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 2 Chg 2**
	Date	**19 November 2015**
	Contract	**NNL13AQ00C**

SPACE PHYSICS RESEARCH LABORATORY

NASA

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 2 Chg 2**
	Date	**19 November 2015**
	Contract	**NNL13AQ00C**

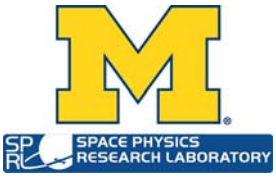
Prepared by: Maria Paola Clarizia, University of Michigan Date: 11/19/2015
Valery Zavorotny, NOAA

Approved by: *Email approval on file* Date: 11/14/2015
Chris Ruf, CYGNSS Principal Investigator

Approved by: *Email approval on file* Date: 11/13/2015
John Scherrer, CYGNSS Project Manager

Approved by: *Email approval on file* Date: 11/18/2015
Randy Rose, CYGNSS Project Systems Engineer

Released by: *Email approval on file* Date: 11/20/2015
Damen Provost, 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;

Table of Contents

1. Introduction and Summary

The CYGNSS Project will implement 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 will 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 largely 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) that 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.

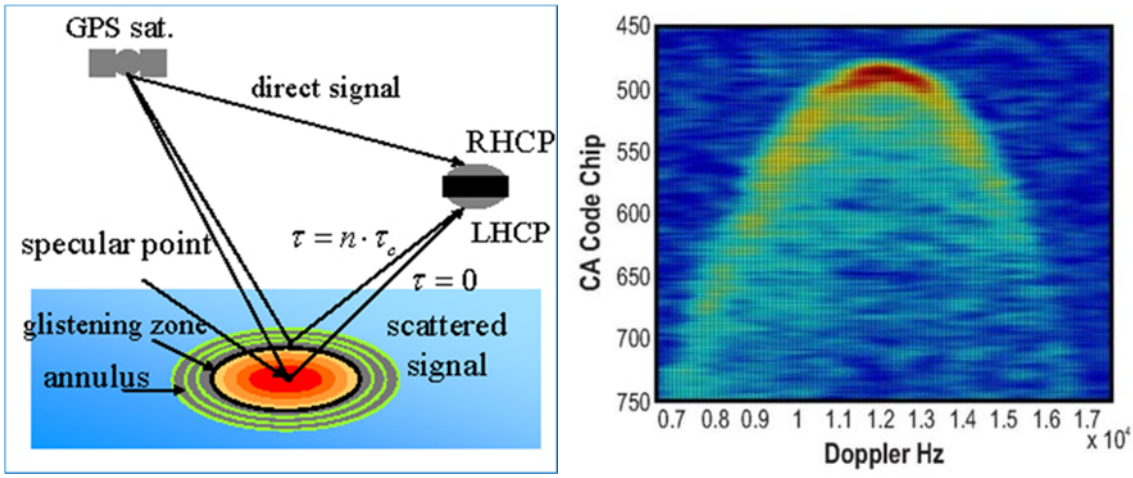


Figure 1. [left] GPS signal propagation and scattering geometries for *ocean surface bistatic quasispecular scatterometry*. The position of the spacecraft is determined from the direct GPS signal; the surface winds are determined by the indirect signal scattered off the ocean surface. Combining the position 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 Figure 1. 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.

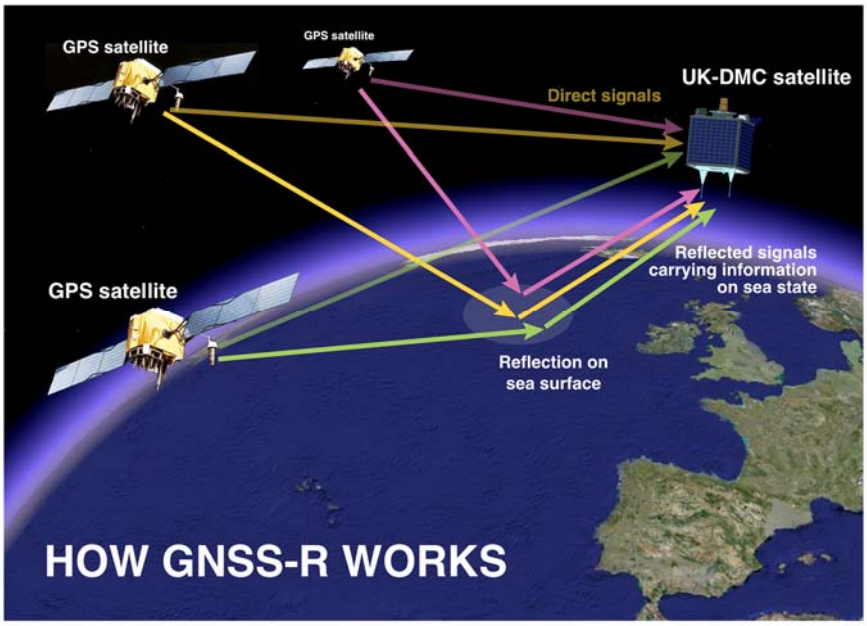


Figure 1. Measurement principle of GPS-Reflectometry over the ocean.

The CYGNSS mission is comprised of 8 microsatellites deployed into two planes of six S/C at inclinations of 35.0 and 37.5° at ~475 km altitude. Each S/C will be able to track up to four SP simultaneously, and generate a 1-second DDM for each SP. This results in 48 wind measurements (48 DDMs) per second across the globe, providing a wind field imagery of TC genesis, intensification and decay with unprecedented spatial and temporal resolution (Figure 2).

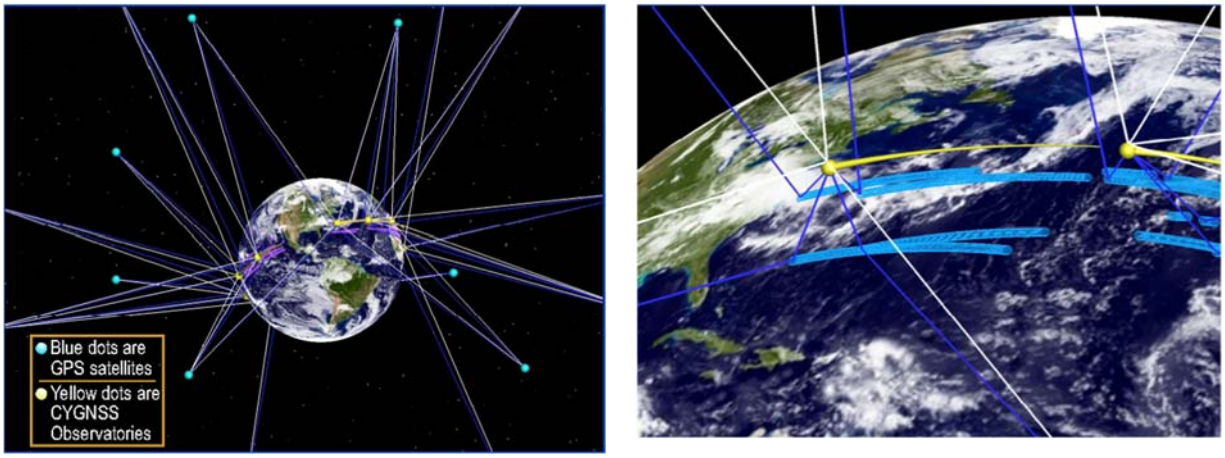


Figure 2. (left) overview of GPS Satellites and CYGNSS observatories. (right) illustration of the four simultaneous specular acquisitions for each observatory.

Each CYGNSS observatory will be equipped with a Digital Doppler Mapping Instrument (DDMI). The DDMI will generate DDMs continuously at a low data rate, which will provide a source for ocean roughness measurements across the ocean. In special situations, such as when passing over an active tropical cyclone, the instrument can be operated in Raw Data 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 usually 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 typical 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 3.

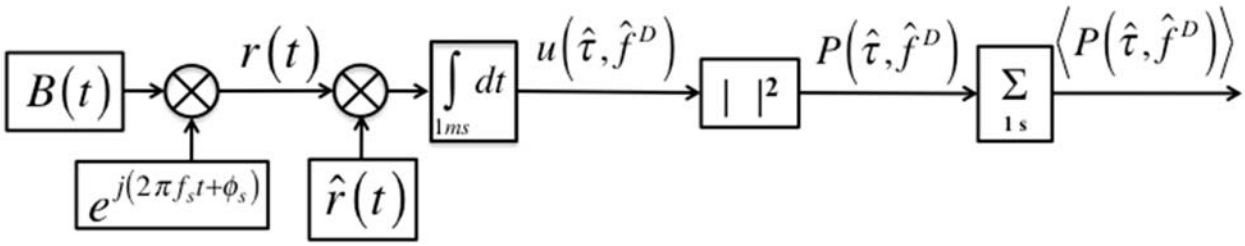


Figure 3. Schematic of a GPS receiver (modified from Gleason et al., 2005).

DDMs constitute the fundamental measurement for spaceborne GNSS-R, to which inversion algorithms can be therefore applied directly to estimate the sea surface wind and roughness.

2.3 Algorithm Approach

There are three popular types of 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 4.4, 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., 2012]), 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., 2013], and the width of the waveform obtained as DDM integration along the delays, known as Integrated Doppler Waveform, or IDoW [Clarizia et al., 2013b] 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 beforehand [Gleason, 2006]. This 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 Memorandums:

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

3.3 Delay Doppler Mapping Instrument

Each CYGNSS spacecraft will carry 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 Fig. 1.

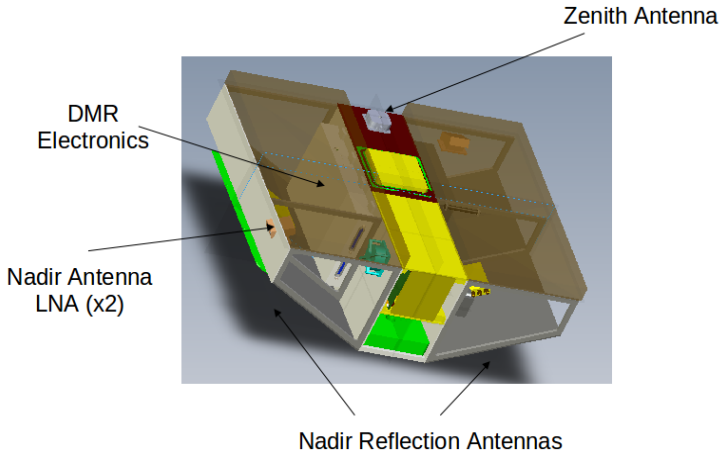


Figure 1. The Delay Doppler Mapping Instrument CYGNSS platform configuration.

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 a great deal of 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) are 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 values 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 will be interfaced to two downward looking nadir antennas, which together result in the surface footprint shown in Fig. 2.

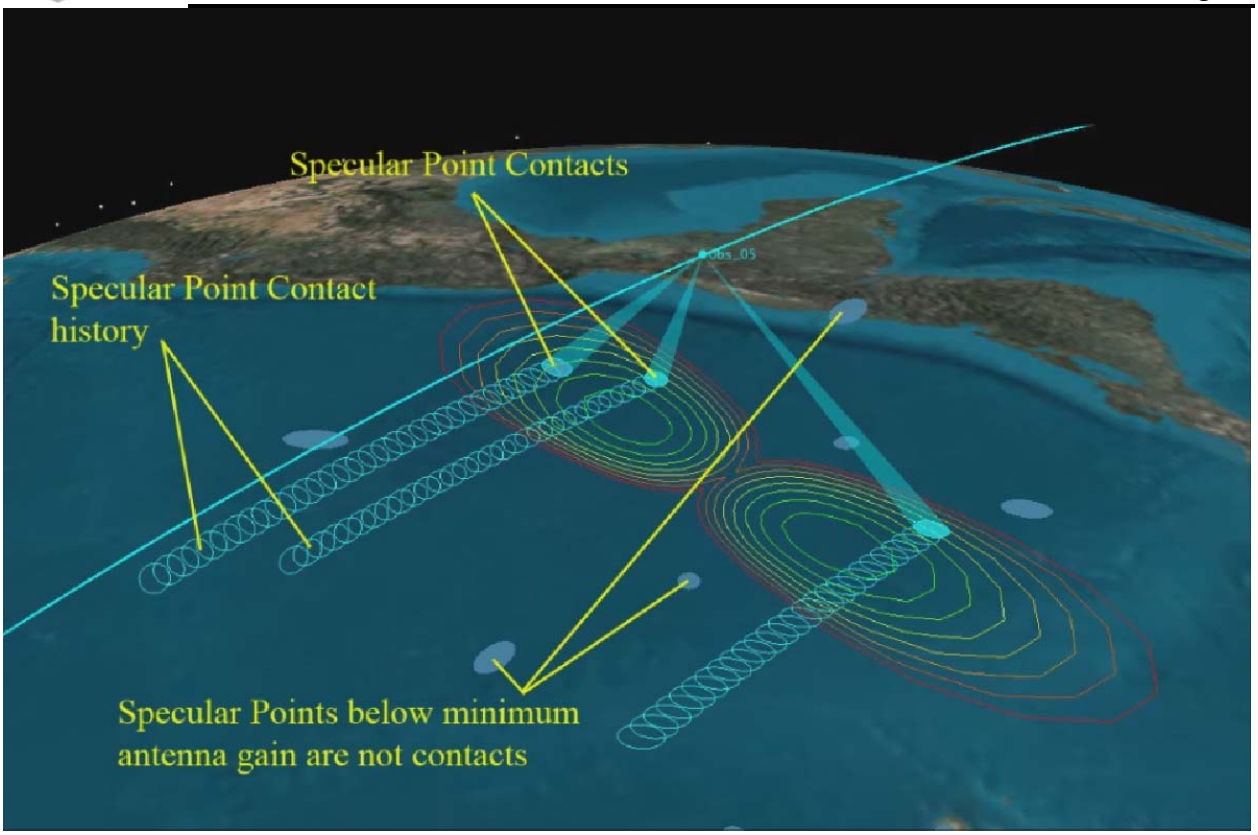


Figure 2, CYGNSS nadir antenna surface coverage footprint. The outermost (red) gain contour represents the threshold for acceptable received signal strength, inside of which scattered signals can be converted to near-surface wind speed with uncertainties that meet or exceed the Level 1 mission requirement.

