N}}. \tag{4.52}$$

Therefore, in this case, the standard deviation of partially averaged SNR is proportional to the average SNR plus one, and can be reduced by $\sqrt{N}$. In a general case of arbitrary τ_{cor} and T_i

$$\sigma_{SNR} = \frac{\sqrt{\langle S \rangle^2 + 2\langle S \rangle + 2T_{int}\tau_{cor}^{-1}}}{\sqrt{N}} \tag{4.53}$$

4.6   Speckle Noise

The reflected signal received by each CYGNSS observatory is formed by contributions from a large number of independent surface scatterers. This random scattering generates multiplicative, self-noise (i.e. Rayleigh fading or speckle noise), which is proportional to the signal. This is in contrast to thermal noise, which is additive. This section describes how this speckle noise is accounted for in the forward model.

Recall that the bistatic radar equation from Section 4.5.1. The expected value of the power of the reflected signal versus delay and Doppler can be rewritten as

$$\left\langle \left| Y_s(t_0, \tau, f) \right|^2 \right\rangle = \iint H(\vec{\rho}) \Lambda^2(\tau, \vec{\rho}) \left| S(f, \vec{\rho}) \right|^2 d^2\rho, \tag{4.54}$$

where

$$H(\vec{\rho}) = T_i^2 \frac{P_T G_T \lambda^2 G_R}{(4\pi)^3} R_0^{-2} R^{-2} \sigma_0(\vec{\rho}) \tag{4.55}$$

represents the contribution of each location on the surface to the total expected power of the reflected signal at a particular delay and Doppler.

In the forward model, DDMs are formed from integrations performed over finite time intervals rather than expected values (such as in equation 4.54). We must model the effect of speckle noise, but, for the surface areas involved in space-borne GPS reflectometry, it would be unrealistic to instantiate the actual random rough surface and use a computational electromagnetics approach. Rather, we have chosen a suitable to accurately capture the effects of speckle noise.

First, we take the square root of the power contribution in equation 4.55 and include a time varying phase term $\phi(t,\rho)$ to make the contribution complex,

$$h(t,\rho) = \sqrt{H(\rho)}e^{j\phi(t,\rho)}. \tag{4.54}$$

This is an approximate representation of the contribution of each location on the surface to the voltage DDM, and can be thought of as the transfer function over the surface. The approximate voltage DDM is given by

$$Y_s(t,\tau,f) = \iint h(t,\vec{\rho})\Lambda(\tau,\vec{\rho})S(f,\vec{\rho})d^2\rho. \tag{4.55}$$

The DDM is formed by integrating for 1 second, t will be between t_0 and (t_0+1), where t_0 is the start of the integration. This produces

$$\left|Y_s(t_0,\tau,f)\right|^2 = \int_{t_0}^{t_0+T} Y_s(t,\tau,f)Y_s^*(t,\tau,f)dt, \tag{4.56}$$

where $T = 1$. The phase term $\phi(t,\rho)$ must be chosen such that the expectation of equation 4.56 is equal to 4.54. Also, it must also result in the temporal correlation of speckle noise.

First, a random phase, $\phi_0(\rho)$, is associated with each location on the surface. This random phase is assumed uniformly distributed between 0 and 2π and represents the phase shift caused by the random rough surface at that location. This phase will evolve in time according to the changing geometry of the satellites. Thus, the total phase associated with the reflection of a particular point on the surface is a combination of the random phase and phase associated the total path length,

$$\phi(\rho) = \phi_0(\rho) + \frac{2\pi}{\lambda}R(t,\rho), \tag{4.57}$$

where λ is the wavelength at the GPS L1 center frequency, and $R(t,\rho)$ is the total path length from the transmitter to the surface location at ρ and up to the receiver at time t. Since it is such a short duration, the time variation in the path length can be accurately approximated using the Doppler at the start of the integration $f_D(t_0,\rho)$,

$$R(t,\rho) = R(t_0,\rho) - (t-t_0)\lambda f_D(t_0,\rho). \tag{4.58}$$

Each point of the surface will exhibit a different time varying phase, depending on the relative motion of the satellites. Over short time delays (i.e. less than one millisecond), the change in geometry will be small, and the speckle noise will exhibit time correlation. For longer delays, the speckle noise will be completely uncorrelated, as is expected from reflections from a real ocean surface.

5 L2 Wind Speed Retrieval Algorithm

The L2 wind speed retrieval algorithm described here is the algorithm implemented in the CYGNSS Science Operation Centre (SOC). It is similar to the algorithm described in [Clarizia and Ruf, 2016], although the observable computation and the time-averaging approach are here slightly different, and some of the filters, which were implemented in [Clarizia and Ruf, 2016] to demonstrate that the resolution requirements are met, are not implemented by the SOC. The Geophysical Model Functions (GMFs) used to map the observables into wind speed are based on empirical parameterized fits to on-orbit measurements, as described in [Ruf and Balasubramaniam, 2018]. Two sets of GMFs are used. One is derived in fully developed seas (FDS) conditions using coincident matchups with ECMWF nurmerical weather prediction model winds speeds. The other is derived using ECMWF matchups at low wind speeds and, at high wind speeds, matchups with coincident NOAA P-3 hurricane hunter observations of hurricane force winds made by their Stepped Frequency Microwave Radiometers. The latter GMF is refered to as Young Seas with Limited Fetch (YSLF). A decription of the empirical GMF development is included in appendix A of this document.

The basic steps for the L2 retrieval algorithm can be summarized as follows:

1. Two DDM "observables", the DD Map Average (DDMA) and the Leading Edge Slope (LES), are derived from L1b DDMs of Radar Cross Section (RCS) and DDMs of effective scattering area, both output by the L1 calibration procedure described in [Gleason et al., 2016; Gleason et al., 2018];
2. Wind speed is estimated from the L1 observable by inversion of the appropriate GMF (either FDS or YSLF), depending on the sea state conditions. This produces both a DDMA- and LES-based wind speed;
3. In FDS conditions, the DDMA and LES winds are optimally combined using a Minimum Variance (MV) Estimator. In YSLF conditions, the DDMA-based wind alone is used, as it has greater sensitivity at high wind speeds;
4. In FDS conditions, time averaging is applied to consecutive samples to produce a consistent 25 km spatial resolution data product, whenever it is appropriate to do so (the appropriate level of averaging depends on the incidence angle of the sample), In YSLF conditions, not time averaging is applied in order to retain the highest possible horizontal spatial resolution in and near the inner core of tropical cyclones;

These steps are described further in the subsequent sections.

5.1 Sample population of GMF training data

The DDMs used to develop the GMFs and to test the L2 retrieval algorithm performance were obtained from on-orbit measurements made by the CYGNSS constellation during the period July - Oct 2017. Included are measurements over the entire globe matched up with coincident reanalysis 10 m referenced ocean surface wind speed by ECMWF, for development of testing of the FDS GMF and low to morderate wind speed retrieval, and matched up with coincident

measured 10 m referenced ocean wind speed by SFMR instruments on NOAA P-3 "hurricane ehunter" aircraft during hurricane eyewall penetrations, for development and testing of the YSLF GMF and hiugh wind speed retrieval. [Ruf et al., 2018]

5.2   *DDM Observables: DDMA and LES*

Here we present a definition of the observables we use for our retrieval algorithm, derived from the outputs of the L1b calibration described in [Gleason et al., 2016; Gleason eta al, 2018].

5.2.1   Definition of Delay-Doppler Map Average (DDMA)

The DDMA is the average of the L1b DDM of the Normalized Radar Cross Section (NRCS) over a given delay/Doppler range window around the specular point [Clarizia et al., 2014, Clarizia and Ruf, 2016]. The DDMA exploits the DDM region which is most sensitive to varying wind speed, namely the scattered power at and around the specular point. Notably, the DDMA has the advantage of mitigating the effect of noise, by averaging the power over the area around the SP, rather than the power value at the single SP pixel. An illustration of a DDM of scattered power simulated with the CYGNSS E2ES, and a qualitative example of the area where the DDMA is calculated (shown as the "black box") is shown in Figure 27.
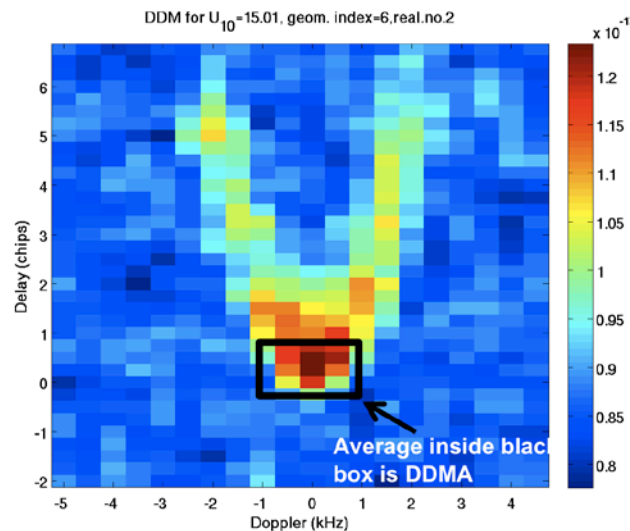

Figure 22. A simulated DDM and the DDM area where the DDMA is calculated.

5.2.2   Definition of Leading Edge Slope (LES)

The LES is the slope of the leading edge of the Integrated Delay Waveform (IDW), calculated over a specific delay range [Clarizia et al., 2014, Clarizia and Ruf, 2016]. IDWs are obtained as incoherent integration of DDMs of NRCS along the Doppler dimension, and over a range of Doppler frequencies. Figure 28 illustrates simulated IDWs of integrated power, for a fixed geometry and for different wind speeds. The corresponding IDWs of NRCS would only differ

from these for a constant scaling factor. Figure 28 highlights how the slope of the IDW rising edge decreases for increasing wind speeds.
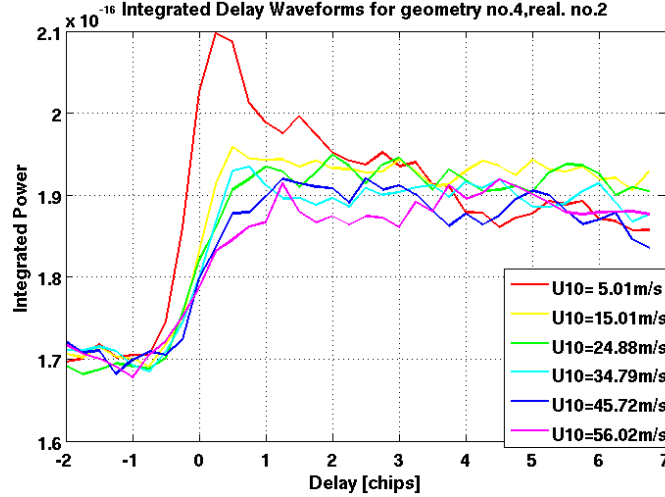

Figure 23. Integrated Delay Waveforms computed from DDMs simulated using different wind speeds.

5.2.3   Calculation of DDMA and LES Observables

The DDMA is computed as the sum of the L1b DDM of Radar Cross Section (RCS) over a delay/Doppler window of respectively -0.25 chips to 0.25 chips, and -1 kHz to 1 kHz, divided by the effective scattering area over the same window. The selected delay/Doppler window is a 3x5 matrix, due to the 0.25 chip delay resolution and the 500 Hz Doppler resolution of the CYGNSS DDMs.

The sum of L1b RCS DDMs can be expressed as follows:

$$\sigma_{DDMA} = \sum_{i=1}^{3}\sum_{j=1}^{5} d(i,j) \tag{5.1}$$

where $d(i,j)$ represents the $(i,j)^{th}$ delay/Doppler bin of the L1b DDM of RCS. The motivation for the choice of such a delay/Doppler window is documented in [Clarizia and Ruf, 2016].

The LES is calculated as the slope of IDWs obtained from L1a DDMs of received power using the same DD window as the DDMA. The IDWs are obtained by summing the 5 L1a DDM columns (along the Doppler axis) that span the Doppler interval from – 1 kHz to 1 kHz. The leading edge slope (LES) of the IDW leading edge is given by

$$a_{LES} = \frac{N\sum_{i=1}^{3} x_i y_i - \sum_{i=1}^{3} x_i \sum_{i=1}^{3} y_i}{N\sum_{i=1}^{3} x_i^{2} - \left(\sum_{i=1}^{3} x_i\right)^{2}} \tag{5.2}$$

where x_i is the *i-th* value of the delay within the *[-0.25 0.25]* chip interval, y_i is the corresponding IDW value, and *N=3* is the number of points of the IDW used for the LES computation.

The effective scattering area of the 3x5 region that is used for the L1b DDMA and LES calculation is computed from the DDMs of ideal area and effective area for individual DDM bins, both outputs of the L1b calibration process. The DDM effective area of individual bins within the selected DD window are not simply summed as this would have the effect of accounting for the spreading of effective area into neighboring bins multiple times, which would result in too big of an effective area. Therefore, the approximation we use for the effective area A_{eff} is:

$$A_{eff} = \sum_{i=1}^{3}\sum_{j=1}^{5} A_{id}(i,j) + \frac{1}{2}\sum_{i=1,3}\sum_{j=1,5}\Delta A(i,j) + \frac{1}{4}\sum_{i=1,3}\sum_{j=2,3,4}\Delta A(i,j)$$
$$\Delta A(i,j) = A_{eff}(i,j) - A_{id}(i,j)$$

(5.3)

where $A_{id}(i,j)$ and $A_{eff}(i,j)$ are the $(i,j)^{th}$ bin of the ideal and effective scattering area DDM, respectively.

The final equations for the DDMA and LES are given by:

$$DDMA = \frac{\sigma_{DDMA}}{A_{eff}}$$

$$LES = \frac{\alpha_{LES}}{A_{eff}}$$

(5.4)

Computation of the DDMA and LES is also explained in detail in [Clarizia and Ruf, 2016], although in that case the DDMA and LES of RCS are normalised by the ideal area, calculated simply as the sum of the DDM ideal area bins within the DD window considered for the observables.

5.3 Time Averaging

An extra processing step that contributes to improve the performances of the algorithm is to apply time averaging to the collected data.

The DD window chosen for the observable computation corresponds to a given Instantaneous Field of View (IFOV), and whose square root is shown as a function of incidence angle in Figure 29.
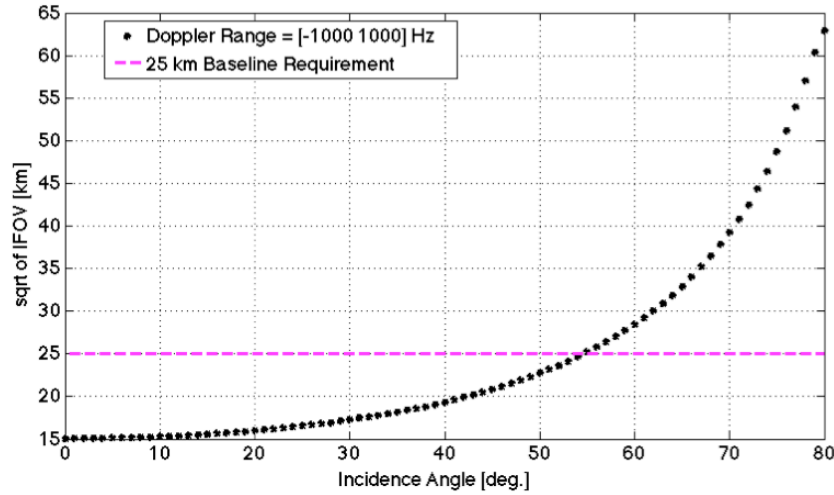
Figure 24. square root of Instantaneous Field of View (IFOV) versus incidence angle, for the DD window spanning -0.25 chip to 0.25 chip, and -1 kHz to 1 kHz (black dots), along with the requirements for the CYGNSS mission (dashed magenta line).

The IFOV is defined here as the physical area at the intersection between the iso-range ellipse at 0.25 chips, and the iso-Doppler lines at -1 kHz and 1 kHz. Full details about the relationship between physical area, IFOV and incidence angle are contained in [Clarizia and Ruf, 2016] and repeated also in Appendix A.

It is clear from Figure 10 that for a range of incidence angles lower than approximately 54.5°, the IFOV of the collected samples is below the spatial resolution requirement of 25 km. For all these cases, it is possible to average a number of consecutive samples in time to achieve the spatial resolution limit of 25 km. The maximum number of samples n that can be averaged can be easily calculated, using some simplifications, and after a few simple calculations it turns out to be [Clarizia and Ruf, 2016]:

$$n = \frac{EFOV^2}{6\sqrt{IFOV}} - \frac{\sqrt{IFOV}}{6} + 1$$

(5.5)

For each incidence angle, it is therefore possible to know the exact number of samples that can be averaged, if a 25km requirement on the spatial resolution must be met. A pre-computed Lookup Table of n as a function of incidence angle is used in the time-averaging process of the algorithm, and this table is reported in Appendix B. Of course, the calculation of n uses simplified assumptions, and does not take into account several aspects, like the real trajectory of the specular point on the surface, or the real shape of the IFOV, which is ideally assumed to be a square in equation (5.5), but which depends instead on the configuration of the iso-delay and iso-Doppler lines at the ranges selected for the computation of the observables. However, these simplifications are valid enough to allow a reliable estimation of the number of samples to be averaged together.