Referring to Figs. 1-2, 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 "right" and "left" of the satellite direction of motion.

3.3.3 Front End Receiver and Backend Digital Processor

The complete DDMI processing chain is shown in Fig. 3, which includes both receiver RF front end components, on-board digital receiver back-end stages and ground based Level 1 data product processing steps.

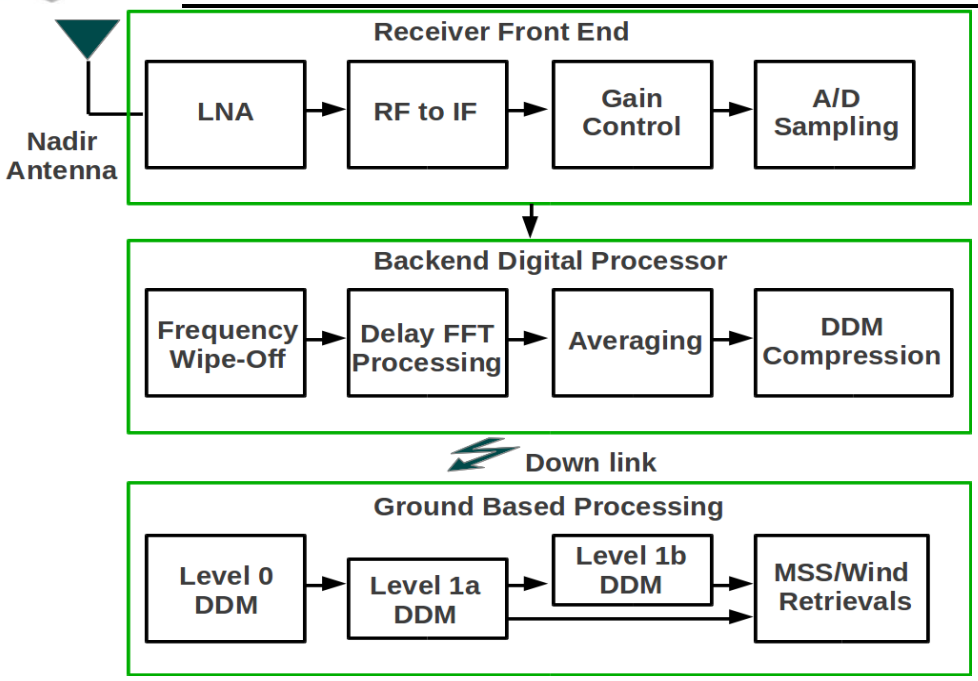


Figure 3, DDMI signal processing chain. Front end receiver, consisting of RF to IF and digitization stages, and the Back-end Digital Processor constitute the stages of DDM generation.

Upon capture 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 3, the hardware and processing stages of the DDMI processing chain are summarized below.

1. The signal is captured 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. Nominally, 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 [ref X]. 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 un-compressed, 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 will output (up to) 4 uncalibrated delay Doppler maps in uncalibrated counts to the spacecraft every second. These DDMs will then 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 2. The algorithms for converting Level 0 DDMs to Level 1 DDMs is described in detail in the Level 1 Calibration ATBD.

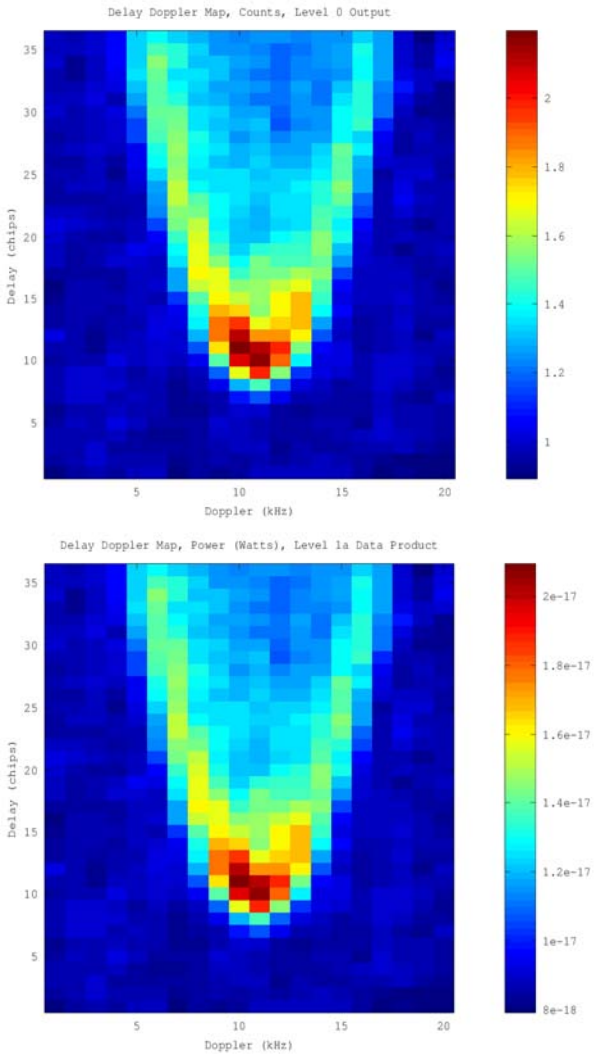


Figure 4, (top) raw Level 0 DDM in uncalibrated counts. (bottom) Level 1a calibrated DDM in units of Watts. DDMs generated by the CYGNSS End-to-End Simulator. Note the different magnitude scales before and after the Level 1a calibration.

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 will be 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 will consist of a DDM calibrated to the units of Watts (in 1 second).

Every DDM produced by the delay mapping receiver will be calibrated to Level 1a product power values. There are three types of calibration environments and scenarios, each with specific input signals 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 will be 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 will involve making measurements with the instrument switched to the black body load source as well as over well modeled ocean noise temperatures.

The beacon calibration will occur occasionally and involve 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 will be 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 will consist 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 will use an extensive set of external meta-data to convert the Level 1a power in Watts to a DDM map of BRCS values. This conversion will be 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 will also be tracked by the CYGNSS spacecraft navigation receiver, which will provide 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 will be 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 will be 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 will be 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 to 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

$$G_{rain} = \exp(-\alpha h(\csc\theta_t + \csc\theta_r)) \tag{4.1.1}$$

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

$$\alpha = aR^b \tag{4.1.2}$$

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 4.1.1 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).

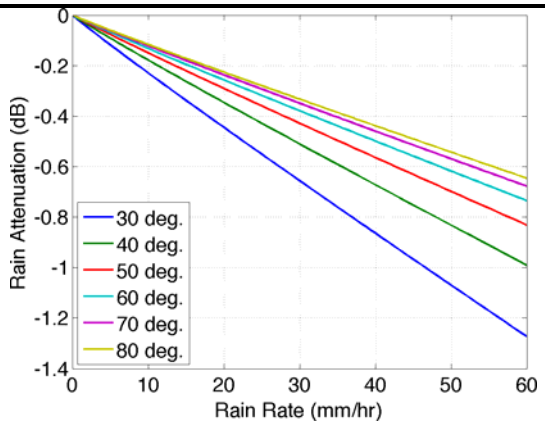


Figure 4.1.1 Rain attenuation versus rain rate for various elevation angles.

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 small regions of high 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 [1, 2]:

$$Y(t_0,\tau,f_c)=\int_0^{T_i} a(t_0+t')u(t_0+t'+\tau)\exp(2\pi i f_c t')dt' \qquad (4.1)$$

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 [3, 4], 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 (4.1), 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 [5]. 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 [2]:

$$u(\vec{R}_r,t)=\int D(\vec{\rho})a[t-(R_0(t)+R(t))/c]g(\vec{\rho},t)d^2\rho \tag{4.2}$$

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:

$$g(\vec{\rho},t) = -\Re(\vec{\rho})q^2 \exp\left[ik\left(R_0(t)+R(t)\right)\right]/4\pi i R_0 R q_z \tag{4.3}$$

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 (4.3) into (4.2), and then into (4.1), and assuming that integration over the accumulation time T_a is equivalent to averaging over a statistical ensemble of surface elevations

$$\left\langle \left|Y(\tau,f)\right|^2 \right\rangle = \frac{1}{T_a}\int_0^{T_a} \left|Y(t_0,\tau,f)\right|^2 dt_0 \tag{4.4}$$

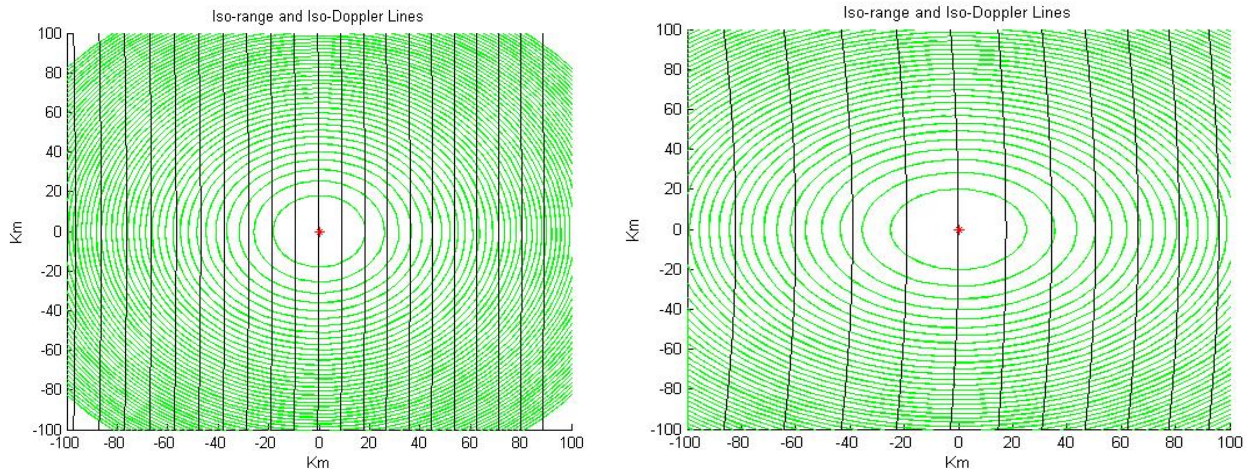
after making some additional assumptions, we arrive at the bistatic radar equation for the delay-Doppler map [2]

$$\left\langle \left|Y_s(t_0,\tau,f)\right|^2 \right\rangle = T_i^2 \frac{P_T G_T \lambda^2 G_R}{(4\pi)^3} \iint F(\vec{\rho})\Lambda^2(\tau,\vec{\rho})\left|S(f,\vec{\rho})\right|^2 R_0^{-2} R^{-2} \sigma_0(\vec{\rho}) d^2\rho. \tag{4.5}$$

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 (5.7) 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 (4.1) is similar to the WAF used in the unfocused SAR technique [6]. For fixed positions of the transmitter and the receiver both WAF and BRCS are functions of reference surface S

coordinates. Looking at (4.1) 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 Figs. 4.2.1a and b.



(a) (b)

Fig. 4.2.1

4.2.2 The Bistatic Radar Cross Section – Geometric Optics Approximation

The effect of surface roughness is described by σ_0, the normalized bistatic radar cross section (BRCS) of the rough surface. In the geometric-optics limit of the Kirchhoff approximation this term is represented by the following expression [5, 7]:

$$\sigma_0 = \pi \left|\Re\right|^2 \left(q / q_z\right)^4 P\left(-q_\perp / q_z\right) \tag{4.6}$$

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 [2]:

$$\Re = \frac{1}{2}\left[\frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} - \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}}\right] \tag{4.7}$$

where ε is the complex dielectric permittivity of sea water, and θ is the local incidence angle. According to Klein and Swift model [8]:

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 (4.6) 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 4.2 (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

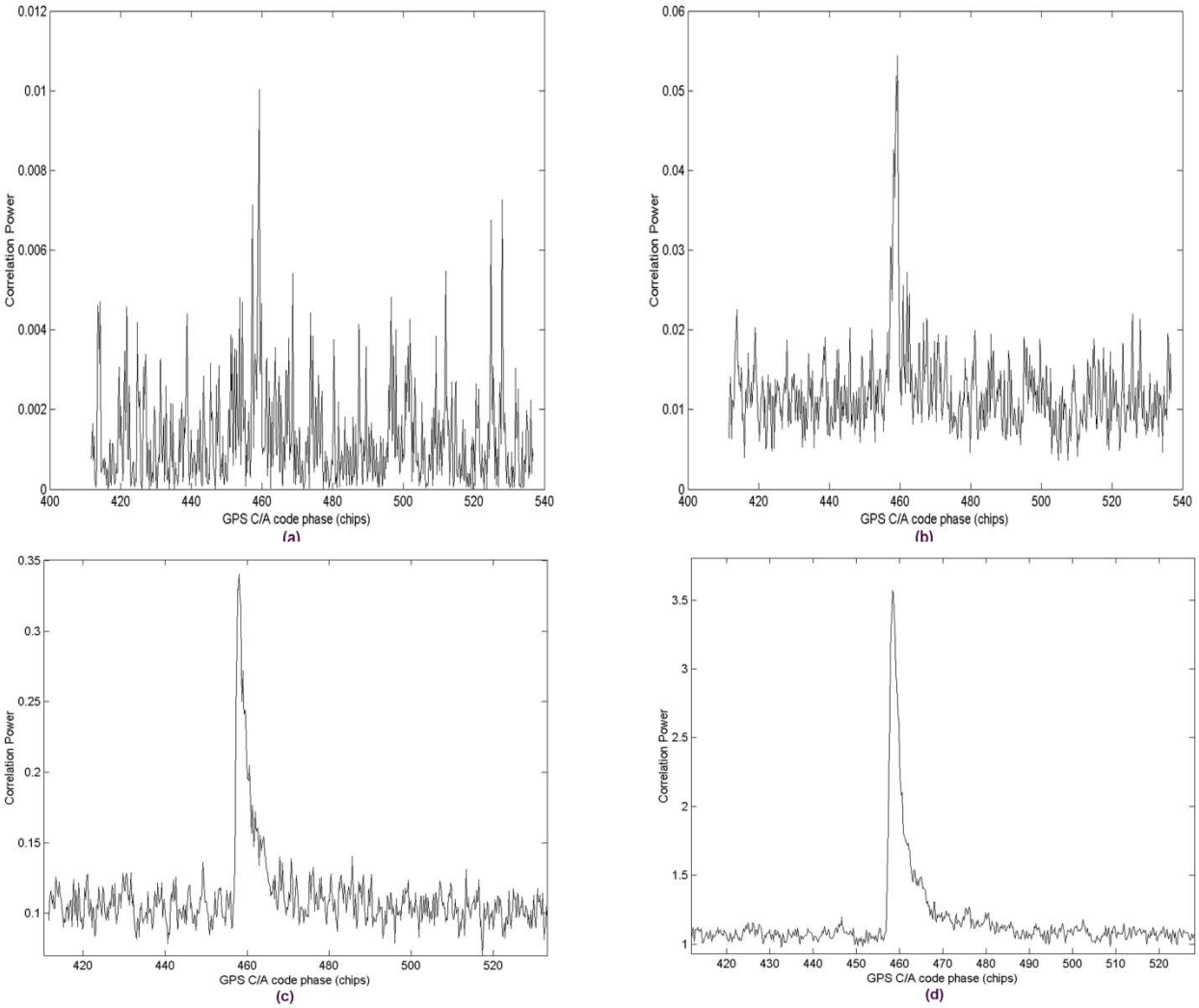


Fig. 4.2.2

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 (4.5) 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 (4.5) 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 [9-11].