Some graphical examples of time averaging are illustrated in Figure 30, where the different SP tracks for a single CYGNSS observatory are shown as blue crosses, the sample considered for

time averaging is shown in red, and the consecutive samples to be averaged together with the red one are highlighted with black circles.
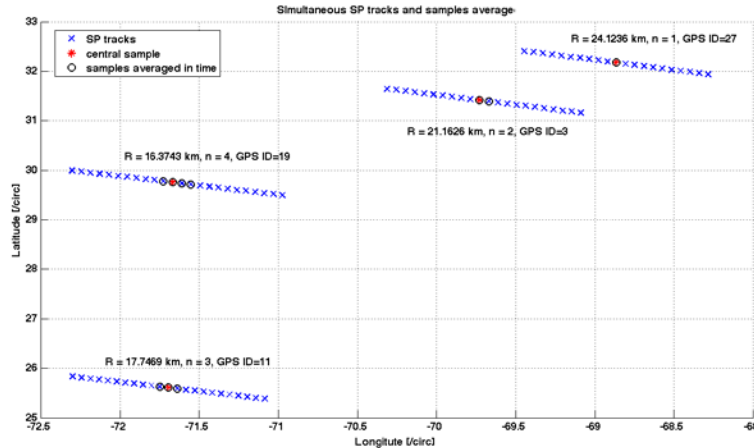

Figure 5. An illustration of how the Time Averaging (TA) algorithm works, for four simultaneous SP tracks acquired by a single CYGNSS observatory.

The time-averaging implementation used in the SOC algorithm follows a few simple rules which mostly aims at avoiding geolocation ambiguities in the resulting time-averaged samples, and that can be summarized as follows:

1) The non-valid samples are excluded from the time averaging. These are either the samples with the overall quality flag set to 0 (see section 6), or those whose DDMA or LES value is not valid (-9999).
2) The number of samples to time-average before the central sample (i.e. preceding the central samples in time along the track) is always equal or higher than the number of samples to time average after the central sample;
3) The difference between number of samples before and number of samples after cannot be greater than 1;

These rules were not applied in the algorithm presented in [Clarizia and Ruf, 2016], and as a result of them there are cases when the number of time-averaged samples (*n*) is less than that prescribed by the incidence angle. As a consequence of that, the amount of time-averaging applied to the SOC algorithm is in some cases less than that applied in [Clarizia and Ruf, 2016].

5.4 Generation of Geophysical Model Function

The dataset used to generate the GMF is formed as the L1b observable, computed over a delay range of [-0.25 0.25] chips and a Doppler range of [-1000 1000] Hz, and acquired with high enough Range-Corrected Gain, or RCG. This is defined as the receiver antenna gain at the specular point multiplied by the range losses there, as given by

$$RCG = \frac{G_{RX}^{SP}}{\left(R_0^{SP} R^{SP}\right)^2} 10^{27}$$

(5.6)

where the range terms are in units of meters and the numeric (not dB) value of the receive antenna gain is used. The scale factor of 10^{27} is applied in order to produce values of order 1-100. The RCG represents those factors affecting the received signal strength which are related to the measurement geometry and are independent of the surface scattering cross section. It is used to filter data according to their expected signal-to-noise quality. Typical ranges of low, medium and high RCG are, respectively, <3, 3-30, and >30.

For GMF generation we use data with a RCG values >3, as the use of "clean" data that do not contain much noise is preferable. The GMF is 2-dimensional, – function of both wind speed and incidence angle.

A GMF is then derived from the high RCG data, to map the selected observables against their ground truth wind speed, which is true wind speed from the simulations, spatially averaged over an area of 25 km x 25 km, centered at the specular point. The GMFs are computed in the form of LUTs of DDMA and LES values corresponding to the 25 km x25 km spatially averaged wind values, and must be a monotonic smooth enough curves that fits well the whole cloud of samples. In our case, we did not force the LUT to be a specific type of function, but we generated it in an empirical manner. The LUT used by the retrieval algorithm is here derived using a tapered approach to the bin widths, starting with very small bin widths for very low or very high winds, since those are the wind ranges with the smallest number of samples. Then, the bin widths gradually increase towards medium winds, i.e. towards wind ranges with the highest number of samples. The LUT function is constructed by taking for each wind bin an average with a triangle weighting of all the true winds falling within that bin as the y-value, and the same type of average of the corresponding DDMA values as the x-value.

Examples of the empirical training data and LUTs for both DDMA and LES samples are shown in Figure 31.
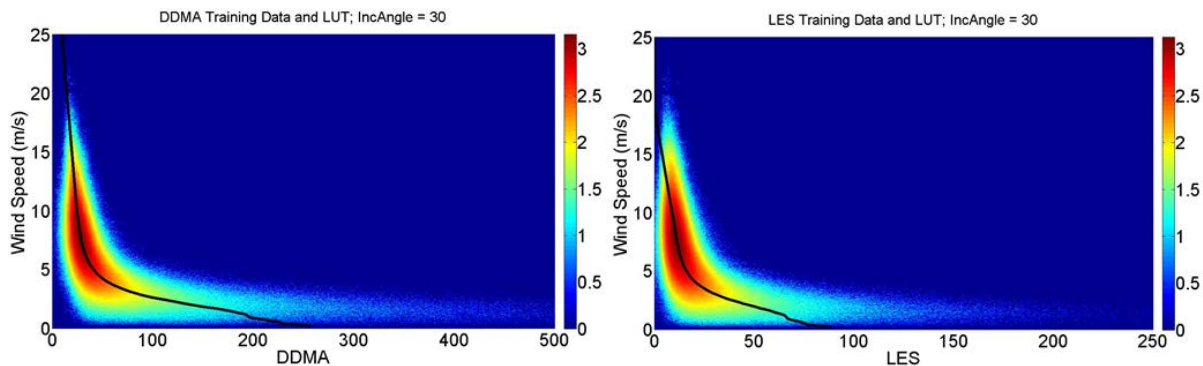

Figure 6. Empirical GMFs for the two Level 1 observables, σ_o or DDMA (top) and LES (bot), at θ_{inc} = 30º, overlaid on log(density) scatter plots of the training data from which they were derived.

5.5   Wind Estimation using the LUT function

For each point of the performance evaluation dataset (either DDMA or LES), a wind speed is estimated using the LUT function. The estimation is done through interpolation, when the observable value falls within the range of values spanned by the LUT, and through extrapolation outside of such range. The mathematical equations to estimate the wind through interpolation is given by (referring to DDMA):

$$\hat{U}_{10}^{DATA} = U_{10}^{LUT} + \alpha \left(DDMA^{DATA} - DDMA \left(U_{10}^{LUT} \right) \right)$$

$$\alpha = \left(V_{10}^{LUT} - U_{10}^{LUT} \right) \Big/ \left(DDMA \left(V_{10}^{LUT} \right) - DDMA \left(U_{10}^{LUT} \right) \right)$$

(5.7)

where $DDMA^{DATA}$ is the DDMA value within the DDMA range of the LUT, $\hat{U}_{10}^{DATA}$ is the wind estimation for DDMA value, the pair of values $\left(U_{10}^{LUT}, DDMA \left(U_{10}^{LUT} \right) \right)$ are the y and x coordinates of the LUT point whose DDMA value (x coordinate) is immediately below $DDMA^{DATA}$, and the pair of values $\left(V_{10}^{LUT}, DDMA \left(V_{10}^{LUT} \right) \right)$ are the y and x coordinates of the LUT whose DDMA value is immediately above $DDMA^{DATA}$.

The mathematical expression for estimating the wind through extrapolation is given by

$$\hat{U}_{10}^{EXT} = U_{10}^{i \, LUT} + \alpha^{EXT} \left(DDMA^{DATA} - DDMA \left(U_{10}^{iLUT} \right) \right) \quad i = 1, n$$

(5.8)

where the values *1* and *n* for the index *i* refer, respectively, to the smallest and largest value of the wind range, $\hat{U}_{10}^{EXT}$ is the wind value to estimate through extrapolation, and α^{EXT} is the slope estimated from the 2 points of the LUT with lowest wind (if *i=1*) or from the 3 points of the LUT with highest wind (if *i=n*).

5.6   Wind Speed MV Estimator

The wind speed estimates from DDMA and LES can be combined together to produce a Minimum Variance (MV) estimator. A MV estimator exploits the degree of decorrelation between the errors in the individual estimates to minimize the RMS error in its wind speed estimate. The advantage of such an estimator lies in the fact that its RMS error will always be better than or equal to the lowest RMS error in the retrieved wind speeds among the individual observables. The lower the correlation between errors in pairs of individual estimators, the better the RMS error performance of the MV estimator. The MV estimator is built as a linear combination of the original estimators, as shown in [Clarizia et al., 2013]:

$$u_{MV} = \boldsymbol{m} \cdot \boldsymbol{u}$$

(5.9)

where u is the vector of individual estimates (from DDMA and LES) and m is the vector of coefficients. The coefficients are obtained by requiring that the MV 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 mathematical derivation of the coefficients for the estimator is illustrated in [Clarizia et al., 2013], [Clarizia and Ruf, 2016] and repeated for convenience in Appendix A.

The covariance matrix could be estimated from all the retrieval errors, however a further improvement in the final performances is obtained when a different covariance matrix of retrieval errors is estimated for different ranges of RCG. This happens because the correlation between the errors in the retrievals from the two observables decreases for noisier data characterized by lower RCG and consequently lower SNR. This allows the MV estimator to contribute more to the performance improvement right for those data where the performances are worse due to a lower RCG, so in a sense the MV approach will help where it is needed most.

Table 1 shows the relative weight that is applied to the DDMA- and LES-based wind speed estimates as they are combined to produce the MV estimate

RCG range	DDMA weight	LES weight
$1 \leq RCG < 3$	0.4370	0.5630
$3 \leq RCG < 5$	0.4025	0.5975
$5 \leq RCG < 10$	0.4358	0.5642
$10 \leq RCG < 20$	0.5540	0.4460
$20 \leq RCG < 30$	0.6251	0.3749
$30 \leq RCG < 70$	0.6474	0.3526
$70 \leq RCG < 110$	0.6805	0.3195
$110 \leq RCG < 150$	0.7151	0.2849
$150 \leq RCG$	0.7450	0.2550

Table 1. Weighting of individual DDMA and LES wind speed estimates used by the Minimum Variance Estimator, for nine different RCG intervals.

A wind speed uncertainty is also associated with each final retrieved wind. The uncertainty values are stored in the form of a LUT indexed by RCG interval and by wind speed interval, and then assigned to each retrieved wind depending on the RCG of the sample and the retrieved wind itself. The numerical values of this LUT are reported in table B4 in Appendix B.

5.7 *CDF-Matching Debias*

A last adjustment is made to the MV wind speed to produce the final 'wind_speed' data product reported in the L2 data files. The MV winds are debiased so their probability distribution, assembled using a global set of measurements over several seasons, matches that of near-coincident wind speeds produced by the ECMWF reanalysis numerical weather prediction model. The debiasing algorithm consists of a re-mapping of the original MV wind speeds to their debiased value. Below 12-15 m/s, the re-mapping forces agreement between CYGNSS MV and ECMWF reanalysis wind speed probability distributions. At higher winds, where sample populations are low and there is less confidence in the representativeness of the empirical

probability distributions, the re-mapping tapers from full adjustment to the ECMWF distribution (at low winds) to no adjustment of the MV winds (at highest retrieved winds). The exact transition wind speed from full to tapered adjustment is determined by the sample size of the wind speed population. Re-mapping parameters are derived independently for each RCG bin used by the MV estimator. The re-mapping parameter LUT is shown in Appendix B.

5.8 *Quality Control Flags in the Retrieval Algorithm*

The retrieval algorithm reports a number of quality control (QC) flags associated with each L2 wind speed estimate. Some are designated as "fatal" and indicate that the reported wind speed is either highly non-physical (e.g. a large negative value) or has an estimated uncertainty that is much higher than npormal. Some are designated as "non fatal" and are triggered by anamolous conditions which may affect the accuracy of the estimate. A complete list of the flags is provided in tabular form in Appendix B.

5.9 *Summary and Conclusions*

Here we summarize the main characteristics of the L2 wind speed retrieval algorithm described here and implemented in the SOC:
  – Generation of observables from L1b DDMs, computed over a fixed DD-range of -0.25 chips to 0.25 chips, and -1 kHz to 1 kHz.
  – Derivation of GMFs under FDS and YSLF sea state conditions to map each observable value into an appropriate wind-speed value given the predominant sea state
  – Time Averaging of observables (FDS only)
  – Wind speed MV Estimator using RCG-dependent coefficients, derived from the covariance matrix of retrieval errors (FDS only);

It is important to also highlight some differences with respect to the algorithm illustrated in [Clarizia and Ruf, 2016]. These main differences are:
  – The time-averaging process follows here some additional rules listed in 5.3, and as a result of that the number of time-averaged observables is in some cases less than that in [Clarizia and Ruf, 2016]
  – No EFOV filter is applied here, or in other words the present algorithm does not exclude samples acquired with an incidence angle greater than 54.5°.

6 References

[Arnold-Bos et al., 2007] A. Arnold-Bos, A. Khenchaf, and A. Martin, "Bistatic radar imaging of the marine environment — Part I: theoretical background," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 11, pp. 3372–3383, Nov 2007.

[Arnold-Bos et al., 2007b] A. Arnold-Bos, A. Khenchaf, and A. Martin, "Bistatic radar imaging of the marine environment — Part II: simulation and result analysis," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 11, pp. 3384–3396, Nov 2007.

[Barrick, 1968] Barrick, D. E., "Relationship between slope probability density function and the physical optics integral in rough surface scattering," *Proc. IEEE*, vol. 56, pp. 1728–1729, 1968.

[Bass and Fuks, 1979] F. G. Bass and I. M. Fuks, *Wave Scattering From Statistically Rough Surfaces*. New York: Pergamon, 1979.

[Brown, 1978] Brown, G.S., "Backscattering from a Gaussian-distributed, perfectly conducting rough surface," *IEEE Trans. Antennas Propag.*, Vol. AP-26, pp. 472–482, May 1978.

[Cardellach and Rius, 2008] Cardellach, E., and A. Rius, "A new technique to sense non-Gaussian features of the sea surface from L-band bistatic GNSS reflections," *Remote Sens. Environ.*, vol. 112, pp. 2927–293, 2008.

[Clarizia et al., 2009] Clarizia, M., Gommenginger, C., Gleason, S., Srokosz, M., Galdi, C., di Bisceglie, M., "Analysis of GNSS-R delay-Doppler maps from the UK-DMC satellite over the ocean", Geophys. Res. Lett. 36, 2009.

[Clarizia, 2012] Clarizia, M. P., "Investigating the Effect of Ocean Waves on GNSS-R Microwave Remote Sensing Measurements", PhD Thesis, University of Southampton, September 2012.

[Clarizia et al., 2013] Clarizia, M.P., Gommenginger C., "Spaceborne GNSS-Reflectometry Instrument and Algorithms: investigating mean square slope retrieval with UK-DMC GNSS-R data", Technical Note, ESTEC contract no. 4000106450/12/NL/FF/LF, August 2013.

[Clarizia et al., 2014] Clarizia, M. P., C. Ruf, C. Gommenginger and P. Jales, "Spaceborne GNSS-R Minimum Variance Wind Speed Estimator", in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6829-6843, 2014.

[Clarizia and Ruf, 2016] Clarizia, M.P., and C. S. Ruf, "Wind Speed Retrieval Algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) Mission," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4419-4432, 2016, doi: 10.1109/TGRS.2016.2541343.

[Cox and Munk, 1954] Cox, C., and W. Munk, "Measurement of the roughness of the sea surface from photographs of the Sun's glitter," *J. Opt. Soc. Am.,* Vol. 44, 1954, pp. 835–850.

[Elachi, 1988] Elachi, C., *Spaceborne Radar Remote Sensing: Applications and Techniques*, IEEE Press, New York, 1988.

[Elfouhaily et al., 1997] Elfouhaily, T., Chapron, B., Katsaros, K., Vandemark, D., "A unified directional spectrum for long and short wind-driven waves", Journal of Geophysical Research 102(C7), 15781–15796, 1997.