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 [2,12,13]:

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{\det(M)}} \exp\left[-\frac{1}{2} \begin{pmatrix} s_x \\ s_y \end{pmatrix} M^{-1} \begin{pmatrix} s_x \\ s_y \end{pmatrix} \right] \tag{4.7}$$

where matrix M is

$$M = \begin{pmatrix} \cos\varphi_0 & -\sin\varphi_0 \\ \sin\varphi_0 & \cos\varphi_0 \end{pmatrix} \cdot \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix} \cdot \begin{pmatrix} \cos\varphi_0 & \sin\varphi_0 \\ -\sin\varphi_0 & \cos\varphi_0 \end{pmatrix} \tag{4.8}$$

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 (4.7) can be re-written in more common fashion:

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{mss_x mss_y (1-b_{x,y}^2)}} \exp\left[-\frac{1}{2(1-b_{x,y}^2)} \left(\frac{s_x^2}{mss_x} - 2b_{x,y} \frac{s_x s_y}{\sqrt{mss_x mss_y}} + \frac{s_y^2}{mss_y} \right) \right] \tag{4.9}$$

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:

$$mss_{x,y} = \left\langle s_{x,y}^2 \right\rangle = \iint\limits_{\kappa<\kappa_*} \kappa_{x,y}^2 \Psi(\vec{\kappa}) d^2\kappa \tag{4.10}$$

$$b_{x,y} = \left\langle s_x s_y \right\rangle / \sqrt{mss_x mss_y} \tag{4.11}$$

$$\left\langle s_x s_y \right\rangle = \iint\limits_{\kappa<\kappa_*} \kappa_x \kappa_y \Psi(\vec{\kappa}) d^2\kappa \tag{4.12}$$

One advantage of a Gaussian distribution is that the variance of slopes in (4.10) 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 do not exactly follow a Gaussian shape at their tails [14]. 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. [15]. The integrand in (4.10) is called a slope spectral density. An example of Elfouhaily et al. slope spectrum taken along the wind direction is shown in Fig. 4.2.3.

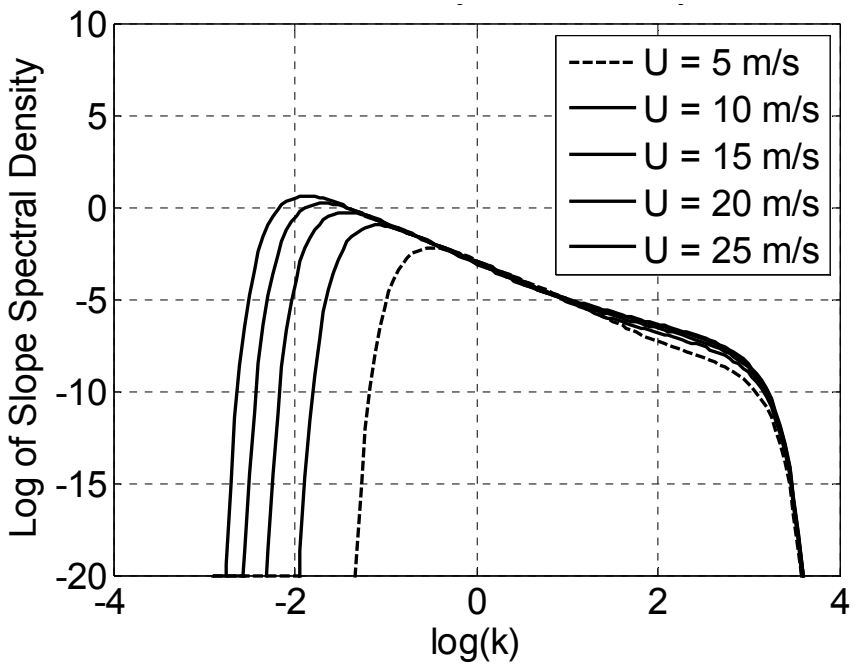


Fig. 4.2.3

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 [16], 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 (4.5). 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 G. Brown [17] 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 [2]. Later on, a reasonable $\kappa_* = \kappa\cos\theta/3$ on the incidence angle θ was assumed in [18]. In the paper [19], 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 (4.5) for various mss values will give the sought dependence mss vs wind speed. This approach was adopted in [20]. The empirical model from [20] gives the following expression:

$$\begin{aligned} mss_{\parallel} &= 0.45\cdot(0.00+0.00316)f(U) \\ mss_{\perp} &= 0.45\cdot(0.03+0.00192)f(U) \end{aligned} \tag{4.13}$$

where

$$f(U)=\begin{cases} U & 0.00 < U < 3.49 \\ 6\cdot\ln(U)-4.0 & 3.49 < U < 46 \\ 0.411\cdot U & 46.0 > U \end{cases} \tag{4.14}$$

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 [20] was rather arbitrary because GPS reflection data were not available for such high winds.

Below in Fig. 4.2.4 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 [18] and [19], and the empirical one from [20].

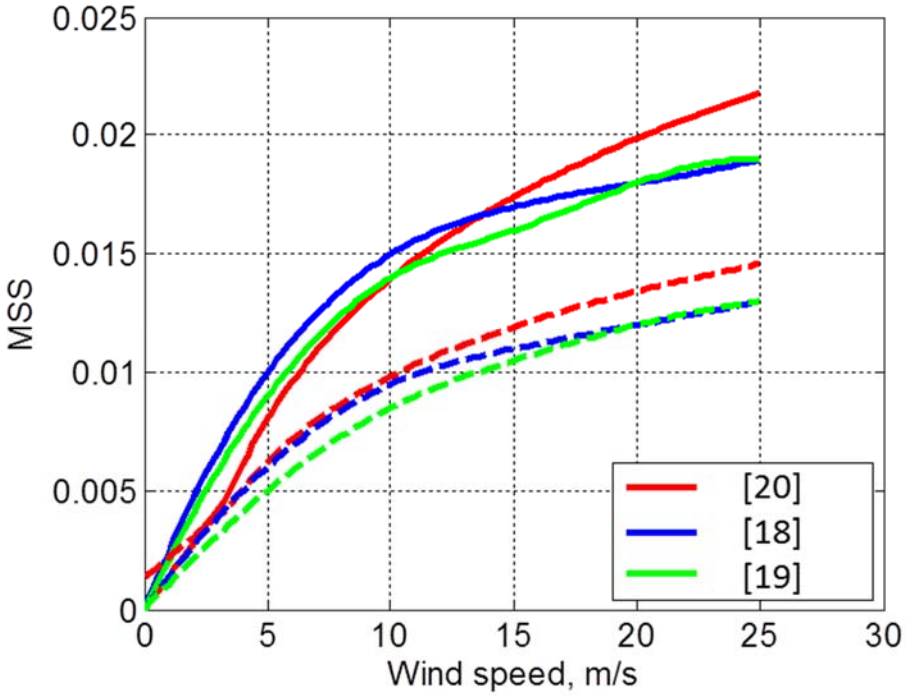


Fig. 4.2.4

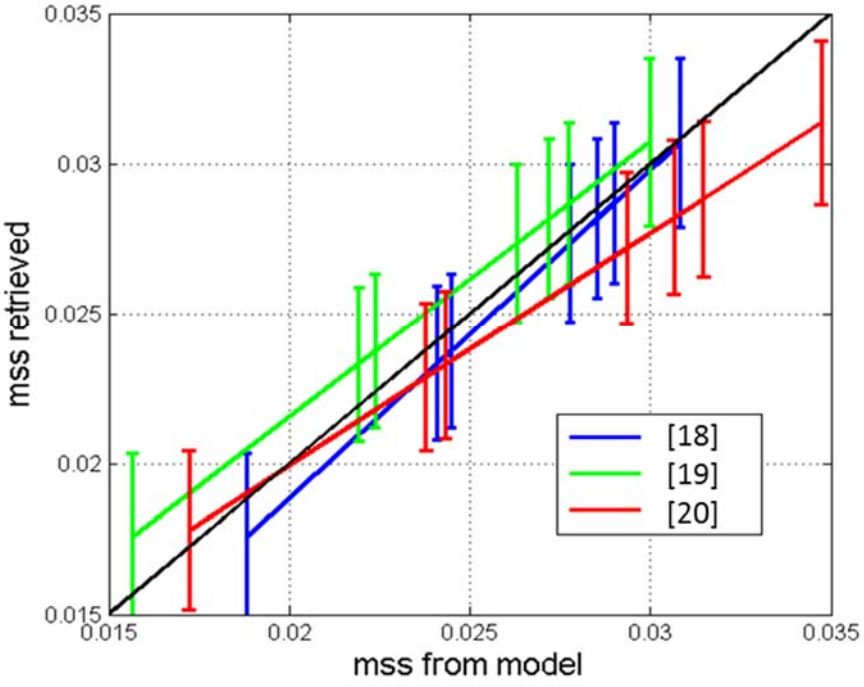


Fig. 4.2.5.

Figure 4.2.4 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 4.2.5 shows a comparison between three modeled mss and mss retrieved from DDM measurements during aircraft experiments [21, 22].

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.

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

The small slope approximation (SSA) was developed earlier in [23, 24] and was used successfully for solving various scattering and radiometric problems (see, e.g., [25-35]). The geometry of the scattering problem is shown in Fig.4.2.6. 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 wave lengths 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

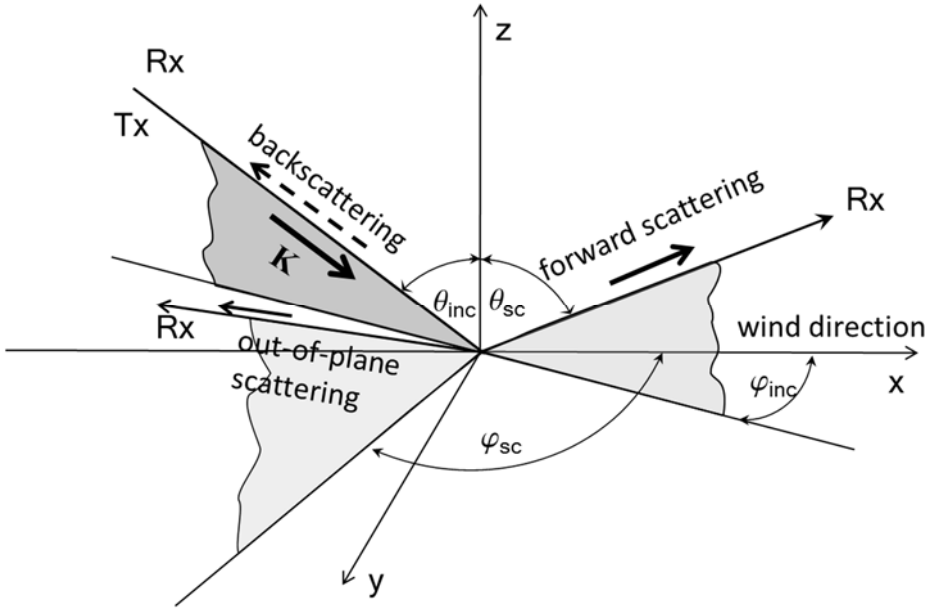


Fig. 4.2.6.

$h(\vec{r})$ 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 [27]:

$$\sigma_{\alpha\beta,\alpha'\beta'}\left(\vec{k},\vec{k}_0\right)=\frac{4q_k^2q_0^2}{\pi\left(q_k+q_0\right)^2}B_{\alpha\beta}\left(\vec{k},\vec{k}_0\right)\overline{B}^*_{\alpha'\beta'}\left(\vec{k},\vec{k}_0\right)\int_{r<r_{\max}}\exp\left[-i\left(\vec{k}-\vec{k}_0\right)\vec{r}-\left(q_k+q_0\right)^2\left(C(0)-C(\vec{r})\right)\right]d\vec{r} \tag{4.15}$$

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 (4.15) 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

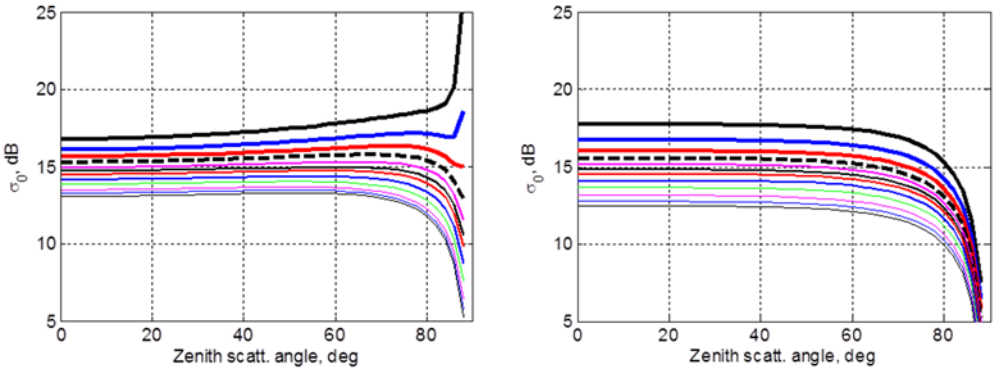
of the medium. Expressions for them can be found in [25]. Expressions for the LHCP bistatic scattering cross section can be expressed through corresponding cross sections for linear polarization as follows [36, 37]:

$$\sigma_{RL} = \frac{1}{4}\left\{\sigma_{11,11} + \sigma_{22,22} + \sigma_{12,12} + \sigma_{21,21} + 2\left[-\mathrm{Re}\,\sigma_{11,22} + \mathrm{Re}\,\sigma_{12,21} - \mathrm{Im}\left(\sigma_{11,12} + \sigma_{11,21} + \sigma_{12,22} + \sigma_{21,22}\right)\right]\right\} \tag{4.16}$$

We performed calculations of BRCS using (4.15) and (4.16) 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.

4.2.4. 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 [18], and another one using the empirical MSS model from [20]. 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 Fig. 4.2.7a, 4.2.7b, and 4.2.8, 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) VZ model **(b) SK model**

Fig. 4.2.7

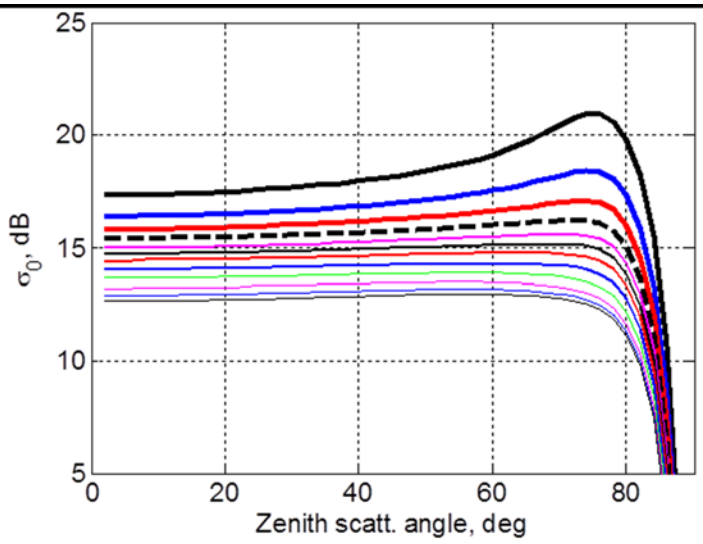


Fig. 4.2.8 SSA model

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 [20] 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 Fig. 4.2.9.

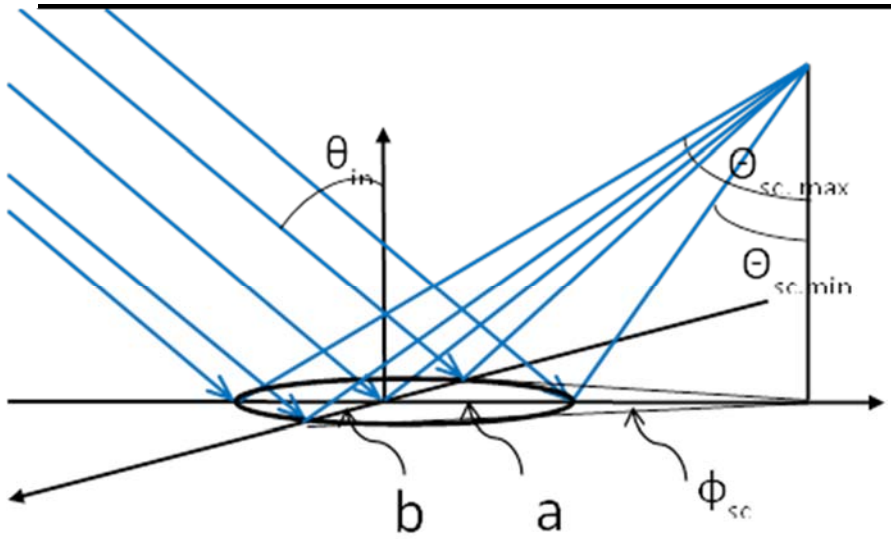


Fig. 4.2.9

Notations in Fig. 4.2.9 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} :

$$a_n = b_n / \cos\theta_{in}, \quad b_n = \left(2nlH / \cos\theta_{in}\right)^{1/2}.$$

Fig. 4.2.10 shows how angles $\theta_{sc,\min}$, $\theta_{sc,\max}$ and ϕ_{sc} from Fig. 4.2.9 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 θ_{in} = 30°, *H* = 600 km, and *l* = 300 m (C/A code).

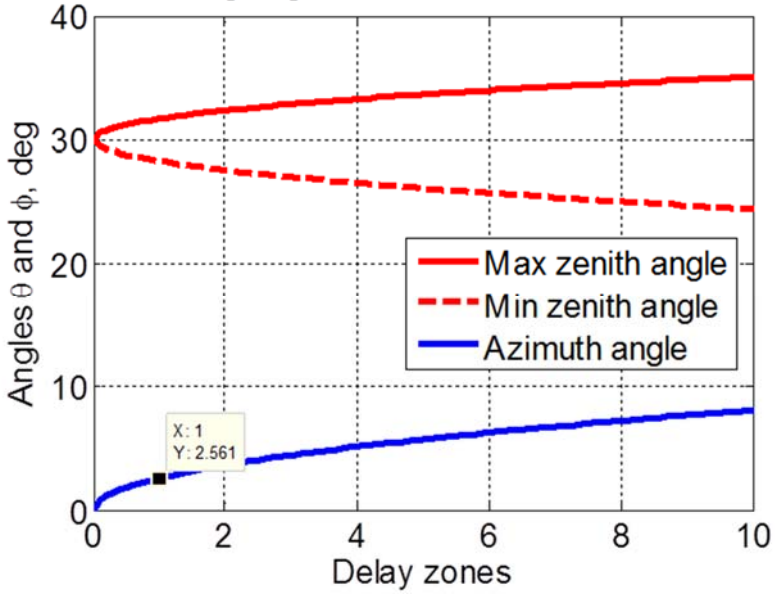


Fig. 4.2.10

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 Figs. 4.2.11(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) $\theta_{sc} = 30^\circ, \phi_{sc} = 0^\circ$; (b) $\theta_{sc} = 30^\circ, \phi_{sc} = 2.56^\circ$; (c) $\theta_{sc} = 28^\circ, \phi_{sc} = 0^\circ$; and (d) $\theta_{sc} = 32^\circ, \phi_{sc} = 0^\circ$. 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.

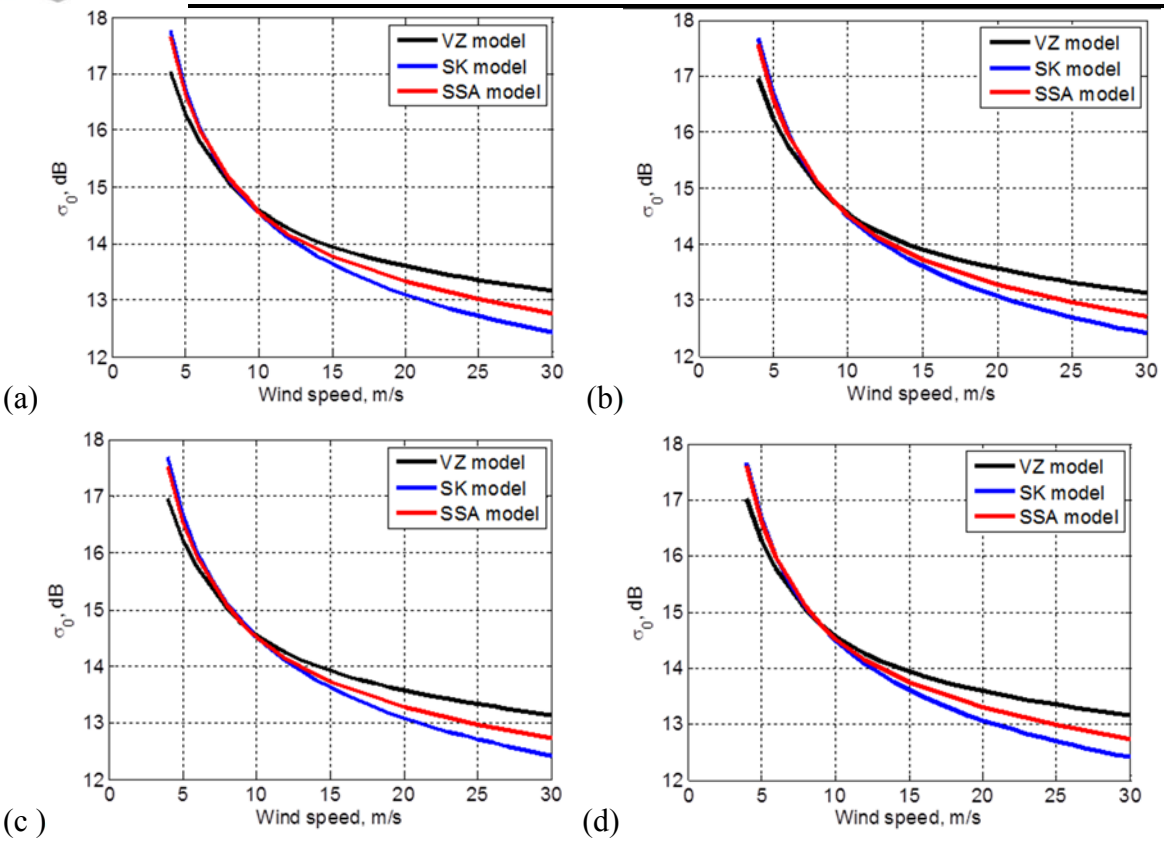


Fig. 4.2.11

One can see that curves in Fig. 4.2.11 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 Figs. 4.2.12(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) $\theta_{sc} = 30^\circ, \phi_{sc} = 0^\circ$ (this plot repeats plot (a) from the previous figure; it is given for comparison purposes); (b) $\theta_{sc} = 30^\circ, \phi_{sc} = 8^\circ$; (c) $\theta_{sc} = 25^\circ, \phi_{sc} = 0^\circ$; and (d) $\theta_{sc} = 35^\circ, \phi_{sc} = 0^\circ$.

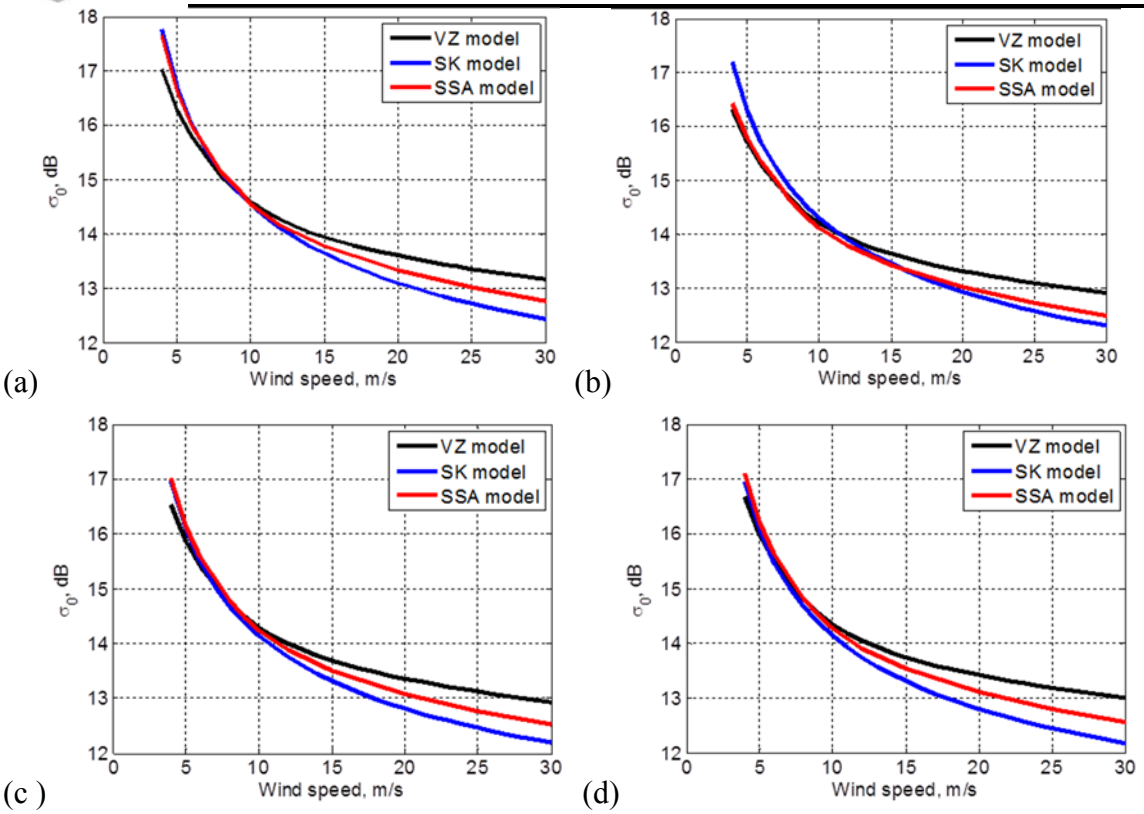


Fig. 4.2.12

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.

4.2.5. Conclusions

In this Section 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., [38]) 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 [39] developed at the NOAA National Centers for Environmental Prediction (NCEP) which is now widely used for hurricane long-wave predictions (see, e.g., [40]). 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.