[Elfouhaily et al., 2002] Elfouhaily T., D.R. Thompson and L. Lindstrom. "Delay-Doppler analysis of bistatically reflected signals from the ocean surface: theory and application," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, No. 3, March 2002.

[Elfouhaily and Guerin, 2004] T. Elfouhaily and C. A. Guérin, "A critical survey of approximate scattering wave theories from random rough surfaces," *Waves Random Media*, vol. 14, pp. R1–R40, 2004.

[Fan et al., 2009] Fan Y., I. Ginis, T. Hara, C.W. Wright, and E. J. Walsh, "Numerical simulations and observations of surface wave fields under an extreme tropical cyclone," *J. Phys. Oceanogr.*, vol. 39, pp. 2097–2116, Sep. 2009.

[Fung et al., 2001] A. K. Fung, C. Zuffada, and C. Y. Hsieh, "Incoherent bistatic scattering from the sea surface at L band," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 5, pp. 1006–1012, May 2001.

[Garrison et al., 1998] Garrison, J., Katzberg, S., Hill, M., "Effect of sea roughness on bistatically scattered range coded signals from the global positioning system", Geophysical Research Letters 25(13), 2257–2260.

[Garrison et al., 2002] Garrison, J., Komjathy, A., Zavorotny, V., S.J.Katzberg, "Wind speed measurements using forward scattered GPS signals", IEEE Transactions on Geoscience and Remote Sensing 40 (1), 50–65, 2002.

[Germain et al., 2004] Germain, O., Ruffini, G., Soulat, F., Caparrini, M., Chapron, B., Silvestrin, P., "The Eddy experiment: GNSS-R speculometry for directional sea-roughness retrieval from low altitude aircraft" Geophys. Res. Letters 31 (L12306), 2004.

[Gleason et al., 2005] Gleason, S., Hodgart, S., Yiping, S., Gommenginger, C., Mackin, S., Adjrad, M., Unwin, M., "Detection and processing of bistatically reflected GPS signals from Low-Earth Orbit, for the purpose of ocean remote sensing", IEEE Transactions on Geoscience and Remote Sensing 43 (6), 1229–1241, 2005.

[Gleason, 2006] Gleason, S. T., "Remote sensing of ocean, ice and land remote sensing using bistatically scattered GNSS signals from low earth orbit", PhD Thesis, University of Surrey, 2006.

[Gleason et al., 2016] Gleason, S., C., S. Ruf, M. P. Clarizia and A. J. O'Brien, "Calibration and Unwrapping of the Normalized Scattering Cross Section for the Cyclone Global Navigation Satellite System," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2495-2509, 2016, doi: 10.1109/TGRS.2015.2502245.

[Gleason et al., 2018] Gleason, S., C. S. Ruf, A. O'Brien, D. S. McKague, "The CYGNSS Level 1 Calibration Algorithm and Error Analysis Based On On-Orbit Measurements," IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., doi: 10.1109/JSTARS.2018.2832981, 2018.

[Guerin et al., 2010] C.-A. Guérin, G. Soriano, and B. Chapron, "The weighted curvature approximation in scattering from sea surfaces," *Waves Random Complex*, vol. 20, no. 3, pp. 364–384, Aug. 2010.

[Johnson, 2005] J. T. Johnson, "A study of ocean-like surface thermal emission and reflection using Voronovich's small slope approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 306–314, Feb. 2005.

[Johnson and Elfouhaily, 2007] Johnson, J.T. and T. M. Elfouhaily, "Computation of oceanlike surface thermal emission and bistatic scattering with the reduced local curvature approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2108–2115, Jul 2007.

[Katzberg et al., 2000] Katzberg, S. J., J.L. Garrison, "Wind speed retrieval of GPS surface reflection data using a matched filter approach, *paper presented at Sixth International Conference, 2000.*

[Katzberg et al., 2006] Katzberg, S. J., O. Torres, and G. Ganoe, Calibration of reflected GPS for tropical storm wind speed retrievals, *Geophys. Res. Lett.*, 33, L18602, doi:10.1029/2006GL026825, 2006.

[Klein and Swift, 1977] L. A. Klein and C. T. Swift, "An improved model for the dielectric constant of sea water at microwave frequencies," *IEEE Trans. Antennas Propag.*, AP-25(1): pp. 104–11, Jan 1977.

[Komjathy et al., 2004] Komjathy, A., Armatys, M., Masters, D., Axelrad, P., Zavorotny, V., Katzberg, S., "Retrieval of ocean surface wind speed and wind direction using reflected GPS signals", Journal of Atmospheric and Oceanic Technology 21, 515–526, 2004.

[Marchan-Hernandez et al., 2008] Marchan-Hernandez, J., Rodriguez-Alvarez, N., Camps, A., Bosch-Lluis, X., Ramos- Perez, I., Valencia, E., "Correction of the sea state impact in the L-band bright- ness temperature by means of delay-doppler maps of global navigation satellite signals reflected over the sea surface", IEEE Transactions on Geoscience and Remote Sensing 46 (10), 2914–2923, 2008.

[Marchan-Hernandez et al., 2010] Marchan-Hernandez, J., Valencia, E., Rodriguez-Alvarez, N., Ramos-Perez, I., Bosch- Lluis, X., Camps, A., Eugenio, F., Marcello, J., "Sea-state determination using GNSS-R data", IEEE Geoscience and Remote Sensing Letters 7(4), 621–625, 2010.

[Nolan et al., 2013]: David S. Nolan, Robert Atlas, Kieran T. Bhatia, and Lisa R. Bucci, "Development and Validation of a Hurricane Nature Run Using the Joint OSSE Nature Run and the WRF Model", *Journal of Advances in Modeling Earth Systems* (accepted for publication), April 2013.

[O'Brien, 2014] O'Brien, A., "End-to-End Simulator Technical Memo", CYGNSS Technical Document, 2014;

[Parkinson et al., 1996] Parkinson B.W., J.J. Spilker, P. Axelrad, and P. Enge (eds.) *Global Positioning System: Theory and Applications,* Vol. I & II, AIAA, Washington, DC, 1996.

[Rodriguez-Alvarez et al., 2013] Rodriguez-Alvarez, N., D.M. Akos, V.U. Zavorotny, J.A. Smith, A. Camps, and C.W. Fairall, "Airborne GNSS-R wind retrievals using delay-Doppler maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 626–641, Jan. 2013.

[Ruf, 2013] Ruf, C. S., "DDM Data Compression and Decimation Algorithm", *SPRL Technical Memorandum No. 148-0046-X1*, May 2013.

[Ruf and Balasubramaniam, 2018] Ruf, C., R. Balasubramaniam, "Development of the CYGNSS Geophysical Model Function for Wind Speed," IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., doi: 10.1109/JSTARS.2018.2833075, 2018.

[Ruf et al., 2018] Ruf, C., S. Gleason, D. S. McKague, "Assessment of CYGNSS Wind Speed Retrieval Uncertainty," IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., doi: 10.1109/JSTARS.2018.2825948, 2018.

[Soriano and Guerin, 2008] G. Soriano and C-A. Guérin, "A cutoff invariant two-scale model in electromagnetic scattering from sea surfaces," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 199-203, Apr 2008.

[Soulat, 2004] Soulat, F., *Sea Surface Remote Sensing With GNSS and Sunlight Reflections*. Ph.D. Thesis. Universitat Politecnica de Catalunya, 2004.

[Thompson et al, 2000] Thompson, D.R., T. M. Elfouhaily, and R. F. Gasparovic, "Polarization dependence of GPS signals reflected from the ocean," In: *Proc. IGARSS'00*, Honolulu, HI, pp. 3099-3101, 2000.

[Thompson et al., 2005] Thompson, D. R., T. M. Elfouhaily, and J. L. Garrison, "An improved geometrical optics model for bistatic GPS scattering from the ocean surface, *IEEE Trans. Geosci. Remote Sens.*, vol. 43, No. 12, pp. 2810–2821, 2005.

[Tolman, 1998] Tolman, H. L., "Validation of a new global wave forecast system at NCEP." *Ocean Wave Measurements and Analysis*, B. L. Edge and J. M. Helmsley, Eds., ASCE, pp. 777–786, 1998.

[Valencia et al., 2013] Valencia, E. V. U. Zavorotny, D. M. Akos, and A. Camps, "Using DDM asymmetry metrics for wind direction retrieval from GPS ocean-scattered signals in airborne experiments," *IEEE Trans. Geosci. Remote Sens.*, 10.1109/TGRS.2013.2278151, 2014 (in press).

[Voronovich, 1994] Voronovich, A.G., "Small-slope approximation for electromagnetic wave scattering at a rough interface of two dielectric half-spaces," *Waves Random Media,* vol. 4, pp. 337–367, 1994.

[Voronovich, 1999] Voronovich, A.G., *Wave Scattering from Rough Surfaces,* 2nd edn (Berlin: Springer), 236p, 1999.

[Voronovich and Zavorotny, 2001] A. G. Voronovich and V. U. Zavorotny, "Theoretical model for scattering of radar signals from rough sea-surface with breaking waves at Ku- and C-bands," *Waves Random Media*, vol. 11, pp. 247–269, 2001.

[Voronovich and Zavorotny, 2014] Voronovich, A.G., and V. U. Zavorotny, "Full-polarization modeling of monostatic and bistatic radar scattering from a rough sea surface," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1362-1371, Mar 2014.

[You et al., 2004] You, H. G. Heckler, J. L. Garrison, and V. U. Zavorotny, Stochastic voltage model and experimental measurement of ocean-scattered GPS signal statistics, *IEEE Trans Geosci. Remote Sens.*, vol. 42, No. 10, pp.2160–2169, Oct. 2004.

[You et al., 2006] You, H., J. L. Garrison, G. Heckler, and D. Smajlovic, "The autocorrelation of delay-Doppler waveforms generated from ocean-scattered GPS signals," *IEEE Geosci. Remote Sens. Lett.*, vol. 3(1), pp.78-82, 2006.

[Zavorotny and Voronovich, 2000] Zavorotny, V., Voronovich, A., "Scattering of GPS signals from the ocean with wind remote sensing applications", IEEE Transactions on Geoscience and Remote Sensing 38 (2), 951–964, 2000.

[Zuffada and Zavorotny, 2001] Zuffada, C., and V. Zavorotny, "Coherence time and statistical properties of the GPS signal scattered off the ocean surface and their impact on the accuracy of remote sensing of sea surface topography and winds," In: *Proc. IGARSS'01,* pp. 3332–3334, 2001.

[Zuffada et al., 2004] Zuffada, C., Fung, A., Parker, J., Okolicanyi, M., Huang, E., "Polarization properties of the GPS signal scattered off a wind-driven ocean", IEEE Transactions on Antennas and Propagation 52 (1), 172–187, 2004.

Appendix A: Emprical Generation of Geophysical Model Function and Retrieval Performance Assessment Using On-orbit Observations

The description presented here of how the GMFs connecting L1 observations to L2 wind speed were generated is extracted from the more detailed description provided in [Ruf and Balasubramaniam, 2018].

Geophysical Model Functions (GMFs) are developed which map the Level 1 observables made by the Cyclone Global Navigation Satellite System (CYGNSS) radar receivers to ocean surface wind speed. The observables are: 1) the normalized bistatic radar cross section (σ_o) of the ocean surface; and 2) the slope of the leading edge of the radar return pulse scattered by the ocean surface. GMFs are empirically derived from measurements by CYGNSS which are nearly coincident with independent estimates of the 10 meter referenced ocean surface wind speed (u_{10}). Two different sources of "ground truth" wind speed are considered – numerical weather prediction model outputs and measurements by the NOAA P-3 hurricane hunter during eyewall penetrations of major hurricanes. The GMFs derived in each case have significant differences that are believed to result from differences in the state of development of the long wave portion of the ocean surface height spectrum that result from characteristic differences in wave age and fetch length near vs. far from a hurricane.

The two CYGNSS receive antenna beams are pointed cross track to the direction of orbital motion. Each antenna is a 2x3 element phased array with a fan beam antenna pattern. The patterns cover incidence angles of approximately 5–65 deg and azimuth angles of 75–105 deg and 255–285 deg on the port and starboard sides of the sub-satellite point, respectively. DDMs are processed into two observables which are used for wind speed retrieval. One is the normalized scattering cross section (σ_o) averaged over an area roughly 25 km in diameter centered on the specular point. The other is the leading edge of the slope of the radar return waveform (LES). Both observables are defined in [Clarizia and Ruf, 2016], which describes the wind speed retrieval algorithm used by the CYGNSS mission and develops an initial, pre-launch, GMF based on simulated measurements.

An empirical GMF is developed here for use by the CYGNSS wind speed retrieval algorithm. It uses near-coincident matchups between the CYGNSS observables and independent estimates of the ocean surface wind speed referenced to a 10 m height (u_{10}). The independent wind speeds are provided by Numerical Weather Prediction (NWP) models at low to moderate wind speeds and by instruments on the NOAA P-3 hurricane hunter aircraft at higher wind speeds in tropical cyclones [24]. The GMF developed here is distinct from those reported previously in two primary respects. With previous airborne applications, the observables were either uncalibrated Signal-to-Noise Ratio (SNR) measurements of the received GPS signal, or normalized measures of the extent of diffuse scattering away from the specular point. The GMF developed here instead uses as its observables the absolutely calibrated σ_o of the surface in the vicinity of the specular and the LES. With previous spaceborne applications, coincident matchup measurements were not available at hurricane force wind speeds from the NOAA P-3 hurricane hunter aircraft. Their availability allows for the extension of the GMF to significantly higher wind speeds than

has been reported previously. The GMF development is presented in two stages. First, a Fully Developed Seas (FDS) version is developed based on matchups with NWP model outputs at low to moderate wind speeds. Then a Young Seas/Limited Fetch (YSLF) version is presented based on matchups with measurements by hurricane hunter aircraft during flights through several 2017 Atlantic hurricanes.

A.1   Fully Developed Seas GMF

The FDS GMF is based on an empirical pairing of CYGNSS Level 1 (L1) observations of σ_o and LES with the 10 meter referenced ocean surface wind speed(u_{10}), as determined by Numerical Weather Prediction (NWP) models. A large population of these pairings is partitioned into "bins" with respect to u_{10} and the incidence angle (θ_{inc}) of the observation. The average values of u_{10} and the L1 observable within a bin are paired together as one discrete sample of the GMF for that observable. An algebraic parametric model is then fit to the discrete GMF samples to produce a continuously varying GMF which is used by the Level 2 (L2) wind speed retrieval algorithm.

A.1.1   Description of Training Data Set: ECMWF & GDAS matchups

Matchup NWP data used to train the empirical GMF are the 10 meter referenced ocean surface wind speeds provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Global Data Assimilation System (GDAS). ECMWF is an independent intergovernmental organization supported by 34 countries to produce current weather forecasts and climate reanalysis products and to perform related research activities [25]. The reanalysis products are used here. GDAS is a system operated by the NOAA National Centers for Environmental Information to organize a variety of surface, balloon, aircraft and spaceborne observations into a gridded model space for use by NOAA's global forecast system [26]. ECMWF and GDAS use a 0.25 deg and 1.0 deg reporting interval, respectively. Bilinear interpolation in space and linear interpolation in time of the reported NWP products are used to estimate u_{10} at the times and locations of the CYGNSS specular point observations.

A merged "ground truth" u_{10} product is used which combines model outputs by both ECMWF and GDAS. For wind speeds below 20 m/s, the ECMWF value for u_{10} is used alone. For wind speeds between 20 and 25 m/s, the arithmetic average of ECMWF and GDAS wind speeds is used. For wind speeds above 25 m/s, the GDAS value alone is used. This merged-product approach is used to accentuate the better accuracy of ECMWF at lower wind speeds and of GDAS at higher wind speeds [27]. In addition, matchups are only used if the u_{10} values for ECMWF and GDAS differ by no more than 3 m/s.

Examples of the training data set are shown in Fig. A1 for the σ_o observable at three values of θ_{inc}. Shown are logarithmic density scatterplots of the merged ground truth u_{10} values vs. the observable. The general trend is for the scattering cross section to decrease as wind speed increases, as expected for bistatic forward scattering from a wind-roughened ocean surface.
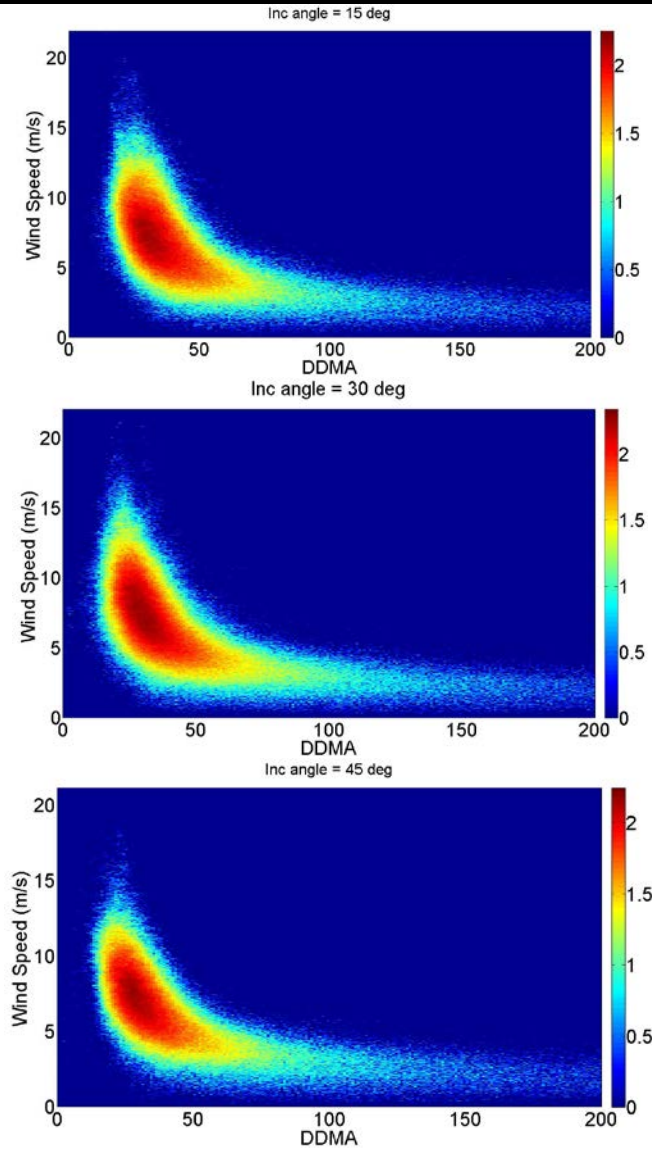
Fig. A1. Log(density) scatterplots of σ_o measured by CYGNSS vs. ground truth u_{10} at incidence angles of 15º (top), 30º (mid) and 45º (bot). The color scale is the log10 of the number density of points.

A.1.2. Binning of matchups for discrete, empirical GMF

The CYGNSS L1 observables are filtered prior to use as part of the training used to derive the empirical GMF. The filters are for reasons of quality control. Specifically:

- The Doppler coordinate of the specular point in the DDM is required to be greater than the lowest possible value in the map and less than the highest possible value. This discards cases where it is at the edge of the map and the computed Doppler coordinate may be incorrect. In practice, this happens less than 0.1% of the time.
- The delay coordinate of the specular point in the DDM is required to be greater than the

lowest possible value in the map and less than the highest possible value. This discards cases where it is at the edge of the map and the computed delay coordinate may be incorrect. In practice, this happens less than 0.1% of the time.

- All NaN values of the observables are discarded. This eliminates samples for which noise in the calibration data can produce non-physical calibrated L1 data. In practice, this happens less than 0.1% of the time.
- The observables are required to be non-negative. This eliminates samples for which noise in the calibration data can produce non-physical calibrated L1 data, as well as measurements which are very close to the measurement noise floor. In practice, this happens less than 0.1% of the time.
- All measurements are discarded for which the spacecraft star tracker is not tracking due to solar contamination. Some reported spacecraft attitude data during sun outages are known to be erroneous (with inaccuracies greater than the error allocation in the L1 calibration algorithm for attitude knowledge). This only occurs when the outage is especially long, but all sun outage data are flagged and removed as a precaution. In practice, this happens less than 1% of the time.
- All data with a CYGNSS Range Corrected Gain (RCG) of less than 10 are discarded. RCG is a composite measure of receive signal strength that combines the receive antenna gain in the direction of the specular point with the R^{-2} propagation range loss from the GPS transmitter to the specular point and from the specular point to the CYGNSS receiver. In practice, data with an RCG > ~1 can typically produce useful wind speed retrievals, but only data with a higher SNR are used to train the empirical GMF.
- All observations resulting from transmissions by the GPS Block Type II-F satellites are discarded. Block II-F is the newest family of GPS satellites, and the one for which the CYGNSS team has the least information about its transmitter antenna gain pattern. There are currently 8 II-F satellites in the constellation, out of 31 total in operation.

The behavior of the empirical GMF as a function of u_{10} and θ_{inc} is smoothed by allowing sequential bins in either dimension to overlap. In the incidence angle dimension, the bin center is incremented every 1 deg from 1 to 70 deg and all samples are included within +/- 2.0 deg of the center. In the wind speed dimension, the bin center is incremented every 0.1 m/s from 0.05 to 34.95 m/s and all samples are included within a variable bin width that varies according to the population density of samples as a function of wind speed. Specifically, the bin widths used are:

- +/- 0.4 m/s ($u_{10} < 2$ m/s)
- +/- 0.3 m/s ($2 < u_{10} < 5$ m/s)
- +/- 0.2 m/s ($5 < u_{10} < 9$ m/s)
- +/- 0.4 m/s ($9 < u_{10} < 11$ m/s)
- +/- 0.6 m/s ($11 < u_{10} < 14$ m/s)
- +/- 0.8 m/s ($14 < u_{10} < 17$ m/s)
- +/- 1.0 m/s ($17 < u_{10}$)

A weighted average of all samples within twice these bin width ranges is performed. Samples within +/- one bin width of the bin center are given twice as much weight as those between

1xbinwidth and 2xbinwidth from the bin center. This tapered weighting approach reduces the introduction of artificial higher frequency components into the GMF than are present in the original discrete empirical samples.

The GMF is also forced to be monotonic as a function of wind speed. The GMF value at 7.05 m/s is computed first (since this is generally the most probable wind speed and so has the largest population of samples in its near vicinity). GMF values are then sequentially computed in steps of 0.1 m/s above and below this value using the averaging scheme described above. However, values are allowed to either decrease or stay the same with increasing wind speed, and increase or stay the same with decreasing wind speed. This limits the introduction of non-physical variations into the GMF due to undersampling of certain parts of the (wind speed, incidence angle) state space. In practice, this monotonicity algorithm is only enforced at the highest and lowest wind speeds in the population, where the sampling density tends to be lowest.

Examples of the empirical GMFs for both observables (σ_o and LES) at θ_{inc} = 30 deg, overlaid on the training data from which they were derived, are shown in Fig. A2. Over the central range of wind speeds where most of the samples occur, the GMF agrees with the highest density part of the scatter plot. At the highest and lowest wind speeds, the lower size of the population makes the behavior of the GMF more susceptible to errors.
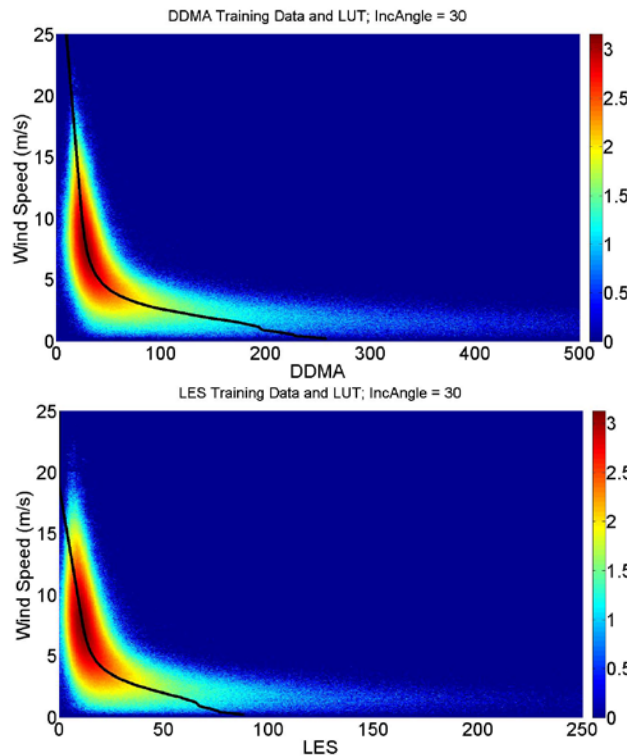

Fig. A2. Empirical GMFs for the two Level 1 observables, σ_o or DDMA (top) and LES (bot), at $\theta_{inc} = 30^o$, overlaid on log(density) scatter plots of the training data from which they were derived.

Examples of the empirical GMFs for both L1 observables across a range of incidence angles are shown in Fig. A3. The general dependence of observable on wind speed is consistent across all incidence angles. The dependence on incidence angle at a given wind speed is also consistent. Note that the slope of the GMF ($dObs/du_{10}$) is highest at low wind speeds, indicating that wind speed retrievals will, in general, perform better and be less susceptible to measurement noise and calibration errors at the lower wind speeds. The general behavior of the empirical σ_o GMF, both as a function of wind speed and of incidence angle, is consistent with scattering model predictions based on the first order small slope approximation method [28], [29].
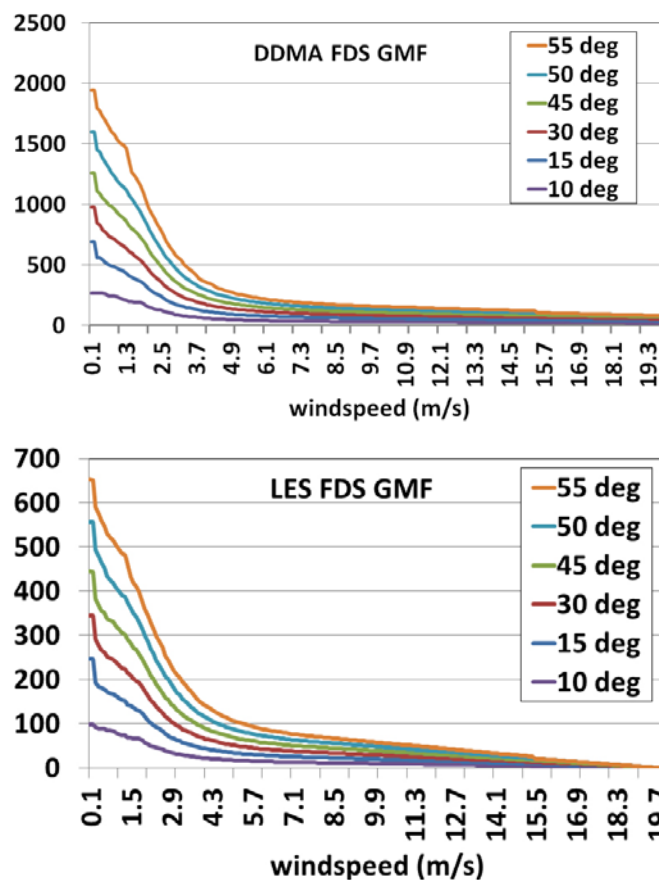

Fig. A3. Empirical GMFs for σ_o (top) and LES (bot) at $\theta_{inc}=10\ 15, \ldots,55^o$.

A.1.3. *Parametric model for continuous, empirical GMF*

An algebraic parametric model is fit in a least-squares sense to the empirical GMF in order to populate the lookup tables used by the CYGNSS Level 2 wind speed retrieval algorithm. This process smooths across some of the behavior in the empirical GMF that is related to measurement noise and insufficient number of samples in the training set. It also interpolates across portions of (wind speed, incidence angle) state space that are not sufficiently populated by the training set. The parametric model assumed for the GMF is divided into two portions based

on the observed behavior of the empirical GMF as a function of wind speed. At low wind speeds, a model is assumed of the form

$$Obs = a_0 + a_1 u^{-1} + a_2 u^{-2} \tag{A.1}$$

where *Obs* is the Level 1 CYGNSS observable (either σ_o or LES), *u* is the ground truth u_{10} wind speed, and a_i are the dependent parameters of the model. At high wind speeds, a model is assumed of the form

$$Obs = b_0 + b_1 u + b_2 u^2 \tag{A.2}$$

where b_i are the dependent parameters of the model. The population of samples used to train these two models is different for each observable. For σ_o, samples at wind speeds below 15 m/s are used to determine a_{0-2} and samples above 15 m/s are used to determine b_{0-2}. For the LES observable, samples below and above 10 m/s are used. The actual transition from one parametric model to the other in the GMF occurs near, but not exactly at, these wind speed values. A transition point is chosen where the first derivatives of the two models are equal (i.e. a spline fit). The model parameters and the spline fit transition point are chosen independently at each incidence angle.

Examples of the parametric model GMF, together with the empirical GMF from which they are obtained, are shown in Fig. A4. At the lowest wind speeds (below ~2 m/s), the empirical and parameteric models tend to deviate (more so for the LES observable). The sensitivity of the LES observable to wind speed drops to zero at wind speeds above ~18 m/s, whereas the σ_o observable retains its sensitivity up to wind speeds of ~30 m/s.
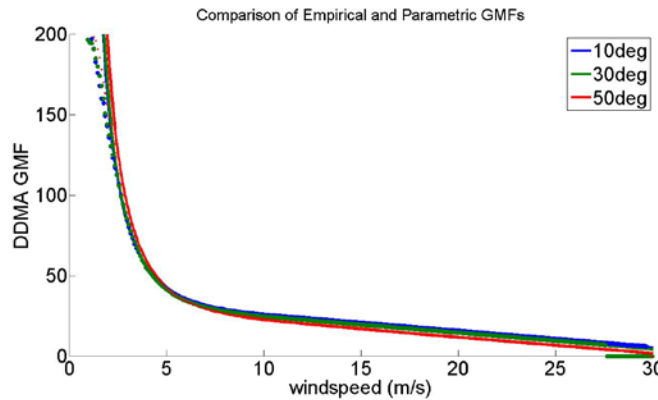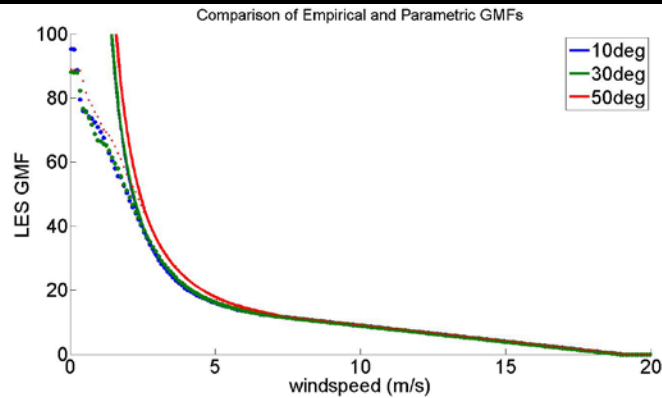
Fig. A4. Discrete empirical GMFs (symbols) and continuous parametric models (solid lines) derived from them, for the two L1 observables σ$_o$ (top) and LES (bot), and for θ$_{inc}$ = 10, 30, 50°.

Examples of the parametric model GMFs for both observables and over a range of incidence angles are shown in Fig. A5. The truncation of the GMF at high wind speeds results from the limitations in the dynamic range of wind speeds included in the training data set.
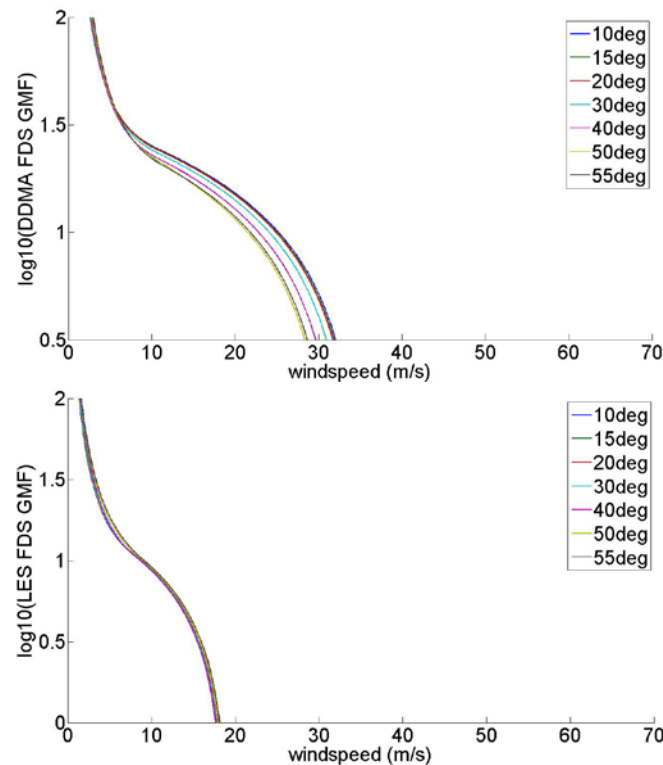

Fig. A5. Parametric model FDS GMF for the Level 1 observables σ$_o$ (top) and LES (bot) at incidence angles of 10, 15, …, 55 deg. The dependence on incidence angle is more pronounced in the case of the σ$_o$ (DDMA) observable. The maximum wind speed at which the observable is sensitive to changes in wind speed is also higher for σ$_o$ than for the LES observable.