References (Section 4.2)

1. 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.
2. Zavorotny, V. U., and A. G. Voronovich, "Scattering of GPS signals from the ocean with wind remote sensing application," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, pp. 951-964, 2000.
3. Gleason, S.T., S. Hodgart, S. Yiping, C. Gommenginger, S. Mackin, M. Adjrad and M. Unwin, "Detection and processing of bistatically reflected GPS signals from low Earth orbit for the purpose of ocean remote sensing," *IEEE Trans. Geosci. Remote Sens*., Vol. 43, No. 6, June 2005, pp. 1229-1241.
4. Gleason, S., *Remote Sensing of Ocean, Ice and Land Surfaces Using Bistatically Scattered GNSS Signals From Low Earth Orbit*, PhD Thesis, University of Surrey, 2006.
5. F. G. Bass and I. M. Fuks, *Wave Scattering From Statistically Rough Surfaces*. New York: Pergamon, 1979.
6. Elachi, C., *Spaceborne Radar Remote Sensing: Applications and Techniques*, IEEE Press, New York, 1988.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Soulat, F., *Sea Surface Remote Sensing With GNSS and Sunlight Reflections*. Ph.D. Thesis. Universitat Politecnica de Catalunya, 2004.
13. 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.
14. 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.
15. T. Elfouhaily, B. Chapron, K. Katsaros, and D. Vandemark, "A unified directional spectrum for long and short wind-driven waves," *J. Geophys. Res.*, vol. 102, pp. 15 781–15 796, 1997.
16. 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.
17. Brown, G.S., "Backscattering from a Gaussian-distributed, perfectly conducting rough surface," *IEEE Trans. Antennas Propag.*, Vol. AP-26, pp. 472–482, May 1978.

18. Garrison, J.L., A. Komjathy, V.U. Zavorotny and S.J. Katzberg, "Wind speed measurements using forward scattered GPS signals," *IEEE Trans. Geosci. Remote Sens.*, Vol. 40, No. 1, pp. 55–65, Jan 2002.
19. 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.
20. Katzberg, S.J., Torres, O. and G. Ganoe. "Calibration of reflected GPS for tropical storm wind speed retrievals," *Geophys. Res. Lett.*, *33*, L18602, doi:10.1029/2006GL026825, 2006.
21. 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.
22. 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).
23. 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.
24. Voronovich, A.G., *Wave Scattering from Rough Surfaces,* 2nd edn (Berlin: Springer), 236p, 1999.
25. 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.
26. 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.
27. 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.
28. 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.
29. C. Bourlier, N. Déchamps, and G. Berginc, "Comparison of asymptotic backscattering models (SSA, WCA, and LCA) from one-dimensional Gaussian ocean-like surfaces, *IEEE Trans.Antennas Propag.,*vol. 53, no. 5, pp. 1640-1653, May 2005.
30. 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.
31. 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.
32. 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.
33. 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.
34. J.T. Johnson and J. Ouellette, "A study of polarization features in bistatic scattering from rough surfaces," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 3, pp. 1616–1626, Mar 2014.

35. 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.
36. 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.
37. C. Zuffada, A. K. Fung, J. Parker, M. Okolicanyi, and E. Huang, Polarization properties of the GPS signal scattered off a wind driven ocean, *IEEE Trans. Antennas Propag*., vol. 52, no. 1, pp. 172–187, Jan. 2004.
38. 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.
39. 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.
40. 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.

4.3 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, as illustrated in section 4.3 [Clarizia, 2012].

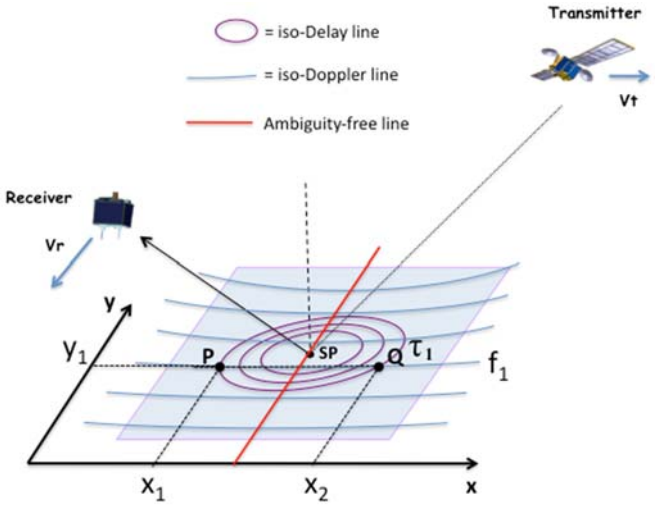


Figure 4. Delay-Doppler coordinates in GNSS-R and their relation to the space coordinate system.

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 (see section 4.3), 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 4, , 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 4. 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 4.

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 5 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).

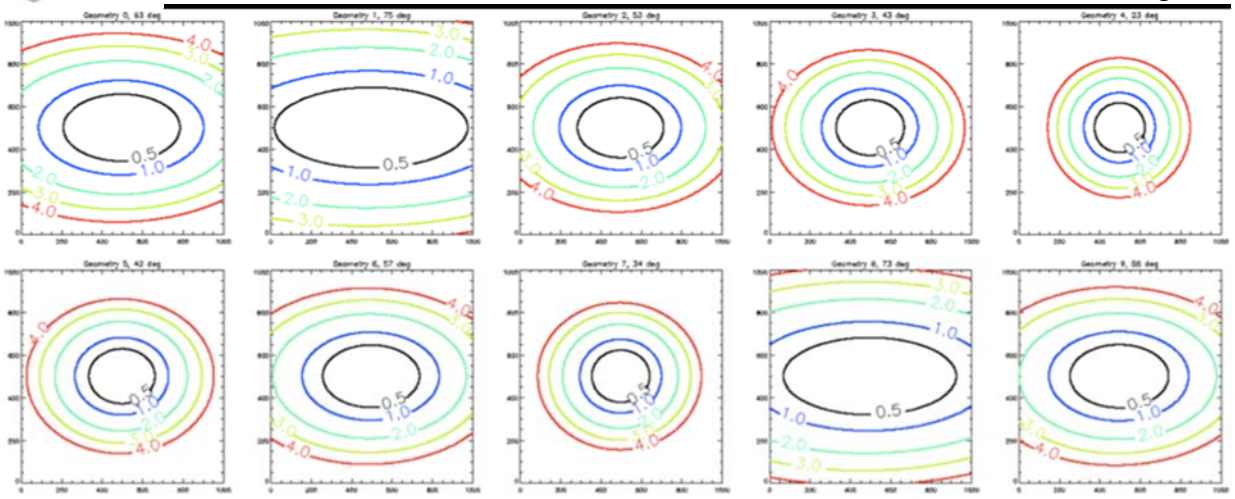
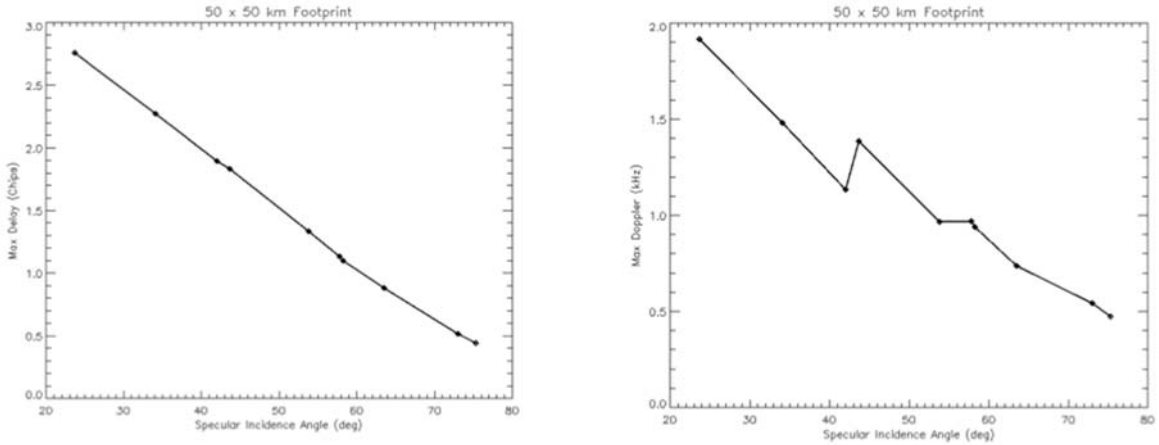


Figure 5. Iso-delay contours over a footprint of 100 x 100 km, for the following incidence angles (top, left to right): 63°, 75°,53°,43°,23°; (bottom, left to right): 42°, 57°, 34°, 73°, 58°.

Figure 6 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 Figure6(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 6. (a) maximum delay as a function of incidence angle, within a 50 x 50 km^2 footprint; (b) maximum Doppler as a function of incidence angle, within the same footprint.

4.4 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:

$$u(t) = s(t) + n(t), \tag{4.4.1}$$

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:

$$Y(t_0, \tau) = \int_0^{T_i} a(t_0 + t')u(t_0 + t' + \tau)dt'. \tag{4.4.2}$$

Taking into account (5.1), we obtain from (5.2) that

$$Y(t_0, \tau) = Y_s(t_0, \tau) + Y_n(t_0, \tau), \tag{4.4.3}$$

where

$$Y_s(t_0, \tau) = \int_0^{T_i} a(t_0 + t')s(t_0 + t' + \tau)dt', \tag{4.4.4}$$

$$Y_n(t_0, \tau) = \int_0^{T_i} a(t_0 + t')n(t_0 + t' + \tau)dt'. \tag{4.4.5}$$

Therefore, the quantities in (5.4) and (5.5) 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 $\left|Y(t_0,\tau)\right|^2$ over a long enough observation time, so both thermal and surface-induced fluctuations are substantially averaged out. The result is

$$\left\langle \left|Y(t_0,\tau)\right|^2 \right\rangle = \left\langle \left|Y_s(t_0,\tau)\right|^2 \right\rangle + \left\langle \left|Y_n(t_0,\tau)\right|^2 \right\rangle. \tag{4.4.6}$$

4.4.1 *The Signal Term*

The first term in Eq. (4.4.6) is known in the literature [1] as the GPS radar bistatic equation:

$$\begin{aligned} \left\langle \left|Y_s(t_0,\tau)\right|^2 \right\rangle &= T_i^2 \frac{P_T G_T \lambda^2 G_R}{(4\pi)^3} \\ &\times \iint F(\vec{\rho})\Lambda^2(\tau,\vec{\rho})\left|S(f_{dop},\vec{\rho})\right|^2 R_0^{-2} R^{-2} \sigma_0(\vec{\rho}) d^2\rho. \end{aligned} \tag{4.4.7}$$

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; $\left|S(f_{dop},\vec{\rho})\right|^2$ 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 cross section of the rough surface, and generally it is a function of two angles, the incidence angle and the scattering angle. Here, in (5.7) 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 zones (function $\left|S(f_{dop},\vec{\rho})\right|^2$). 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., [2]). With regard to our problem it has the following form [3]

$$\sigma_0(\vec{\rho}) = \frac{\pi |\Re|^2 q^4}{q_z^4} P\left(-\frac{\vec{q}_\perp}{q_z}\right). \quad (4.4.8)$$

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:

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{mss_x mss_y (1-b_{x,y}^2)}} \exp\left[-\frac{1}{2(1-b_{x,y}^2)}\left(\frac{s_x^2}{mss_x} - 2b_{x,y}\frac{s_x s_y}{mss_x mss_y} + \frac{s_y^2}{mss_y}\right)\right] \quad (4.4.9)$$

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:

$$mss_{x,y} = \left\langle s_{x,y}^2 \right\rangle = \iint\limits_{\kappa<\kappa_*} \kappa_{x,y}^2 W(\vec{\kappa}) d^2\kappa \quad (4.4.10)$$

$$b_{x,y} = \left\langle s_x s_y \right\rangle / \sqrt{mss_x mss_y} \quad (4.4.11)$$

$$\left\langle s_x s_y \right\rangle = \iint\limits_{\kappa<\kappa_*} \kappa_x \kappa_y W(\vec{\kappa}) d^2\kappa \quad (4.4.12)$$

Here, $W(\vec{\kappa})$ is the surface elevation spectrum. For calculations we adopt the Elfouhaily spectrum [3]. 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 [4], since the author claimed that it gave the best fit with measurements:

$$\kappa_* = k\cos\theta\left(1 + U_{10}/20\right)/7.5 \quad (4.4.13)$$

In (5.8) also enters $\Re$, the LHCP Fresnel reflection coefficient for sea water at L-band [1]:

$$\Re = \frac{1}{2}\left[\frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} - \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}}\right] \tag{4.4.14}$$

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 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.

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

$$\sigma_0 = \frac{|\Re|^2}{2\sqrt{mss_x mss_y (1 - b_{x,y}^2)}} \tag{4.4.15}$$

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 $\Re$.

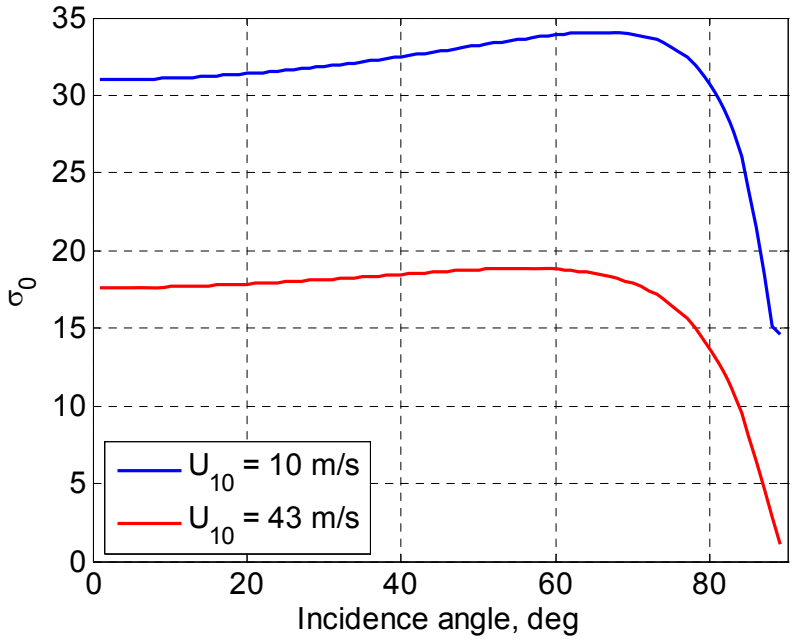


Figure 1. The bistatic radar cross section at the specular direction.