A.1.4. *Validation and performance characterization*

As a means of assessing whether the derived GMFs properly represent the response of the Level 1 observables to changes in ocean surface wind speed, they are used as the basis for a wind speed retrieval algorithm. The algorithm inverts the GMF to estimate wind speed given the measured observable [19]. The error in this retrieval algorithm (groundtruth – retrieval) is considered as a function of the two coordinates of the GMF, incidence angle and wind speed. Note that this co;mparison uses the same population of data as was used to train the GMFs and so should be considered a test of internal consistency in the generation of the GMFs. A more independent assessment of retrieval performance is presented in [30].

The dependence of retrieval error on incidence angle is shown in Fig. A6. The σ_o-based retrieval on the left shows a positive statistical retrieval bias (retrieved values are larger than groundtruth more often than they are smaller). The LES-based retrieval on the right shows a more unbiased distribution of retrieval errors. Notably, in terms of incidence angle dependence, the highest density of retrieval errors occurs near an error of zero and this is true at all incidence angles and for both L1 observables.
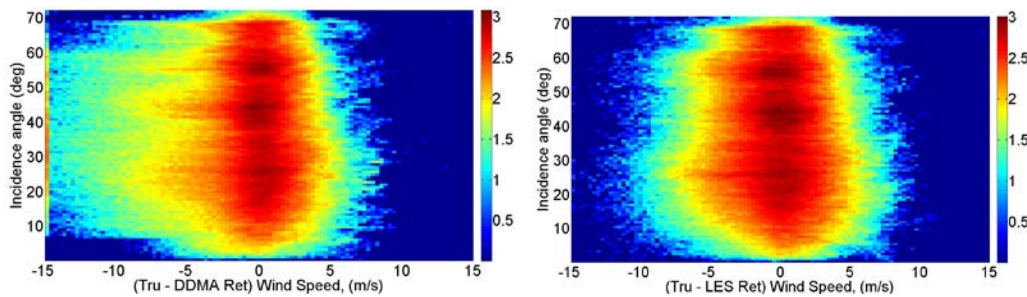

Fig. A6. Log(density) scatter plot of wind speed retrieval error vs. incidence angle for retrievals using the σ_o (left) and LES (right) Level 1 observable. There is no significant dependence on incidence angle.

The dependence of retrieval error on the ground truth wind speed is shown in Fig. A7. Significant positive retrieval biases (retrieved values are larger than groundtruth more often than they are smaller) can be seen at ground truth wind speeds of 5-15 m/s for σ_o and at 3-10 m/s for LES. A possible cause for this behavior, and a corresponding mitigation strategy, are considered next.
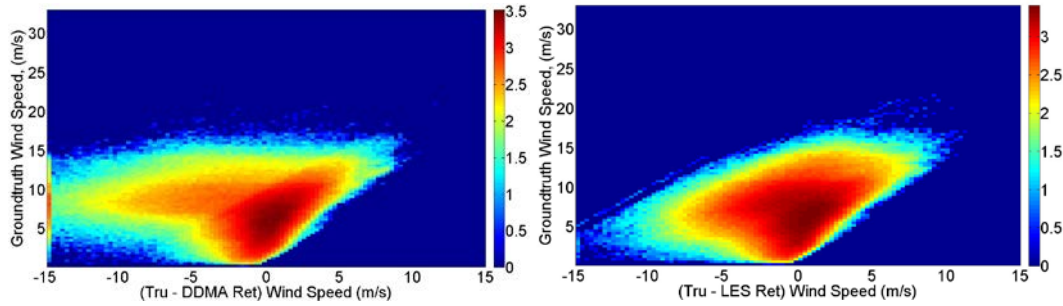
Fig. A7. Log(density) scatter plot of wind speed retrieval error vs. ground truth wind speed for retrievals using the σₒ (left) and LES (right) L1 observable.

Fig. A8 shows the dependence of retrieval error on the difference between the wind speeds retrieved using the two L1 observables. Larger retrieval errors tend to be highly correlated with the difference between the two retrievals. The large positive bias in retrievals based on the σₒ observable tends to coincide with cases where the σₒ retrieval is much larger than that from the LES observable. Alternately, samples for which the σₒ retrieval is much smaller than the LES retrieval tend to coincide with large positive biases in the LES retrieval. The root cause of this behavior may be related to the fact that the two observables respond to different aspects of the sea state, only part of which is forced by local wind speed. If, for example, they respond in different ways to the longer swell portion of the surface wave spectrum, this could explain their different dependence on the retrieval error. One hypothesis is that young sea conditions (not fully developed) may coincide with instances where the two retrievals have significant differences.
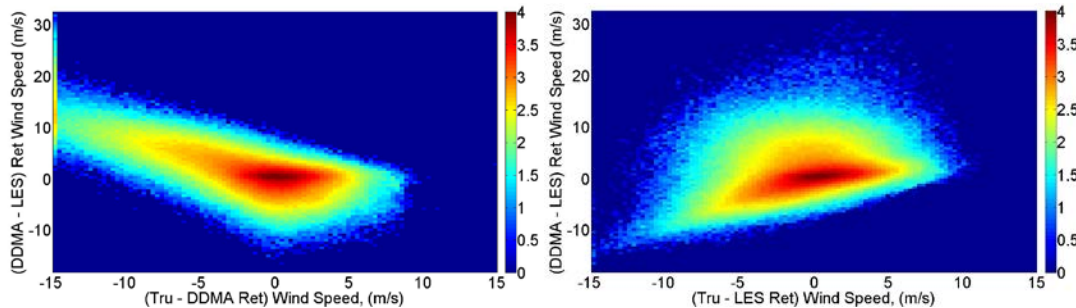

Fig. A8. Log(density) scatter plot of wind speed retrieval error using the σₒ (left) and LES (right) L1 observable vs. the difference between the two retrieved wind speeds.

In terms of mitigation of this behavior, and improvement in the overall wind speed retrieval performance, the dependence of retrieval error on the difference between the σₒ and LES retrieved wind speeds can be used as a quality control (Q/C) filter. This is illustrated in Fig. A9, which shows the dependence of the retrieved Minimum Variance (MV) wind speed on the difference. The MV retrieval is a weighted average of the σₒ and LES retrievals, weighted by the inverse variance of the error in wind speed retrieved by each of the individual observables [19]. Large errors in the MV retrieval can be seen in Fig. 9 to correlate with large differences between the σₒ and LES retrievals. A simple Q/C filter could, for example, exclude all retrievals for which the difference is greater than 6 m/s. This filter threshold is illustrated in Fig. A9. In

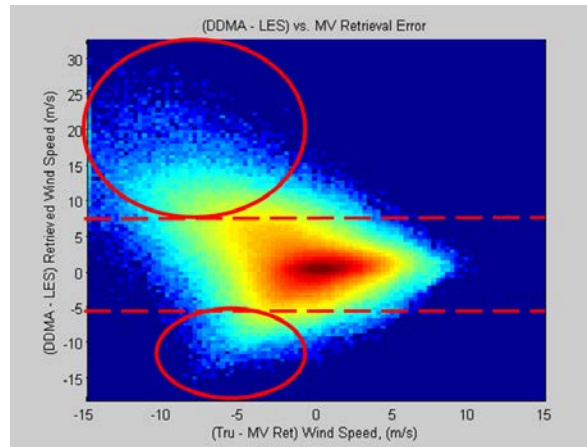practice, this Q/C filter discards approximately 4% of the samples.

Fig. A9. Log(density) scatter plot of minimum variance wind speed retrieval error vs. the difference between the wind speeds retrieved by the two individual L1 observables. A Q/C filter that discards retrievals with differences greater than 6 m/s (shown by the two dashed red lines) will eliminate the circled regions with large retrieval errors.

The effectiveness of the Q/C filter, and the overall quality of the MV retrieval algorithm, is illustrated in Fig. A10, which compares the groundtruth and retrieved wind speeds as a scatter plot and by their RMS and mean difference. The large retrieval biases evident in Fig. A7 have been largely removed by this Q/C filter. Fig. A10 also illustrates the performance of the wind speed retrieval beow 20 m/s. The RMS difference is ~2 m/s at low wind speeds and grows to ~4 m/s at 20 m/s.
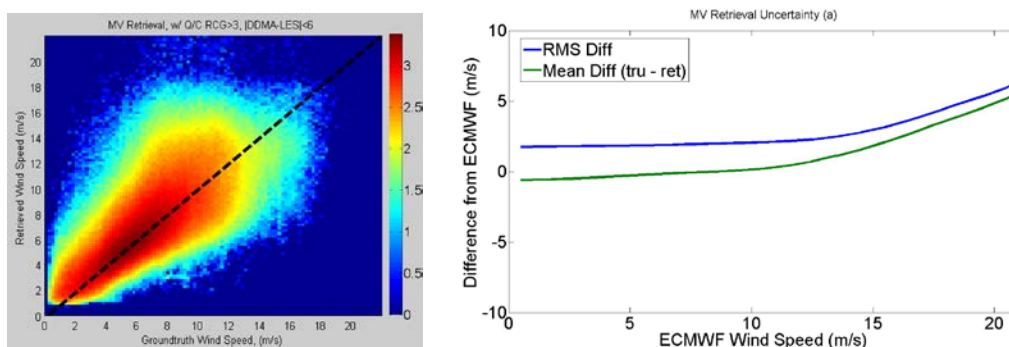

Fig. A10. Log(density) scatter plot of groundtruth vs. MV retrieved wind speed (left) with black-dashed line of 1:1 agreement, and of RMS and mean retrieval error vs. groundtruth wind speed (right).

A.2   Young Seas/Limited Fetch GMF

The YSLF GMF is based on matchups between measurements by CYGNSS made during overpasses of 2017 Atlantic hurricanes and near-coincident ocean surface wind speed

measurements made by the Stepped Frequency Microwave Radiometer (SFMR) on NOAA P-3 hurricane hunter aircraft [24]. These matchups demonstrate a fairly consistent sensitivity of the CYGNSS L1 observables to changes in wind speed at high (30-60 m/s) levels. The mean high wind sensitivities ($d\sigma_o/du_{10}$ and $dLES/du_{10}$) are used to define a YSLF GMF that is consistent with the fully developed seas GMF at low wind speeds but whose high wind behavior is replaced by the YSLF sensitivities derived from the SFMR matchups over hurricanes.

A.2.1   Description of Training Data Set: NOAA P-2 SFMR matchups

Twenty-five (25) coincident overpasses of hurricanes between CYGNSS and NOAA P-3 aircraft occurred during the 2017 Atlantic hurricane season. Coincidence was defined by locating the aircraft ground track during one of its eyewall penetrations that was closest to a CYGNSS specular point track for that overpass and requiring that they occurred within 60 min of one another. The 25 cases identified in this way include overpasses of Hurricanes Harvey, Irma and Maria. The maximum SFMR wind speed recorded across all cases was 73 m/s and the range of CYGNSS incidence angles covered 13-67 deg. Histograms of the SFMR u_{10}, CYGNSS θ_{inc}, and CYGNSS σ_o measured across all 25 overpasses are shown in Fig. A11.
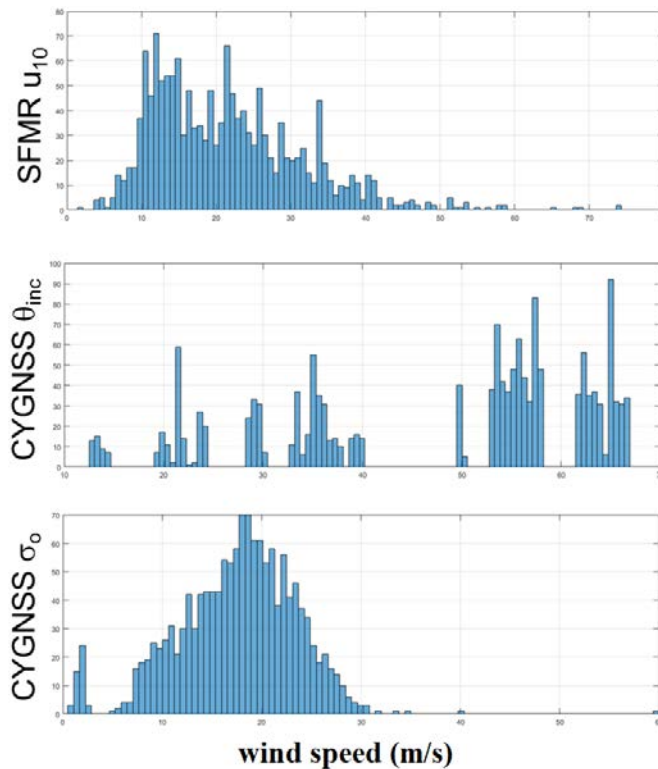

Fig. A11. Histograms of the SFMR u_{10}, CYGNSS θ_{inc}, and CYGNSS σ_o measured across all 25 coincident hurricane overpasses that are used to determine the high wind portion of the limited fetch GMF.

An example of one of the coincident overpasses, occurring over Hurricane Maria on 24 Sep

2017, is shown in Fig. A12. The CYGNSS measurement of σ_o can be seen to decrease roughly monotonically as the wind speed measured by SFMR increases.
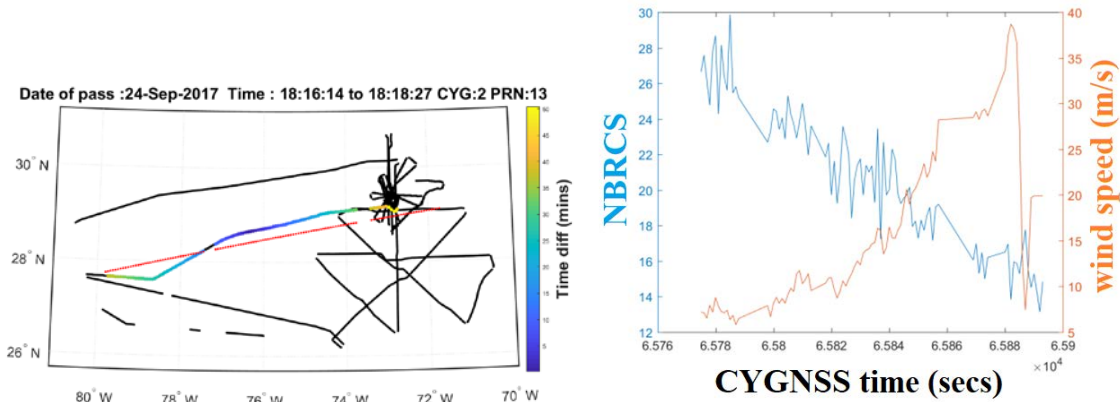

Fig. A12. Coincident CYGNSS/P-3 overpass of Hurricane Maria on 24 Sep 2017 at 1817 UTC. (left) The P-3 ground track is shown in black. The CYGNSS specular point track is shown in red. The colored portion of the P-3 track is color coded by their time difference. (right) CYGNSS L1 σ_o (labeled NBRCS) and SFMR wind speed measured along the coincident track.

A.2.2.   *Regression of coincident overpasses to determine GMF sensitivity ($d\sigma_o$ /du_{10})*

For each of the 25 coincident hurricane overpasses, the L1 observables are related to the coincident SFMR wind speeds by linear regression. One example of this is shown in Fig. A13, for the case illustrated in Fig. A12. The slope of the linear regression is taken as the sensitivity of the observable to changes in wind speed.

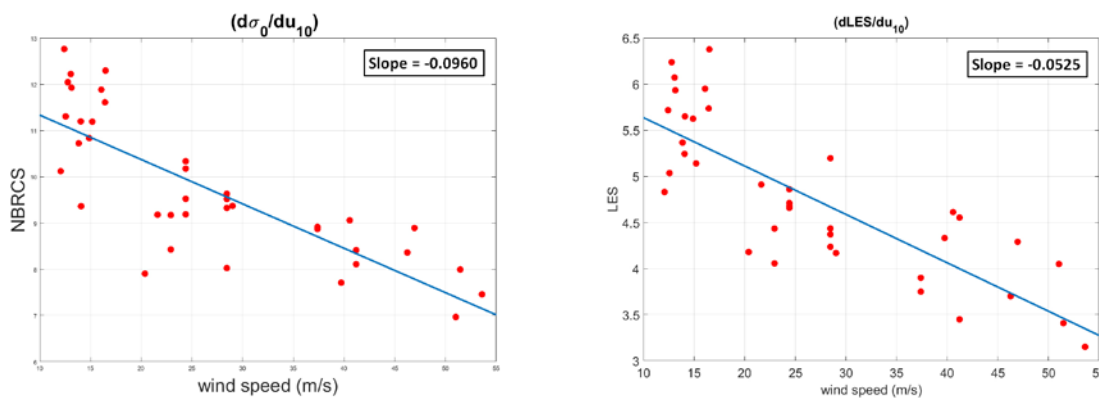*Fig. A13. Linear regression of CYGNSS L1 σ_o and LES observables against SFMR wind speed for Hurricane Maria overpass on 24 Sep 2017 at 1817 UTC. The slope of the linear regression is noted in each plot.*

Fig. A14 shows the slope of the linear regression determined for each of the 25 cases and for both L1 observables. The set of all 25 regression slopes is averaged together to determine the

sensitivity of the YSLF GMF. The resulting sensitivity factors are

$$d\sigma_o/du_{10} = -0.1880 \ (m/s)^{-1}$$
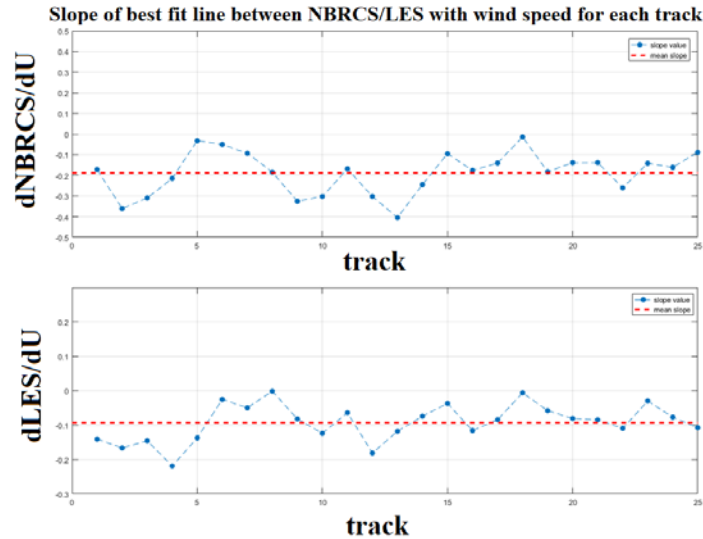$$dLES/du_{10} = -0.0929 \ (m/s)^{-1} \tag{A.3}$$

Fig. A14. Slope of the linear regression determined from each of the 25 coincident hurricane overpasses for both the σ_o (top) and LES (bot) L1 observable.

A.2.3.　　　*Parametric model with hi/lo wind partitions*

As with the fully developed seas GMF, an algebraic parametric model is assumed for the YSLF GMF. The parametric model is again divided into low and high wind speed portions. At low wind speeds, a similar model is assumed as for the FDS GMF

$$Obs = a_0 + a_1 u^{-1} + a_2 u^{-2} \tag{A.4}$$

and the a_{0-2} coefficients are again determined using the fully developed seas training set. At high wind speeds, a linear model is assumed

$$Obs = c_0 + c_1 u \tag{A.5}$$

with slope coefficient (c_1) given by eqn. (A.3) as determined from the linear regression of hurricane overpass matchups. The transition between low and high wind speed segments is again selected as the wind speed where the first derivatives of the two models are equal. The low wind speed model parameters and the spline fit transition point are chosen independently at each incidence angle. A common high wind speed sensitivity is assumed for all incidence angles, since there was no clear dependence of sensitivity on incidence angle evident in the coincident hurricane overpass data.

Examples of the YSLF GMF are shown in Fig. A15, together with the FDS GMFs at the same incidence angles. Several characteristics are noteworthy. The models agree at low wind speeds (by design). At wind speeds in the range 15-25 m/s, the sensitivity ($dObs/du_{10}$) is markedly stronger in the FDS case. The limiting wind speed, above which the value of the observable is zero, is much higher in the YSLF case. And finally, there is a range of values of both observables over which two different wind speeds correspond to the same measurement. This represents an ill-posed, multi-valued inversion problem. In practice, it may be necessary to have some *a priori* knowledge about the fetch or sea age of the conditions under observation in order to uniquely convert L1 observable measurements to wind speed.
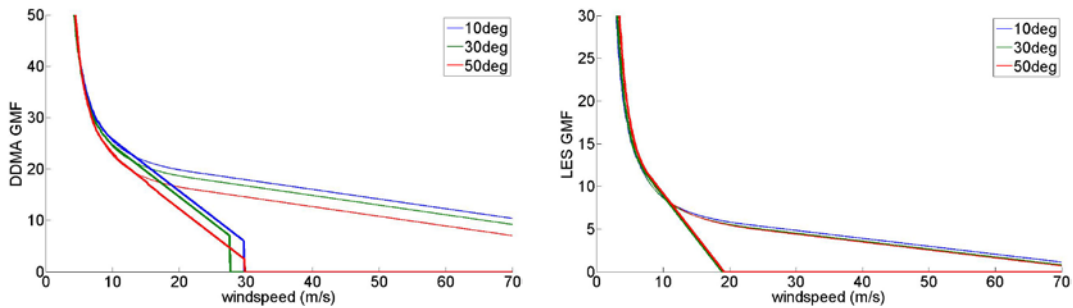

Fig. A15. Young Seas/Limited Fetch GMF (thin lines) for σ_o (left) and LES (right), together with the Fully Developed Seas GMF (thick lines), at θ_{inc} = 10, 30, 50°. The models agree at low wind speeds but have very different high wind speed dependencies.

The multi-valued nature of the mapping from L1 observable to wind speed is illustrated in Fig. A16. The FDS and YSLF GMFs agree below ~12 m/s. Above 12 m/s, the behavior of the FDS GMF is derived from matchups with ECMWF and GDAS away from major storms. The
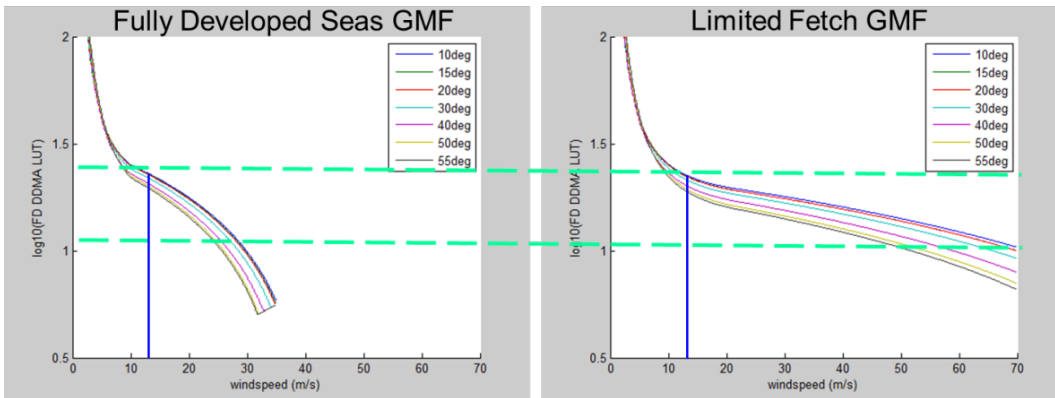
Fig. A16. Side-by-side comparison of GMFs for σ_o that are appropriate for Fully Developed Seas (left) and Young Seas/Limited Fetch conditions in hurricanes (right). The region of the L1 observable bounded by the two dashed green lines maps onto two distinct wind speeds from the same value of σ_o. The wind speed marked by the vertical blue line (12 m/s) represents the highest wind speed with a common mapping by both GMFs.

behavior of the YSLF GMF above 12 m/s is derived from matchups with P-3 SFMR wind speed measurements in major storms. The YSLF GMFs for both L1 observables are shown in Fig. A17.
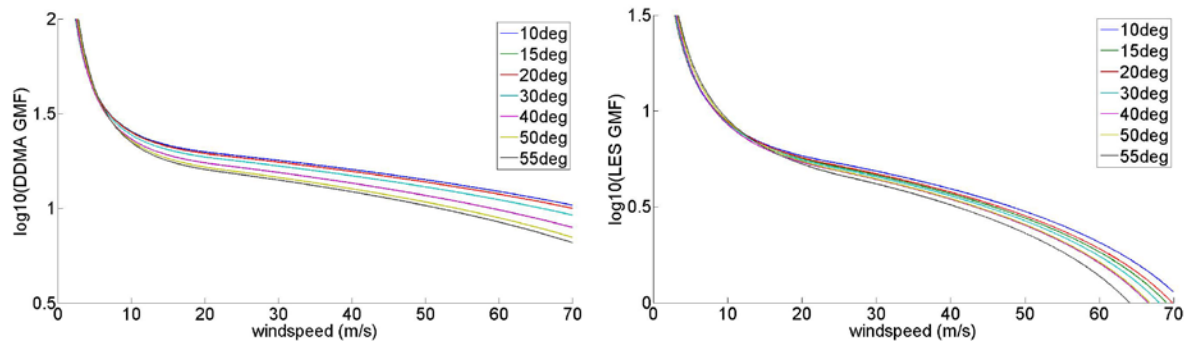

Fig. A17. Representative *YSLF GMFs for the Level 1 observables σ_o (left) and LES (right) at incidence angles of 10, 15, …, 55 deg. The dependence on incidence angle is more pronounced at higher wind speeds with both observables.*

A.2.4. D. *Validation and performance characterization; repeatability of limited fetch conditions*

The adequacy of the YSLF GMFs to represent the response of the Level 1 observables to changes in ocean surface wind speed in hurricanes is assessed by using it in a wind speed retrieval algorithm. The algorithm inverts the GMF to estimate wind speed given the measured observable, in the same manner as the previous assessment for the FDS GMF. In this case, given the limited number of coincident hurricane overpasses that are available, individual case studies are considered rather than overall performance statistics. The wind speed retrieval performance for each hurricane overpass is evaluated using CYGNSS/P-3 matchups similar to the one shown in Fig. A12. The retrieval results for four overpasses are shown in Figs. A18a-d, presented as a time series of the CYGNSS retrieved wind speed overlaid with the nearest SFMR and merged ECMWF&GDAS wind speeds.

In general, the lower wind speed portions of the CYGNSS tracks agree well with the ECMWF&GDAS wind speeds and the high wind speed portions near the storm center agree with the SFMR wind speeds. Note that the CYGNSS retrievals at the highest wind speed levels tend to be noisier, as can be expected given the lower slope of the GMF at high wind speeds. The wind speed retrieved by CYGNSS using the FDS GMF is also included in Fig A18 for comparison. The CYGNSS YSLF and FDS winds agree closely at low wind speed speeds, which is consistent with their very similar GMFs at low winds. At high wind speeds, the YSLF retrievals are higher than the FDS ones, which is also consistent with the higher values for σ_o in the YSLF than the FDS GMF at the same wind speed (see Fig. A16). Occasional drop-outs in the reported FDS winds are evident in Fig. A18 near the storm center. They result from quality control filters which flag the retrievals as unreliable when the individual FDS DDMA and LES retrievals differ by more than 6 m/s.
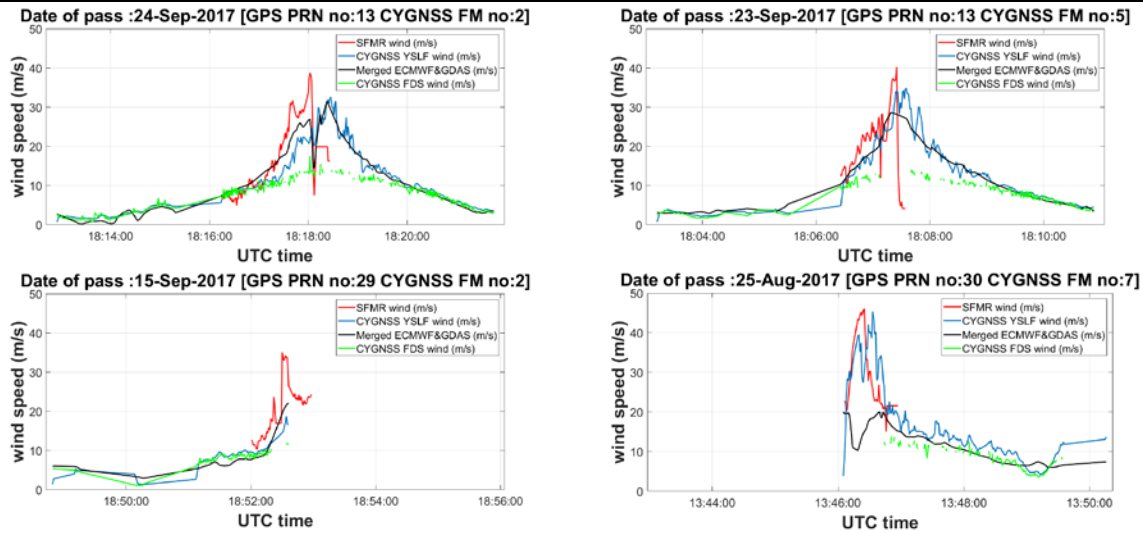
Fig. A18. Coincident hurricane overpasses by CYGNSS and underflights by the NOAA P-3 hurricane hunter aircraft of: (a)Hurricane Maria on 24 Sep 2017, 1817 UTC; (b) Hurricane Maria on 23 Sep 2017, 1807 UTC); (c) Hurricane Maria Coincident Overpass on 15 Sep 2017, 1852 UTC); and (d) Hurricane Harvey on 25 Aug 2017, 1346 UTC. Four wind products are shown: CYGNSS FDS and YSLF retrievals, Stepped Frequency Microwave Radiometer (SFMR) on the P-3 aircraft, and merged numerical weather predictions by ECMWF and GDAS models.

The dependence of CYGNSS measurements on the local wind speed at the location of the measurement is multi-valued in the sense that different wind speeds can result in the same values for σ_o and LES. The relationship appears to be strongly dependent on sea age, with fully developed seas generally exhibiting a repeatable, single-valued mapping. This is demonstrated by the RMS difference between CYGNSS retrieved winds and coincident NWP matchups. As shown in Fig. A10, the RMS difference is between 2 and 3 m/s for NWP wind speeds below 15 m/s, then begins to rise in conditions that are more likely to include partially developed seas. For the young seas with limited fetch that are more typical of conditions in and near tropical cyclones, coincident matchups with airborne SFMR measurements indicate significantly higher values for the σ_o and LES measurements than are observed in the fully developed seas cases, given the same wind speed. This general trend continues for matchups near the inner core of the tropical cyclones, at wind speed values which are above those reported by NWP models.

A detailed assessment and characterization of the performance of a CYGNSS wind speed retrieval algorithm based on the GMF developed here is reported in [30]. To evaluate performance below 20 m/s, a large (~30 million) population of retrieved winds using the FDS GMF is compared to near-coincident winds reported by ECMWF. The RMS difference between them is found to be 2.0 m/s and the component of that difference due to uncertainty in the CYGNSS wind speed retrieval is estimated to be 1.4 m/s. Above 20 m/s, performance is evaluated by comparisons between winds retrieved using the YSLF GMF and near-coincident winds measured by the SFMR instrument on the P-3 hurricane hunter aircraft during eyewall penetrations. In this case, the population of intercomparisons contains 674 samples. The RMS

difference between samples is found to be 6.5 m/s and the component of the difference ascribed to uncertainty in the CYGNSS retrievals is 5.0 m/s. The significantly larger uncertainty in CYGNSS retrievals at high wind speeds is believed to result from two primary causes. One is the decrease in sensitivity of the L1 observables to changes in wind speed as the winds increase. This is illustrated in Fig. A16 by the decrease in slope of both the FDS and YSLF GMFs as wind speed increases. A second cause for the increased retrieval uncertainty at high winds is the sensitivity of the L1 observables to sea state conditions not directly related to wind speed – in particular, to sea age or fetch length in and near tropical cyclones.

The multi-valued dependence on wind speed can be explained by considering the general relationship between GNSS-R measurements and sea state, and the relationship between the sea state at a particular location and the local wind speed there. In general, local winds tend to generate surface roughness nearly instantaneously at the smaller, capillary, wavelength end of the surface height spectrum. The influence of winds on the longer wavelength, swell, portion of the spectrum takes longer to develop, both in time and in fetch length. For this reason, young seas in limited fetch conditions will tend to have a smaller long wave portion of their spectrum. In fully developed seas, the wind speed has sufficient time to influence the full roughness spectrum and the relationship between capillary and swell waves is more consistent. This general behavior is consistent with the fact that the derived empirical GMFs map values for σ_o and LES to higher wind speeds in young sea/limited fetch conditions than in fully developed seas. The existence of multiple GMFs presents implementation challenges for a wind speed retrieval algorithm in terms of deciding which one to use and under what conditions. Use of the FDS version appears to perform well in most cases away from major storms. Likewise, the YSLF version performs well with most coincident hurricane overpass cases, but not all. The use of a single non-FDS GMF should be considered an approximation to the true dependence of the GMF on sea age or fetch length. It reflects an effective average of the relationship between L1 observables and sea state across the young seas/limited fetch conditions that were present in the 2017 Atlantic hurricanes from which the GMF was derived. A more accurate accounting for the departure from a fully developed sea state might, for example, use a fetch-dependent parametrization of the YSLF GMF, or it might modify the L1 observables based on sea age or fetch length in order to estimate an effective FDS values. These are possible improvements that are under consideration for future development of the GMF.