4.4.2 *Thermal Noise*

The second term in Eq. (4.4.6) is the background noise term. It can be written as double integral over the coherent integration time:

$$\left\langle \left| Y_n(t_0,\tau,f) \right|^2 \right\rangle = \int_0^{T_c} dt' \int_0^{T_c} dt'' a(t_0+t',f)a(t_0+t'',f). \times \left\langle n(t_0+\tau+t')n^*(t_0+\tau+t'') \right\rangle. \tag{4.4.16}$$

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

$$\left\langle n(t')n^*(t'') \right\rangle = kT^\circ B_n b_n (t'-t''), \tag{4.4.17}$$

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;

$$b_n(t) = \int W_n(f)\exp(2i\pi ft)df; \qquad \int_0^{T_i} b_n(t)dt = T_{cor}. \tag{4.4.18}$$

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:

$$\left\langle \left|Y_n(t_0,\tau,f)\right|^2 \right\rangle = T_i^2 kT^{\circ} B_D, \tag{4.4.19}$$

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

$$\begin{aligned}\left\langle Y_n(t_0,\tau,f)Y_n^*(t_0,\tau',f')\right\rangle = \int_0^{T_i} dt' \int_0^{T_i} dt'' a(t_0+t',f)a^*(t_0+t'',f').\\ \times\left\langle n(t_0+\tau+t')n^*(t_0+\tau'+t'')\right\rangle.\end{aligned} \tag{4.4.20}$$

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. (4.4.11) can be reduced to

$$\left\langle Y_n(t_0,\tau,f)Y_n^*(t_0,\tau',f')\right\rangle = T_i^2 kT^{\circ} B_D \Lambda^2(d\tau)\left|S(df)\right|^2 \tag{4.4.21}$$

In the forward model, zero mean white Gaussian noise is generated with respect to delay and Doppler using the power level in (4.4.10) 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:

$$SNR = \left\langle \left| Y_s(t_0,\tau) \right|^2 \right\rangle \Big/ \left\langle \left| Y_n(t_0,\tau) \right|^2 \right\rangle. \qquad (4.4.22)$$

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.

4.4.2.1 *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, according to Eq.(5.1)

$$U(t) \equiv |u(t)|^2 = [s(t)+n(t)][s^*(t)+n^*(t)] = S(t)+N(t)+C(t)+C^*(t), \qquad (4.4.23)$$

where

$$C(t) = s(t)n^*(t). \qquad (4.4.24)$$

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):

$$\overline{U}(t) \equiv \overline{S}(t) + \overline{N}(t) + \overline{C}(t) + \overline{C}^*(t) = \frac{1}{T}\int_{-T/2}^{T/2}\left[S(t+t') + N(t+t') + C(t+t') + C^*(t+t')\right]dt'. \tag{4.4.25}$$

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:

$$\widetilde{S}(t) = \overline{U}(t) - \overline{N}(t_0). \tag{4.4.26}$$

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

$$\sigma_{\widetilde{S}}^2 = \sigma_{\overline{U}}^2 + \sigma_{\overline{N}}^2. \tag{4.4.27}$$

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

$$\left\langle \overline{U} \right\rangle = \left\langle S \right\rangle + \left\langle N \right\rangle. \tag{4.4.28}$$

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

$$\sigma_{\overline{U}}^2 = \frac{1}{T^2}\left\langle \left| \int_{-T/2}^{T/2} \left[S(t') + N(t') + C(t') + C^*(t') \right] dt' \right|^2 \right\rangle - \left\langle \overline{U} \right\rangle^2. \tag{4.4.29}$$

The variance of the noise power is:

$$\sigma_{\bar{N}}^2 = \frac{1}{T^2}\left\langle \left| \int_{-T/2}^{T/2} N(t')dt' \right|^2 \right\rangle - \langle N \rangle^2 . \qquad (4.4.30)$$

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:

$$\frac{\sigma_{\tilde{S}}}{\langle S \rangle} = \left[\left(1 + 2\langle S \rangle + 2T_{\text{int}} / \tau_{cor} \langle S \rangle^2 \right) / N \right]^{1/2} . \qquad (4.4.31)$$

For $T >> \tau_{\text{cor}}$ 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 [5-7].

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:

$$r_{cor} = \lambda R / D \qquad (4.4.32)$$

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, (5.19) simplifies:

$$\frac{\sigma_{\tilde{S}}}{\langle S \rangle} = \left(1 + 2\langle S \rangle^{-1} + 2T_{\text{int}} \tau_{cor}^{-1} \langle S \rangle^{-2} \right)^{1/2} N^{-1/2} \approx \frac{1 + 1/\langle S \rangle}{\sqrt{N}} . \qquad (4.4.33)$$

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

$$\sigma_{SNR} == \frac{\langle S \rangle + 1}{\sqrt{N}}. \tag{4.4.34}$$

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

$$\sigma_{SNR} = \frac{\sqrt{\langle S \rangle^2 + 2\langle S \rangle + 2T_{\text{int}}\tau_{cor}^{-1}}}{\sqrt{N}} \tag{4.4.35}$$

4.5 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 4.1. The expected value of the power of the reflected signal versus delay and Doppler can be rewritten as

$$\left\langle \left| Y_s(t_0, \tau, f) \right|^2 \right\rangle = \iint H(\vec{\rho}) \Lambda^2(\tau, \vec{\rho}) \left| S(f, \vec{\rho}) \right|^2 d^2\rho, \tag{4.5.1}$$

where

$$H(\vec{\rho}) = T_i^2 \frac{P_T G_T \lambda^2 G_R}{(4\pi)^3} R_0^{-2} R^{-2} \sigma_0(\vec{\rho}) \tag{4.5.2}$$

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 4.5.1). 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 4.5.2 and include a time varying phase term $\phi(t,\rho)$ to make the contribution complex,

$$h(t,\rho)=\sqrt{H(\rho)}e^{j\phi(t,\rho)}. \tag{4.5.3}$$

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

$$Y_s(t,\tau,f)=\iint h(t\ ,\vec{\rho})\Lambda(\tau,\vec{\rho})S(f,\vec{\rho})d^2\rho. \tag{4.5.4}$$

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

$$\left|Y_s(t_0,\tau,f)\right|^2=\int_{t_0}^{t_0+T} Y_s(t,\tau,f)Y_s^*(t,\tau,f)dt, \tag{4.5.5}$$

where T =1. The phase term $\phi(t,\rho)$ must be chosen such that the expectation of equation 4.5.5 is equal to 4.5.1. 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,

$$\phi(\rho) = \phi_0(\rho) + \frac{2\pi}{\lambda} R(t,\rho), \tag{4.5.6}$$

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 $f_D(t_0,\rho)$,

$$R(t,\rho) = R(t_0,\rho) - (t-t_0)\lambda f_D(t_0,\rho). \tag{4.5.7}$$

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.

References

41. Zavorotny, V. U., and A. G. Voronovich, "Scattering of GPS signals from the ocean with wind remote sensing application," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 951-964, 2000.
42. 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.
43. Elfouhaily, T., B. Chapron, K. Katsaros, and D. Vandemark, "A unified directional spectrum for long and short wind-driven waves." *J. Geophys. Res*.102:15781-15796, 1997.
44. 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.
45. 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," *Proc. IGARSS'01,* pp. 3332-3334, 2001.
46. You, H. G. Heckler, J. L. Garrison, and V. U. Zavorotny (2004), Stochastic voltage model and experimental measurement of ocean-scattered GPS signal statistics, *IEEE Trans Geosci. Remote Sens*., vol. 42, No. 10, pp.2160-2169.
47. 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.

48. Edgar, C., D.B. Goldstein, P. Bentley, "Current Constellation GPS Satellite Ground Received Signal Power Measurements," ION NTM 2002, 28-30 January 2002, San Diego, CA.
49. Katzberg, S.J., Torres, O. and G. Ganoe. "Calibration of reflected GPS for tropical storm wind speed retrievals," *Geophys. Res. Lett.*, *33*, L18602, doi:10.1029/2006GL026825, 2006.

5. L2 Wind Speed Retrieval Algorithm

The L2 wind speed retrieval algorithm described here is a revised and improved version of the algorithm illustrated in [Clarizia et al., 2013]. While the overall approach remains the same (i.e. regression-based wind retrieval using empirical GMFs), the present algorithm contains a number of improvements to the calculation of the observables, as well as to the actual retrieval method. 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 scattering area, which are generated as explained in [Gleason, 2014];
2. A population of simulated samples, produced by the CYGNSS Project End-to-End Simulator (E2ES) applied to a 13-day nature run of tropical cyclone and background wind fields, is split into training and test subsets, using odd and even minutes of sample time;
3. A statistical inversion algorithm is trained using the training data subset, by constructing an empirical Geophysical Model Function (GMF) from measured and truth matchups;
4. De-biasing is applied to wind retrievals produced by the DDMA and LES observables;
5. The test data subset is mapped to retrieved wind speed using a Minimum Variance (MV) Estimator;
6. 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);
7. Samples with effective Field of View (EFOV) above the spatial resolution requirement are removed (EFOV filter);
8. The performance of the retrieval algorithm is evaluated, in terms of Root Mean Square (RMS) Error between the true wind and the retrieved wind;

Steps 2, 3, 7 and 8 represent new additions to the previous (pre-release) version of the retrieval algorithm. They significantly improve the overall performance, as will be shown later. Each step is described in detail in the subsequent paragraphs.

5.1 DDM Observables: DDMA and LES

Here we present a definition of the observables we use for our retrieval algorithm, along with the overall set of DDMs from which they are calculated, and the L2a corrections applied to the DDMs. The DDMs used have all been simulated under realistic conditions, using the CYGNSS E2ES.

5.1.1 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. 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 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 7.

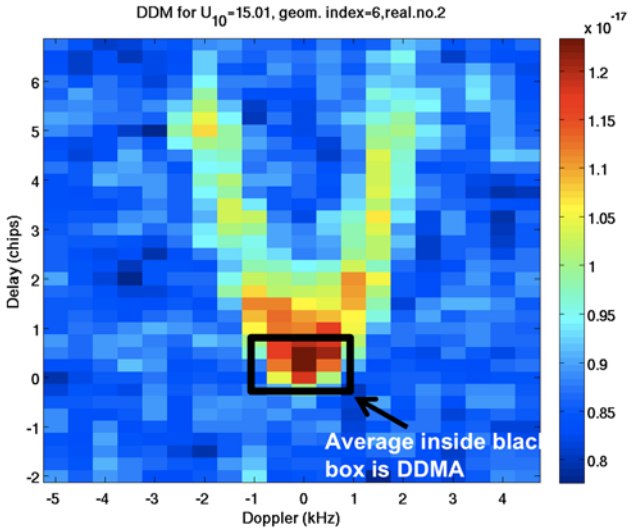


Figure 7. A simulated DDM and the DDM area where the DDMA is calculated.

The DDMA is calculated from the DDM after the additive noise floor has been removed. The noise floor is estimated from the DDM in a region where there is no signal (i.e. at delay values just before the specular point and before entering the horseshoe shape), and is subtracted from each DDM pixel. The noise floor was originally computed as the average of the DDM region over a number of rows where no signal is present (i.e. for all the Doppler frequencies, and for chips<0). In particular, the first four rows were considered for the noise floor computation, since the noise floor begins to rise if more rows are included in the computation, suggesting that some signal is present in those rows [Clarizia et al., 2013c]. The four rows used correspond to chip delays from -2 chips to -1.25 chips. However, some recent investigations on the statistics of each individual row have highlighted a non-gaussian behavior of the distribution of power values for the first three of these rows (chip delays from -2 chip to -1.5 chips), as illustrated in Figure 8. This is unexpected, since the DD region where no signal is present should be only affected by thermal noise which is Gaussian distributed. As opposed to the first three rows, the fourth row (chip delay of -1.25 chips) exhibits the expected gaussian distribution, and it is therefore the only one used to estimate the noise floor.

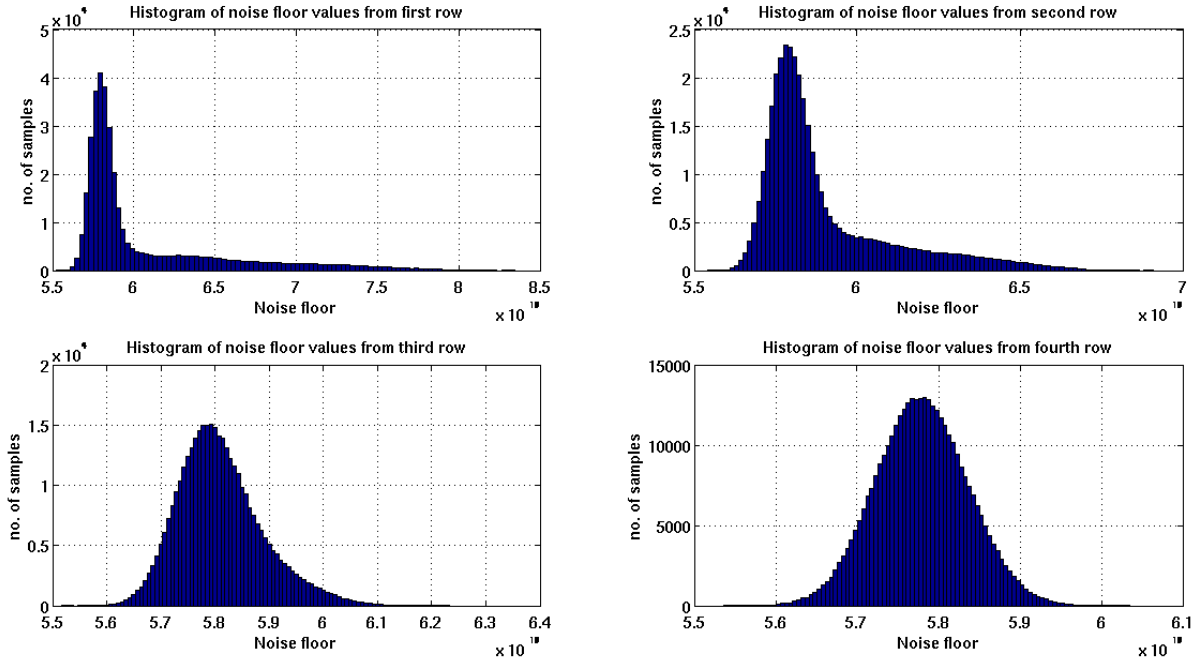


Figure 8. Histograms of noise floor computed for individual rows of the DDMs. The noise floor values used for the histograms have been computed from a 3-hour subset of DDMs taken from the full 13-day Nature Run.