Appendix B: Implementation of On-orbit Wind Speed Rretrieval Algorithm

B.1   Choice of Delay/Doppler Range for Observable Calculation

A very important aspect for the computation of DDMA and LES is the delay and Doppler range, over which the observables are computed. The choice of the delay and Doppler ranges is a trade-off between the improvement in SNR that results from averaging across more of the diffuse scattered signal in the glistening zone, versus the improvement in spatial resolution that results from only averaging over a limited region. The delay and Doppler range depends on the requirement on spatial resolution that one wishes to comply with. For CYGNSS, the baseline requirement on the spatial resolution of the retrieved winds is 25 km x 25 km, ad therefore this is the spatial resolution we focus on. Figure A6 illustrates two examples of how, for a spatial resolution requirement of 25x25 km, the corresponding delay and Doppler range varies. Note that the iso-range ellipses become closer to each other as the delay increases; furthermore, they widen and stretch out with increasing incidence angles without changing their orientation, so the geometrical parameter that mostly influences the configuration of the iso-delay lines is the incidence angle. In the case of iso-Doppler lines, the spacing between them also increases with increasing incidence angle, but they are also strongly affected by the velocity vectors of the transmitter and above all of the receiver, which change their orientation.
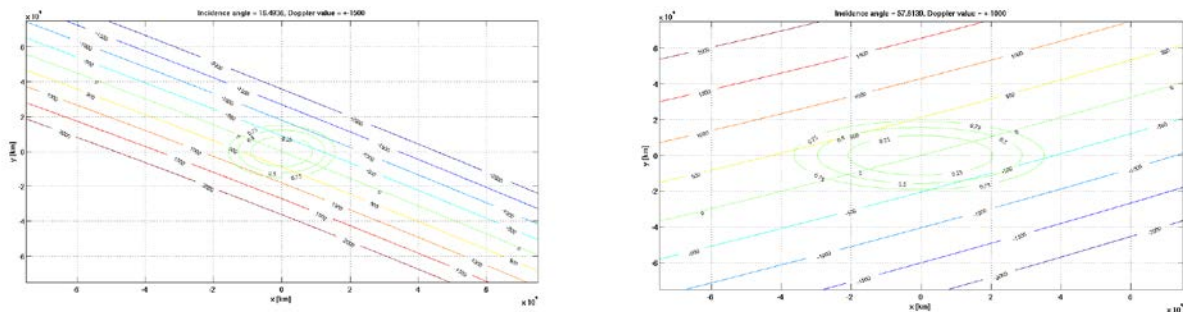

Figure B6. Iso-delay and Iso-Doppler lines for an incidence angle of 16.5° (left) and of 57.6° (right).

Figure B7 shows curves of the square root of the Instantaneous Field of View (IFOV) versus the incidence angle, for different delay ranges. All the delay ranges considered begin one delay sample before the 0 chip sample to improve the SNR, since the samples adjacent to the specular point one still contain a good amount of the scattered power from the specular point pixel, due to the power spreading caused by the Woodward Ambiguity Function (WAF). The IFOV is defined here as the physical area included in the iso-range ellipse corresponding to a given delay. It is clear from Figure A7 that for a 25 km x 25 km requirement on the spatial resolution (shown as black continuous line) the only suitable choice is a delay range from -0.25 to 0.25.
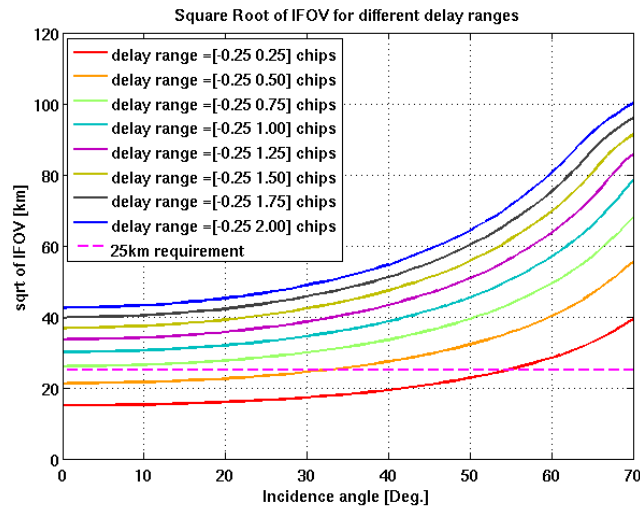
Figure B7. illustration of curves of square root of IFOV versus incidence angle, for a variety of delay ranges. The dashed magenta line shows the 25 km requirement.

Having selected the delay range, the Doppler range has to be chosen such that the iso-Doppler lines will not truncate some of the scattered signal within the iso-delay ellipse at 0.25 chips, but also will not lie too far from the iso-delay ellipse at 0.25 chips and introduce additional noise without adding more signal. Hence, the Doppler range is then chosen to try to satisfy the following two conditions:

- Iso-Doppler lines are the closest possible to the iso-delay line selected
- Iso-Doppler lines are always outside the iso-delay line selected

Since the iso-delay ellipses change depending on geometry, and iso-Doppler lines change their distance to one another and their orientation for different geometries (as shown in Figure B6), the Doppler range that satisfies the above conditions for a given iso-delay line is not unique, and depends on the particular geometry. Furthermore, the Doppler range has an overall much lower influence on the final IFOV, as it is illustrated in Figure B8, where curves of square root of IFOV versus incidence angle are shown for a single delay range, and a number of different Doppler ranges. It is interesting to observe that the first case of Doppler range of [-250 250] Hz is different from the others, and noisier, since the very small Doppler range chops off part of the area within the iso-delay ellipse at 0.25 chips. Instead, small differences can be observed for the other cases, and mostly at lower incidence angles. This happens because at higher incidence angles, the iso-Doppler lines stretch out more rapidly than the iso-delay lines and tend to fall quickly outside the 0.25 iso-delay ellipse, thus the IFOV for higher incidence angles is entirely determined by the delay range. Furthermore, there is no difference in the IFOV between the [-1000 1000] Hz Doppler range, and the [-1500 1500] Doppler range, suggesting that these iso-Doppler lines fall outside the 0.25 chip iso-Delay lines for all geometries.
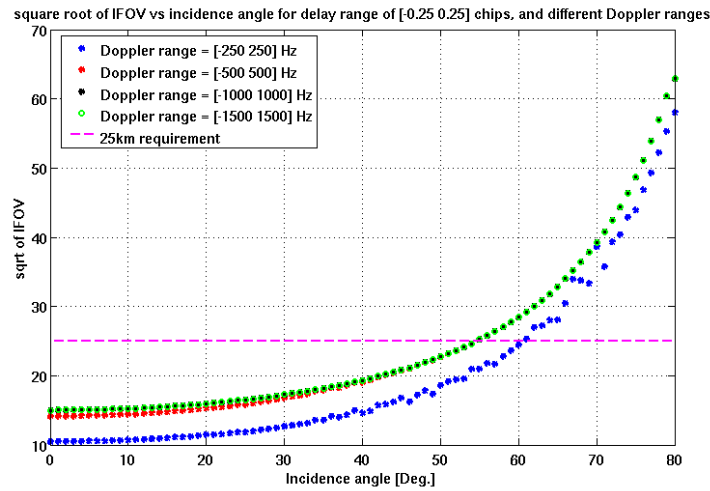
Figure B8. illustration of curves of square root of IFOV versus incidence angle, for a fixed delay range of [-0.25 0.25] chips, and different Doppler ranges. The dashed magenta line shows the 25 km requirement.

Thus, the final choice has been for a Doppler range of [-1000 1000] Hz. This choice has then been confirmed by applying the full L2 retrieval algorithm to DDMA observables computed using the three different Doppler ranges, and by verifying that the Doppler range of [-1000 1000] is the one providing the lowest RMS error among the three

Note that in this case only the physical area included in the iso-delay and iso-Doppler lines is considered. Several additional factors need to be taken into account in order to properly relate the delay and Doppler ranges to the spatial region These include: (1) The spatial boundaries defined by lines of constant iso–delay and iso–Doppler do not conform to a line of constant iso-distance from the specular point. An effective spatial resolution, based, for example, on equal area coverage, needs to be defined; (2) The Woodward ambiguity function and Doppler filter impulse response define weighted response functions for the contribution of different regions of the delay/Doppler domain to the measurements. These weightings should also be accounted for by a suitable definition of the effective spatial resolution; and (3) Multiple samples of the DDM can be averaged together in ground-processing to reduce measurement noise. This will produce spatial smearing in the direction of motion of the specular point. Each of these factors will be included in a more complete definition of the spatial resolution, which is currently under development.

B.2   *Derivation of Coefficients for Minimum Variance Estimator*

The coefficients to combine wind estimates from different observables to form a minimum variance estimator are given by [Clarizia et al., 2014]

$$m = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i,j}^{-1} \right)^{-1} C^{-1} \mathbf{1} \qquad \text{(B.6)}$$

where $\mathbf{1}$ is a vector of ones, C^{-1} is the inverse of the covariance matrix between the individual retrieval errors, and $c_{i,j}^{-1}$ are its elements. The variance of the MV estimator is given by

$$\sigma_{MV}^2 = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i,j}^{-1} \right)^{-1} \qquad \text{(B.7)}$$

The MV estimator requires knowledge of the covariance matrix of the individual retrieval errors. The covariance is estimated empirically from the retrieval errors, and can be factored into two component matrices, as

$$C = SRS \qquad \text{(B.8)}$$

where S is a diagonal matrix of standard deviations of the retrieval errors for each observable (i.e. the square root of the diagonal elements of the covariance matrix), and R is the matrix of correlation coefficients, whose elements are always between -1 and 1.

In practice, the correlation between DDMA and LES retrieval errors is found to vary as a function of the SNR of the measurements. For this reason, the covariance matrix assumed by the estimator is varied accordingly, in a manner referred to as Adaptive Covariance (AC). This approach consists of estimating a different covariance matrix, and therefore a different pair of coefficients for the linear combination, for different ranges of the SNR. The can SNR is estimated using the Range Corrected Gain (RCG), and the estimator with the AC approach included becomes:

$$u_{MV}^i = m^i \cdot u^i \qquad RCG_{low} < RCG < RCG_{high} \qquad \text{(B.9)}$$

where i refers to each of the RCG intervals used to compute the coefficients in the linear combination, m is the vector of coefficients defined in (1) and u is the vector of wind speed estimates from DDMA and LES.

The DDMA and LES wind estimates are linearly combined in [Clarizia and Ruf, 2016] into a MV estimator (called best-weighted estimator in the paper), using coefficients calculated according to equation (B.6), for the following intervals of RCG:

1) $1 \leq RCG < 3$
2) $3 \leq RCG < 5$
3) $5 \leq RCG < 10$
4) $10 \leq RCG < 20$
5) $20 \leq RCG < 30$
6) $30 \leq RCG < 70$
7) $70 \leq RCG < 110$
8) $110 \leq RCG < 150$
9) $RCG \geq 150$

The covariance matrices used to compute the best weighted estimate coefficients are derived empirically from the population of retrievals, since neither the individual PDFs of DDMA and LES wind retrieval errors nor the joint PDF between the two retrieval errors are available in analytical form. The bias between true and retrieved winds is first computed for each observable, and for each RCG interval, and removed, so as to compute the coefficients from unbiased DDMA and LES wind estimates. The dependence of bias on wind speed was found not to be significant and therefore was not included. The coefficients *m* as well as the bias are calculated from the training dataset (50% of data), where the true wind speeds are known, and then used in the test dataset, where the wind speeds are not known, to assess algorithm performance. This means that the coefficients may be suboptimal when applied to the test dataset, but we expect the performance to be asymptotically optimal for increasing number of observations, and assuming that the statistical properties of the training dataset are the same as those of the test dataset.

B.2.1   Minimum Variance Coefficients LUT

The coefficients used to combine the DDMA and LES wind estimates to obtain the final estimated wind are reported in Table B3 for the different RCG intervals considered. These coefficients are currently being used by the algorithm implemented in the SOC, and to produce version 1.0 of the wind product. They have been derived from simulations using Nature Run data.

MV coefficients [k1 k2]	RCG interval
[0.4370 0.5630]	$1 \leq RCG < 3$
[0.4025 0.5975]	$3 \leq RCG < 5$
[0.4358 0.5642]	$5 \leq RCG < 10$
[0.5540 0.4460]	$10 \leq RCG < 20$
[0.6251 0.3749]	$20 \leq RCG < 30$
[0.6474 0.3526]	$30 \leq RCG < 70$
[0.6805 0.3195]	$70 \leq RCG < 110$
[0.7151 0.2849]	$110 \leq RCG < 150$
[0.7450 0.2550]	$150 \leq RCG$

Table B3. MV coefficients for the different RCG intervals

B.3   *Debiasing LUT*

The LUT of debiasing coefficients used by the FDS MV wind speed retrieval algorithm is shown in Fig. B9. Coefficients which lie along the 45º "one-to-one" line represent wind speed ranges where the probability distributions already match well and no adjustment is needed. Coefficients above and below the 1:1 line represent remappings where a range of retrieved wind speeds is compressed or expanded to more closely match the likelihood of occurrence of the ECMWF wind speed for the same wind speed interval.
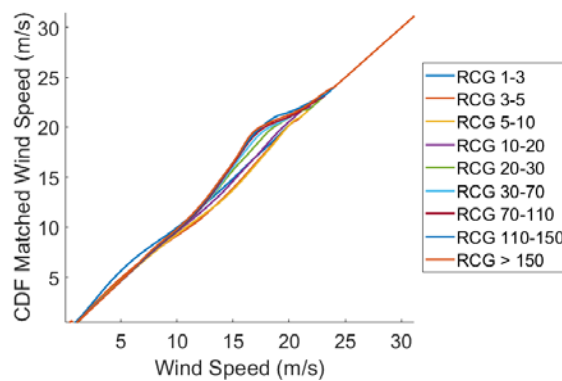

Figure B9. Debiasing coefficients used to match the global probability distribution of CYGNSS L2 fully developed seas minimum variance wind speeds to those of the ECMWF reanalysis numerical weather prediction product. Coefficients are determined independently for different intervals of the range corrected gain (RCG).

B.4   Time-averaging LUT

The time-averaging LUT contains the number of samples to time-average as a function of incidence angle. The incidence angle axis is defined as going from 0° to 89°, with a step of 1 °. Values are reported in Table B1 for specific incidence angle intervals.

Number of samples to time-average	Incidence angle interval [°]
5	$0° < \theta \leq 17°$
4	$17° < \theta \leq 31°$
3	$31° < \theta \leq 41°$
2	$41° < \theta \leq 48°$
1	$\theta > 48°$

Table B1. Number of samples to time-average as a function of incidence angle

B.5   Standard Deviation of the retrieval error LUT

The standard deviation of the error in the retrieved wind, associated with each RCG interval, wind speed interval, and GPS satellite block type is reported in Tables B4. The values are empirically derived from statistical comparisons between retrieved winds and matchup ground truth winds provided by ECMWF reanalaysis NWP products (for the FDS MV retrievals) and by NOAA P-3 hurricane hunter SFMR measurements (for the YSLF retrievals at the higher wind speed intervals).

Table B4.1. Standard Deviation of FDS MV wind speed retrieval error			
σ_{FDS_MV}	$0 \leq u_{10} < 10$ m/s	$10 \leq u_{10} < 20$ m/s	$20 \leq u_{10} < 40$ m/s**
$1 \leq RCG < 3$	2.5	2.5*	3.5*
$3 \leq RCG < 10$	2.0	2.0*	3.0*
$10 \leq RCG < 110$	1.5	2.0*	3.0*
$110 \leq RCG$	1.5	1.75*	2.5*
* Increase by 50% if GPS Block Type II-F transmitter			
** Uncertainty $\equiv$ 99 if $u_{10} > 40$ m/s for FDS wind speeds			

Table B4.2. Standard Deviation of YSLF wind speed retrieval error			
$\sigma_{YSLF\_DDMA}/\sigma_{YSLF\_LES}$	$1 \leq RCG < 10$	$10 \leq RCG < 30$	$30 \leq RCG$
$0 \leq u_{10} < 10$ m/s	2.5/2.5	2.5/2.5	2.5/2.5
$10 \leq u_{10} < 20$ m/s	3.0/3.0*	3.0/3.0*	3.0/3.0*
$20 \leq u_{10} < 40$ m/s	4.5/4.5*	3.5/4.0*	3.5/4.0*
$40 \leq u_{10}$ m/s	6.0/6.0*	5.0/6.0*	4.0/5.0*
* Increase by 50% if GPS Block Type II-F transmitter			

B.6 L2 Data Dictionary

Table B5 lists all the variables in the Level 2 wind speed NetCDF files.