A static noise floor value has been estimated as the mean of the averaged power values of row 4 computed for a large number of DDMs from a 3-hour subset of data, selected from the full 13-day Nature Run. This value is then subtracted from all the DDMs of the Nature Run, prior to computing the observables.

5.1.2 Definition of Leading Edge Slope (LES)

The Leading Edge Slopes (LES, previously called Delay Waveform Slope, or DWS in [Clarizia et al., 2013c], is the slope of the leading edge of the Integrated Delay Waveform (IDW). IDWs are obtained as incoherent integration of DDMs of NRCS along the Doppler dimension, and over a range of Doppler frequencies. An illustration of IDWs, and how the slope of their rising edge differ for different wind speeds, is illustrated in Figure 9.

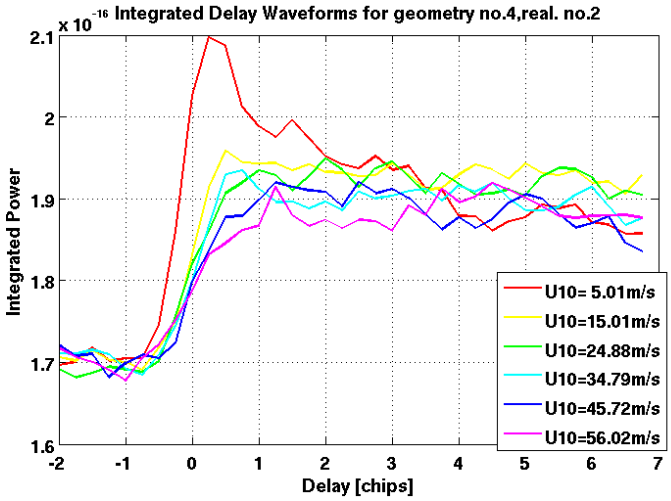


Figure 9. Integrated Delay Waveforms computed from DDMs simulated using different wind speeds

The DDMA is computed in practice as the average of the L1b DDM of Radar Cross Section (RCS) over a selected DD window, divided by the average of the DDM of scattering area over the same window. Similarly, the LES is calculated as the slope of IDWs obtained from L1b DDMs of RCS, divided by the average of DDMs of scattering area. Both the DDMs of RCS and the DDMs of scattering area are output from the L1b calibration as described in [Gleason, 2014].

5.2 13-Day Nature Run Dataset

The dataset used to develop and test the L2 wind speed retrieval algorithm is a very large set of simulated DDMs collected from all the CYGNSS observatories over a time frame of 13 days, during which a full life-cycle of a Tropical Cyclone has been simulated using the ECMWF/WRF Nature Run model [Nolan et al., 2013]. The Nature Run model generates physically realistic, long-lived storms with a very fine grid spacing (1 km reporting interval; ~4 km spatial resolution), a very high frequency output (1 output every 6 minutes, over a standard 13-day total run), and it outputs the complete set of ocean and atmosphere state variables. An example of four snapshots of the TC, spaced 6 hours apart, is shown in Figure 10. These are data grids of 480x480 km.

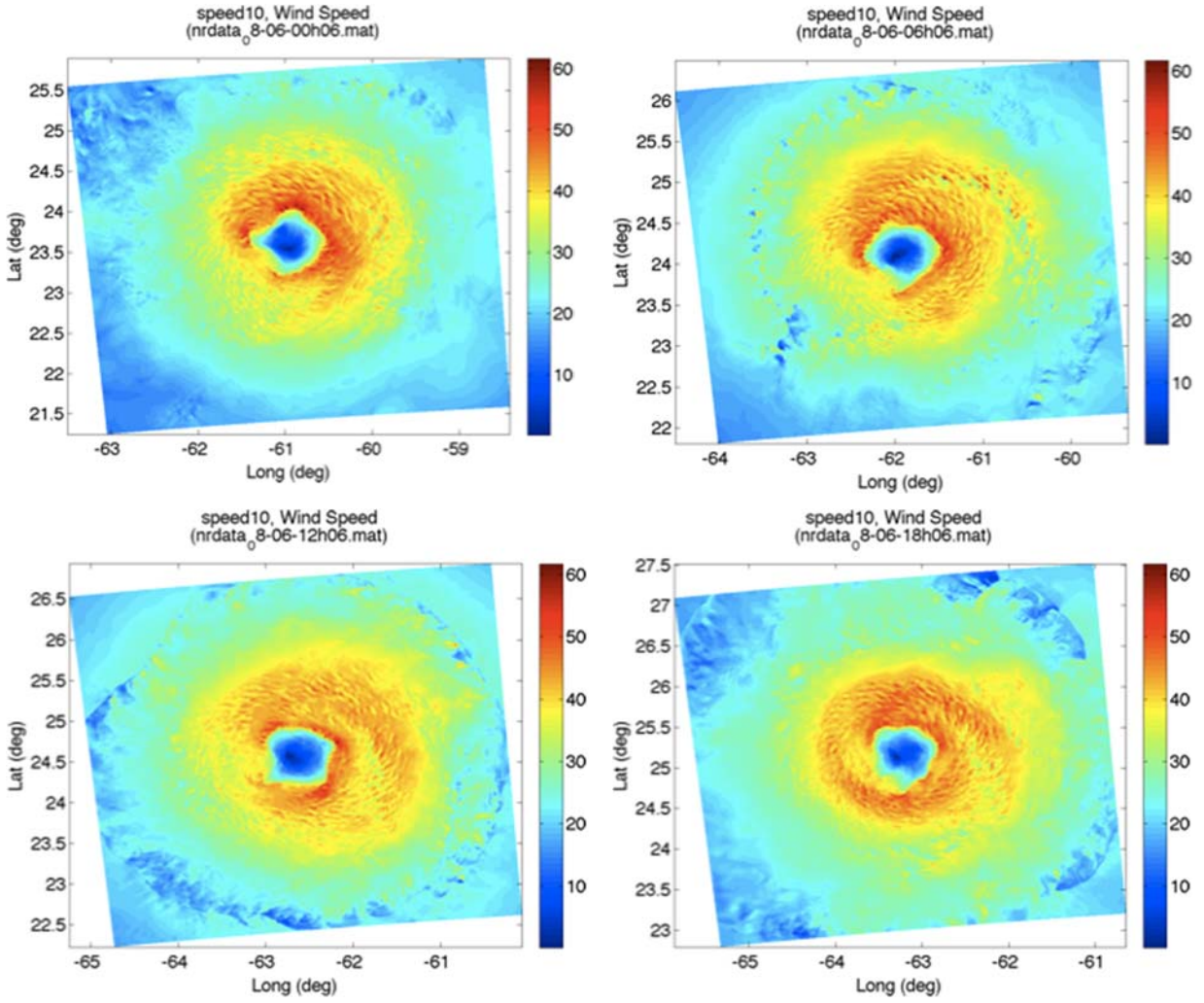


Figure 10. Snapshots of simulated realistic Tropical Cyclones from the Nature Run model.

Figure 11 shows the histogram of the wind speeds for the four snapshots illustrated in Figure 10.

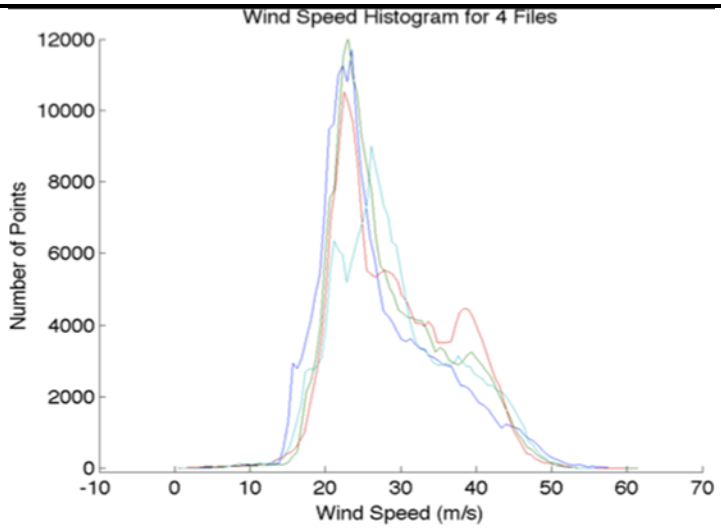


Figure 11. Wind speed histograms for the four TC snapshots.

The wind distribution is not uniform, and exhibits a very strong peak between 20 m/s and 30 m/s in all cases. This represents the wind distribution of a realistic TC. A plot of the locations of the inner core of the simulated TC, registered every 3 hours, is shown in Figure 12, along with plots of minimum surface pressure, and peak wind speed. The TC is simulated from July 29 2005 up until august 10 2005, and the maximum winds are registered in August 3 and 4. A plot of the 25 km x 25 km spatially averaged wind speeds from the full 13-day Nature Run is shown in Figure 13.

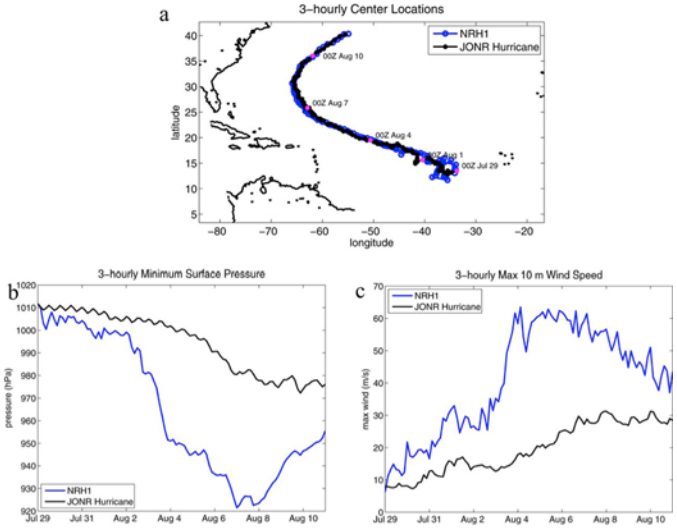


Figure 12. Illustration of 3-hourly center locations of the simulated TC (top), and plots of 3-hourly minimum surface pressure (bottom left) and maximum wind speed (bottom right) for the 13-day Nature Run simulation considered for this study. NRH1 are the values by the nature run numerical model, whereas JONR are the values estimated by the US Navy's Office of Naval Research's hurricane prediction center. The nature run shows higher winds and a lower pressure depression due to its higher spatial and temporal resolution.

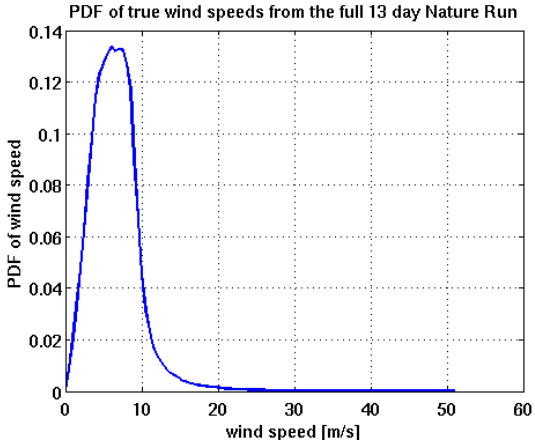


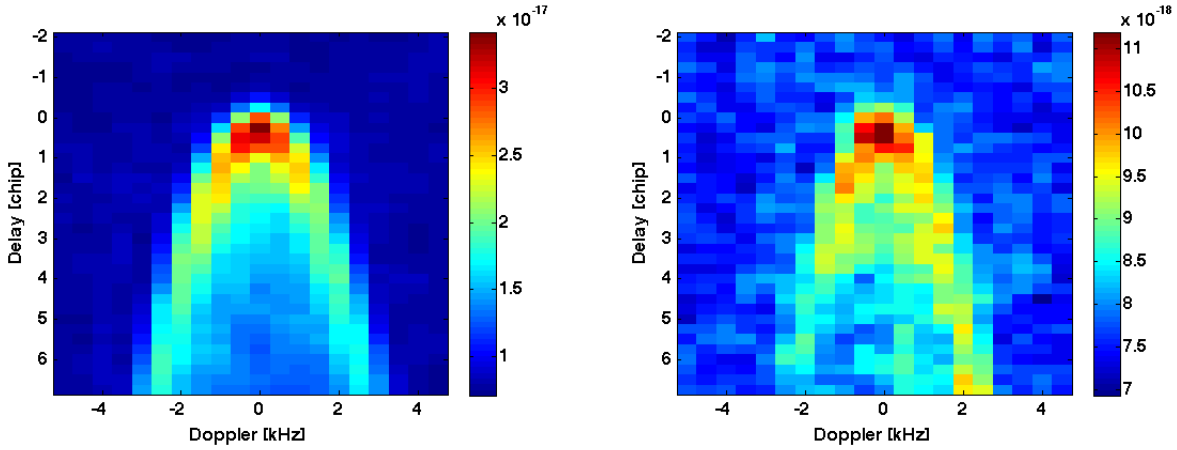
Figure 13. PDF of 25 km x 25 km spatially averaged wind speeds from the full 13-day Nature Run.

5.3 Generation of DDMs and L1b Observables

A large set of DDMs was obtained using a realistic 13-day TC scenario generated with the Nature Run (NR) model, and using realistic geometries for the 8 CYGNSS observatories, generated through orbital simulations.

A DDM was generated for each second of acquired data, and for those specular points acquired during an overpass of the satellites on the TC, resulting in more than 2 million DDMs. Each DDM is therefore characterized by different incidence angles and antenna gains at the specular point. The altitude of the transmitter and receiver for the four geometries is respectively ~20200 km and ~500 km, and the Right-Hand Circularly Polarized (RHCP) transmitter antenna gain is assumed to be constant and equal to 13 dB. Each DDM was also associated with a ground truth wind, computed as the 25 km x 25 km spatial average of the wind simulated from the NR.