Name	Long Name	netCDF Type	CF Conventions Units	netCDF Dimensions	Comment
Global Values					
time_coverage _start	<none>	file attribute, string	<none>	<none>	sample_time of the first sample in the file in ISO-8601 form
time_coverage _end	<none>	file attribute, string	<none>	<none>	sample_time of the last sample in the file in ISO-8601 form
time_coverage _duration	<none>	file attribute, string	<none>	<none>	The time interval between test_coverage_start and test_coverage_end in ISO1806 form
time_coverage _resolution	<none>	file attribute, string	<none>	<none>	The nominal time interval between samples in ISO1806 form
ddm_source	Level 0 data source	byte	<none>	<none>	The source of the Level 0 DDM raw counts and metadata utilized to derive wind_speed. 0 = E2ES (CYGNSS end-to-end simulator) 1 = GPS signal simulator 2 = CYGNSS spacecraft 3 = Source Unknown
nbrcs_les_sel_ lookup_tables _version	<none>	file attribute, string	<none>	<none>	The GMF NBRCS and LES selection lookup table version number.
time_averagin g_lookup_tabl es_version	<none>	file attribute, string	<none>	<none>	The GMF time-averaging lookup table version number.
nbrcs_wind_l ookup_tables_ version	<none>	file attribute, string	<none>	<none>	The GMF NBRCS to wind speed lookup table version number.

les_wind_look up_tables_ver sion	<none>	file attribute, string	<none>	<none>	The GMF LES to wind speed lookup table version number.
covariance_lo okup_tables_v ersion	<none>	file attribute, string	<none>	<none>	The GMF minimum covariance lookup table version number.
standard_devi ation_lookup_ table_version	<none>	file attribute, string	<none>	<none>	The GMF standard deviation lookup table version number.
l2_algorithm_ version	<none>	file attribute, string	<none>	<none>	L2 processing algorithm version number.
source	<none>	file attribute, string	<none>	<none>	Level 1 netCDF source file names.
Per-Sample Values					
spacecraft_id	CCSDS spacecraft identifier	short	1	sample	The CCSDS spacecraft identifier: 0xF7 (247): CYGNSS 1 0xF9 (249): CYGNSS 2 0x2B (43): CYGNSS 3 0x2C (44): CYGNSS 4 0x2F (47): CYGNSS 5 0x36 (54): CYGNSS 6 0x37 (55): CYGNSS 7 0x49 (73): CYGNSS 8 0x00 (0): E2ES 0x0E (14): engineering model 0x0D (15): default 0xFF (255): unknown
spacecraft_nu m	CYGNSS spacecraft number	byte	1	sample	The CYGNSS spacecraft number: Ranges from 1 through 8 and 99. 1 through 8 are on-orbit spacecraft. 99 is the CYGNSS end-to-end simulator.
prn_code	GPS PRN code	byte	1	sample	The PRN code of the GPS signal associated with the DDMs utilized to derive wind_speed. Ranges from 0 to 32. 0 = reflectometry channel idle. 1 to 32 represents PRN code.
sv_num	GPS space vehicle number	short	1	sample	The GPS unique space vehicle number that transmitted prn_code.

antenna	Receive antenna	byte	<none>	sample	The CYGNSS nadir antenna that received the reflected GPS signal associated with the DDMs utilized to derive wind_speed. 0 = none 1 = zenith (never used) 2 = nadir_starboard 3 = nadir_port
sample_time	Sample time	double	seconds since time_coverage_start	sample	The mean of ddm_timestamp_utc of the DDMs that were utilized to derive wind_speed. Note that the DDM sampling period is not synchronized with the UTC change of second so sample_time can occur at any time relative to the UTC change of second.
lat	Latitude	float	degrees_north	sample	The mean of the specular point latitudes of the DDMs that were utilized to derive wind_speed, degrees North.
lon	Longitude	float	degrees_east	sample	The mean of the specular point longitudes of the DDMs that were utilized to derive wind_speed, degrees East.
sc_lat	Subsatellite point latitude	float	degrees_north	sample	The mean of the subsatellite point latitudes of the DDMs that were utilized to derive wind_speed, degrees North.
sc_lon	Subsatellite point longitude	float	degrees_east	sample	The mean of the subsatellite point longitudes of the DDMs that were utilized to derive wind_speed, degrees East.
sc_alt	Spacecraft altitude	int	meter	sample	The mean of the satellite altitudes above the WGS-84 ellipsoid of the DDMs that were utilized to derive wind_speed, meters.
wind_speed	Retrieved wind speed using Minimum Variance estimator applied to Fully Developed Seas retrievals from NBRCS and LES	float	m s-1	sample	The average surface wind speed of the 25 x 25 km cell centered on lat and lon derived from both the NBRCS and the LES observables using the fully developed seas geophysical model function, m/s. Multiple DDMs are utilized to derive wind_speed. The number of utilized DDMs ranges from 1 to 5.
fds_nbrcs_wind_speed	Fully Developed Sea retrieval of wind speed from NBRCS	float	m s-1	sample	The average surface wind speed of the 25 x 25 km cell centered on lat and lon derived only from the NBRCS observable using the fully developed seas geophysical model function, m/s.

					Multiple DDMs are utilized to derive fds_nbrcs_wind_speed. The number of DDMs ranges from 1 to 5.
fds_les_wind_speed	Fully Developed Seas retrieval of wind speed from LES	float	m s-1	sample	The average surface wind speed of the 25 x 25 km cell centered on lat and lon derived only from the LES observable using the fully developed seas geophysical model function, m/s. Multiple DDMs are utilized to derive fds_les_wind_speed. The number of DDMs ranges from 1 to 5.
wind_speed_uncertainty	Uncertainty in MV FDS wind speed	float	m s-1	sample	Standard deviation of the additive wind speed error in the minimum variance fully developed seas wind speed retrieval (dependent on the RCG of the specular point location, the block type of the GPS satellite, and the wind speed), in m/s.
yslf_nbrcs_wind_speed	Young Seas/Limited Fetch retrieval of wind speed from NBRCS	float	m s-1	sample	The surface wind speed centered on lat and lon derived only from the NBRCS observable using the young seas/limited fetch geophysical model function, m/s. Multiple DDMs are utilized to derive wind_speed. The number of utilized DDMs ranges from 1 to 5.
yslf_les_wind_speed	Young Seas/Limited Fetch retrieval of wind speed from LES	float	m s-1	sample	The surface wind speed centered on lat and lon derived only from the LES observable using the young seas/limited fetch geophysical model function, m/s. Multiple DDMs are utilized to derive wind_speed. The number of utilized DDMs ranges from 1 to 5.
yslf_nbrcs_wind_speed_uncertainty	Uncertainty in NBRCS YSLF wind speed	float	m s-1	sample	Standard deviation of the additive wind speed error in yslf_nbrcs_wind_speed (dependent on the RCG of the specular point location, the block type of the GPS satellite, and the wind speed), in m/s.
yslf_les_wind_speed_uncertainty	Uncertainty in LES YSLF wind speed	float	m s-1	sample	Standard deviation of the additive wind speed error in yslf_les_wind_speed (dependent on the RCG of the specular point location, the block type of the GPS satellite, and the wind speed), in m/s.
mean_square_slope	Mean square slope	float	1	sample	The average MSS of the 25 x 25 km cell centered on lat and lon, unitless.
mean_square_slope_uncertai	Mean square slope	float	1	sample	The uncertainty of mean_square_slope, unitless.

nty	uncertainty				
incidence_angle	Incidence angle	float	degree	sample	The mean of the incidence angles of the specular points of the DDMs that were utilized to derive wind_speed, degrees.
azimuth_angle	Azimuth angle	float	degree	samle	The mean of the orbit frame azimuth angles of the specular points of the DDMs that were utilized to derive wind_speed, degrees.
nbrcs_mean	NBRCS mean	float	1	sample	The mean of the DDM NBRCS values that were utilized to derive wind_speed, unitless.
les_mean	LES mean	float	1	sample	The mean of the DDM LES values that were utilized to derive wind_speed, unitless.
range_corr_gain	Range corrected gain	float	1e-27 dBi meter-4	sample	The mean of the RCGs of the DDMs that were utilized to produce wind_speed. Individual RCGs are equal to the receive antenna gain in the direction of the specular point multiplied by 1e27 divided by the square of the receiver to the specular point range and the square of the transmitter to specular point range. Units: 1e27 *dBi * m^-4
fresnel_coeff	Fresnel power reflection coefficient	float	1	sample	The square of the left hand circularly polarized Fresnel electromagnetic voltage reflection coefficient at 1575 MHz for a smooth ocean surface at lat, lon. See UM document 148-0361 for a description of the calculation of the Fresnel coefficient, unitless.
num_ddms_utilized	Number of DDM utilized	byte	1	sample	The number of DDMs averaged together to produce wind_speed. The number of DDMs utilized depends on the incidence angle. Ranges from 1 to 5.
sample_flags	General status flags for the sample	short	<none>	sample	Set of flags indicating general conditions for the sample, set to 1 if condition is true. Flag bit masks: 1 = low_quality_gps_ant_knowledge. The directional gain pattern of the GPS transmit antenna, and hence the value of its gain in the direction of the specular point, is less well known for block type II-F GPS satellites. This flag indicates that Level 1 calibration was based on a GPS antenna gain value with a higher than normal uncertainty. The impact on retrieved wind speed values is typically less than 1 m/s at

					low to moderate wind speeds, so this is considered a non-fatal flag. The increase in uncertainty at higher wind speeds is reflected in the uncertainty data fields.
fds_sample_flags	Sample status flags for Fully Developed Seas wind speed retreval	short	<none>	sample	Set of FDS status flags, set to 1 if condition is true. Flag bit masks: 1 = fatal_composite_wind_speed_flag, logical OR of fatal FDS flags (fatal_neg_wind_speed OR fatal_high_wind_speed OR fatal_retrieval_ambiguity OR fatal_low_range_corr_gain) 2 = non_fatal_neg_wind_speed_flag, -5 < wind_speed < 0 m/s 4 = non_fatal_neg_fds_nbrcs_wind_speed, -5 < fds_nbrcs_wind_speed < 0 m/s 8 = non_fatal_neg_fds_les_wind_speed, -5 < fds_les_wind_speed < 0 m/s 16 = fatal_neg_wind_speed, wind_speed <= -5 m/s 32 = fatal_neg_fds_nbrcs_wind_speed, fds_nbrcs_wind_speed <= -5 m/s
					64 = fatal_neg_fds_les_wind_speed, fds_les_wind_speed <= -5 m/s 128 = fatal_high_wind_speed, fatal_high_fds_nbrcs_wind_speed and fatal_high_fds_les_wind_speed are both one 256 = fatal_high_fds_nbrcs_wind_speed, nbrcs_mean corresponds to a wind speed > maximum FDS NBRCS lookup table wind speed at incidence_angle 512 = fatal_high_fds_les_wind_speed, les_mean corresponds to a wind speed > maximum FDS LES lookup table wind speed at incidence_angle

					1024 = non_fatal_ascending, satellite is on the ascending node of the orbit (subsatellite point latitude is increasing) 2048 = fatal_retrieval_ambiguity, wind_speed was derived from both fds_nbrcs_wind_speed and fds_les_wind_speed AND the absolute value of the difference between fds_nbrcs_wind_speed and fds_les_wind_speed is > 10.0 m/s 4096 = non_fatal_single_observable, wind_speed was derived from a single observable, either fds_nbrcs_wind_speed or fds_les_wind_speed, but not both 8192 = fatal_low_range_corr_gain, range_corr_gain < 1 16384 = non_fatal_low_quality_gps_ant_knowledge. The directional gain pattern of the GPS transmit antenna, and hence the value of its gain in the direction of the specular point, is less well known for block type II-F GPS satellites. This flag indicates that Level 1 calibration was based on a GPS antenna gain value with a higher than normal uncertainty. The impact on retrieved wind speed values is typically less than 1 m/s at low to moderate wind speeds, so this is considered a non-fatal flag. The increase in uncertainty at higher wind speeds is reflected in the 'wind_speed_uncertainty' data field.
yslf_sample_flags	Sample status flags for Young Seas/Limited Fetch wind speed retreval	short	<none>	sample	Set of YSLF status flags, set to 1 if condition is true. Flag bit masks: 1 = fatal_composite_yslf_wind_speed, logical OR of fatal YSLF flags (fatal_neg_yslf_wind_speed OR fatal_high_yslf_wind_speed OR fatal_yslf_retrieval_ambiguity OR fatal_low_range_corr_gain) 2 = spare_3, always zero 4 = spare_4, always zero 8 = fatal_neg_yslf_wind_speed, non_fatal_neg_yslf_nbrcs_wind_speed and non-fatal_neg_yslf_les_wind_speed are both one 16 = non_fatal_neg_yslf_nbrcs_wind_speed, yslf_nbrcs_wind_speed <= -5 m/s 32 = non_fatal_neg_yslf_les_wind_speed, yslf_les_wind_speed <= -5 m/s

					64 = fatal_yslf_retrieval_ambiguity, the absolute value of the difference between yslf_nbrcs_wind_speed and 1.43*yslf_les_wind_speed is > 12.0 m/s 128 = fatal_high_yslf_wind_speed, either fatal_high_yslf_nbrcs_wind_speed or fatal_high_yslf_les_wind_speed is one (or both are one) 256 = fatal_high_yslf_nbrcs_wind_speed, nbrcs_mean corresponds to a YSLF wind speed >= 99.9 m/s 512 = fatal_high_yslf_les_wind_speed, les_mean corresponds to a YSLF wind speed >= 99.9 m/s
					1024 = non_fatal_ascending, satellite is on the ascending node of the orbit (subsatellite point latitude is increasing) 2048 = spare_6, always zero 4096 = spare_7, always zero 8192 = fatal_low_yslf_range_corr_gain, range corrected gain of the DDM used for YSLF winds is < 1 16384 = non_fatal_low_quality_gps_ant_knowledge. The directional gain pattern of the GPS transmit antenna, and hence the value of its gain in the direction of the specular point, is less well known for block II-F GPS satellites. This flag indicates that Level 1 calibration was based on a GPS antenna gain value with a higher than normal uncertainty. The impact on retrieved wind speed values is typically less than 1 m/s at low to moderate wind speeds, so this is considered a non-fatal flag. The increase in uncertainty at higher wind speeds is reflected in the 'yslf_nbrcs_wind_speed_uncertainty' and 'yslf_les_wind_speed_uncertainty' data fields.
sum_neg_brcs _values_used_ for_nbrcs_flag s	Sum of negative BRCS values used for NBRCS flags	byte	1	sample	The number of DDMs utilized to produce wind_speed that used at least one negative BRCS value to calculate NBRCS. Ranges from 1 to 5.
Per-DDM Values					
Wind retrievals are produced utilizing from one to five DDMs. The values below are five					

element arrays which contain per-DDM values. The DDMs that were utilized for wind retrieval are indicated by the ddm_obs_utilized_flag array. Unutilized DDMs are assigned fill values.					
ddm_obs_utilized_flag	DDM-utilized flags	byte	1	sample, ddm	A five element array, one element per DDM. Each element is a flag set to one if the corresponding DDM was utilized to produce wind_speed.
ddm_sample_index	Level 1 netCDF sample indices	int	1	sample, ddm	A five element array, one per DDM. Contains the Level 1 netCDF sample index of the corresponding DDM. Can be utilized together with ddm_channel, spacecraft_num and 'source' to look up the corresponding Level 1 DDM data and metadata.
ddm_channel	Level 1 DDM reflectometry channels	byte	1	sample, ddm	A five element array, one per DDM. Contains the Level 1 netCDF reflectometry channel of the corresponding DDM. Can be utilized together with the ddm_sample_index, spacecraft_num and 'source' to look up the corresponding Level 1 DDM data and metadata.
ddm_les	DDM leading edge slope observables	float	1	sample, ddm	A five element array, one element per DDM. Contains the corresponding DDM leading edge slope value, unitless.
ddm_nbrcs	DDM normalized bistatic radar cross secton observables	float	1	sample, ddm	A five element array, one element per DDM. Contains the corresponding DDM normalized bistatic radar cross section value, unitless.

Table B5. List of data fields contained in the Level 2 wind retrieval data files