DDMs are simulated using the CYGNSS End-to-End Simulator (E2ES). The E2ES simulates DDMs using an advanced implementation of the Zavorotny-Voronovich Model [Zavorotny and Voronovich, 2000], to which thermal noise and speckle noise are added. Furthermore, the simulated bistatic Normalized Radar Cross Section (NRCS) accounts for the non-uniformity of the wind field across the 100 km x 100 km spatial area that is used to generate the DDM. A detailed description of the E2ES can be found in [Clarizia et al., 2013d]. Two examples of simulated L1a DDMs, for low wind speed (5 m/s) and for high wind speed (40 m/s) from the CYGNSS E2ES are shown in Figure 14.



(a) **(b)**

Figure 14. DDM simulated with the E2ES, using geometry 2, baseline configuration, and a wind speed of 5 m/s. (b) DDM simulated using geometry 4, baseline configuration, and a wind speed of 40 m/s.

The delay and Doppler resolution for the simulated DDMs is respectively 0.25 chips and 500 Hz. The delay axis extends from -2 chips to 6.75 chips, whereas the Doppler axis spans -5 kHz to 4.5 kHz.

L1a DDMs are L1b calibrated, and a DDMA and LES observable is then computed for each of them. A very important aspect for this computation 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 15 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.

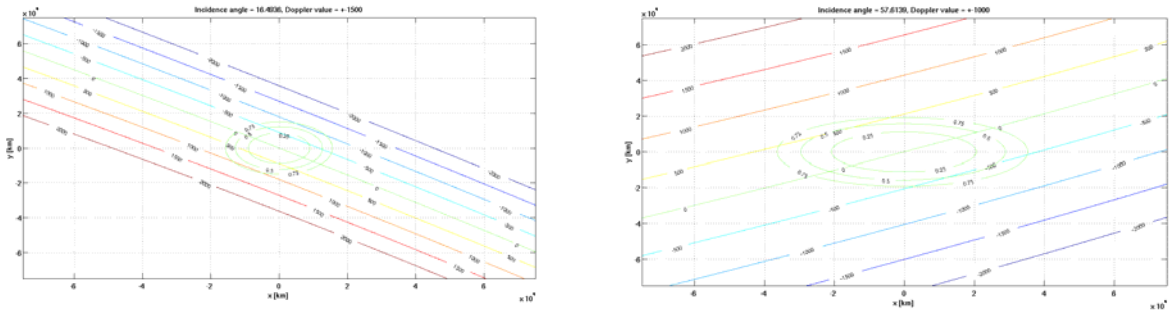


Figure 15. Iso-delay and Iso-Doppler lines for an incidence angle of 16.5° (left) and of 57.6° (right).

Figure 16 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 14 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.

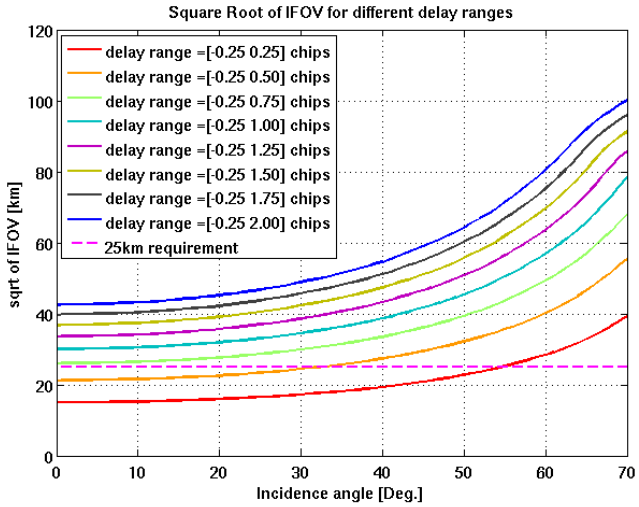


Figure 16. illustration of curves of square root of IFOV versus incidence angle, for a variety of delay ranges. The dashed magenta line shows the 25 km requirement.

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 15), 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 17, 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.

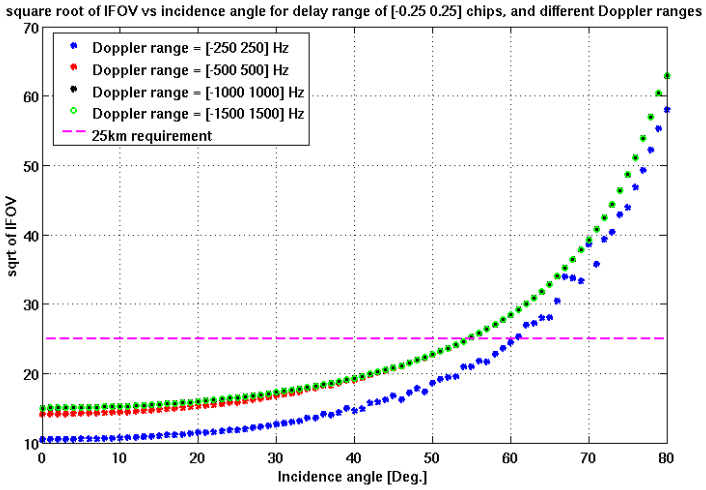


Figure 17. illustration of curves of square root of IFOV versus incidence angle, for a fixed delay range of [-0.25 0.25] chips, and different Doppler ranges. The dashed magenta line shows the 25 km requirement.

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 (Figure 18).

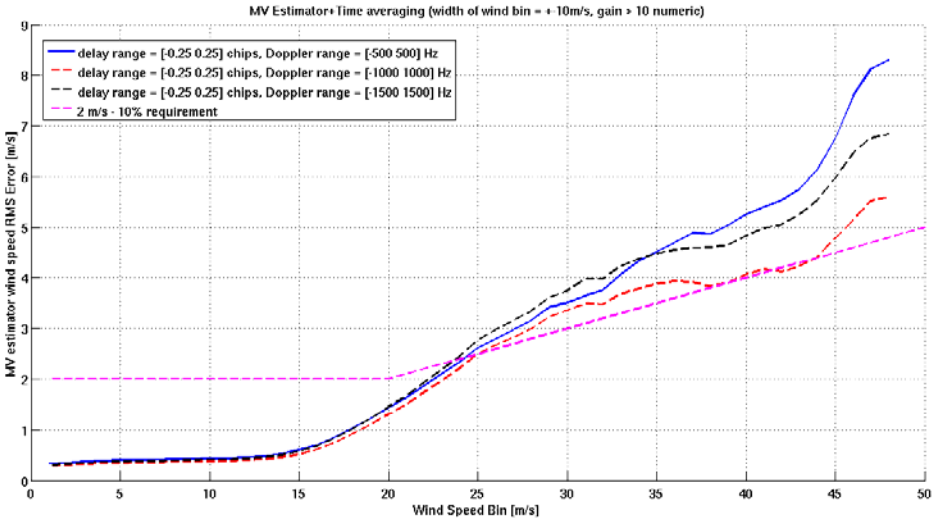


Figure 18. curves of RMS error as a function of wind speed, obtained for a fixed delay range, and three different Doppler ranges, using only samples with receiver antenna gain higher than a numerical value of 10. The lowest RMS error is provided by the Doppler range of [-1000 1000] Hz.

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. The DDMA and LES are therefore computed using a delay range of [-0.25 0.25] chips, and a Doppler range of [-1000 1000] Hz, and they are referred to as L1a observables, since they represent the very first stage of the retrieval algorithm, and they are not corrected for any antenna or geometrical affects. A plot of the L1b DDMA and LES values, against the ground-truth wind speeds, is shown in Figure 19. Here the noisiest observables have been filtered out, by selecting only those with a high enough Range Corrected Gain, or RCG. This is defined as the receiver antenna gain, multiplied by the range losses, at the specular point, as follows:

$$RCG = \frac{G_{RX}^{SP}}{\left(R_0^{SP} R^{SP}\right)^2} \tag{5.1}$$

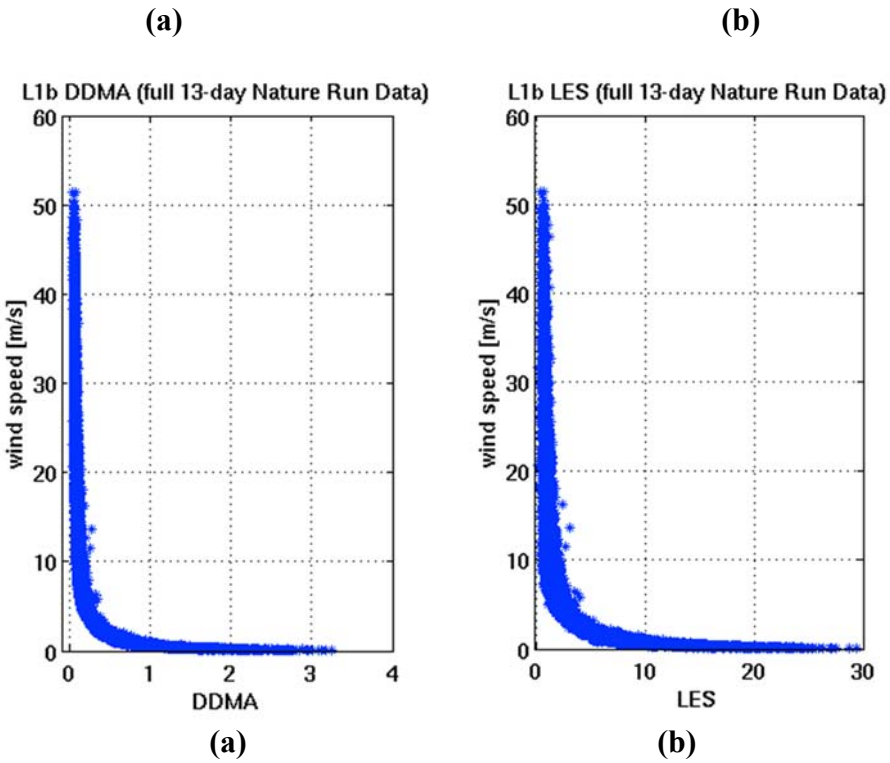


Figure 19. Illustration of ground-truth wind speeds vs L1b DDMA (a) and LES (b), for RCG > 20.

The RCG represents an improved definition of gain, which takes into account both the effect of the receiver antenna and the effect of the attenuation due to range losses, and will be used throughout the whole study to filter data with low, medium or high gain, and analyze different results for different data types. Typical values of low, medium and high RCG are respectively $3*10^{-27}$, $10*10^{-27}$ and $20*10^{-27}$. For brevity, we will often refer to RCG values of 3, 10 and 20, omitting the 10^{-27} term. The observables in Figure 19 have been selected as those with a RCG higher than 20, thus they represent data with very low noise.

6. Specular Point Selection Algorithm

After having applied the L1b corrections to the observables, and before applying the wind speed retrieval algorithm, a Specular Point Selection Algorithm (SPSA) has been applied to the data, to filter out those with lower RCG. Each CYGNSS observatory is capable of tracking up to four simultaneous specular points, and the onboard selection is based on the highest RCG values, so that only those four SPs with highest RCG value are retained, and the others are discarded. The CYGNSS E2ES generates data for all the specular points that are within the receiver antenna footprint, so a SPSA needs to be applied to reproduce realistic on-orbit conditions. Figure 20 shows a random selection of cases with multiple specular points, where the value on the y-axis is the

RCG value for the specular point. It highlights in red those retained after the SPSA is applied to the data (note that all the SPs are retained if they are less than four).

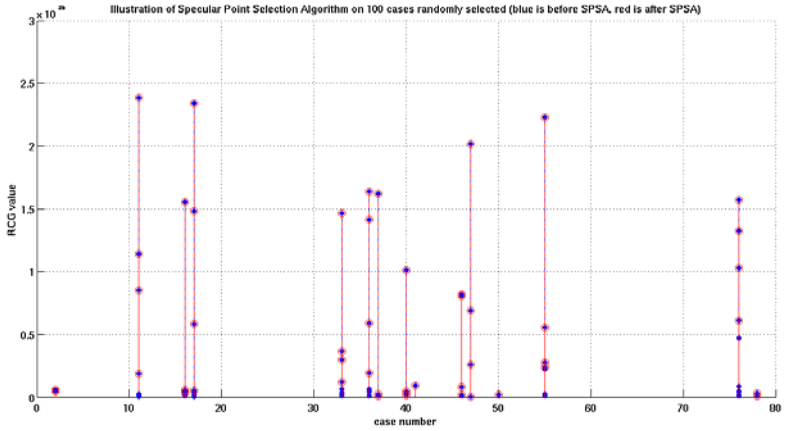


Figure 20. Illustration of specular point selection algorithm, as RCG value versus the acquisition number, where the acquisition refers to a single acquisition by a given CYGNSS satellite for a given second. The blue points represent the totality of specular points within the antenna pattern of a given CYGNSS observatory, and the red circles represent those SPs that survive after the SPSA is applied.

7. Overview of L2 Wind Speed Retrieval Algorithm

The overall approach of the wind speed retrieval algorithm is similar to the one presented in [Clarizia et al., 2013c], with some additional steps in the data processing that improve the final performance of the retrieval algorithm itself. The retrieval algorithm follows a statistical inversion approach that uses an empirical GMF derived in this case from the true wind matchups against their respective DDM observable, and it is applied individually to each observable [Clarizia et al., 2013c]. However, compared to [Clarizia et al., 2013c], the GMF here is not only a function of wind speed, but a function of both wind speed and incidence angle. This is due to the dependence existing between the L1b observables and the incidence angle, which has been observed in L1b observables simulated without thermal noise and speckle, and which can be found in the equation of the Normalized Radar Cross Section (NRCS) [see Zavorotny and Voronovich, 2000]. Two approaches exist to deal with such dependence: one is to remove the incidence angle dependence from the L1b observables (the so-called L2a correction), and another is to generate the GMF function such that it accounts for the incidence angle dependence too, other than the wind speed. While both approaches work quite well with simulated data, the second approach is to be preferred, since it is consistent with scatterometry where GMFs are functions of wind speed and incidence angle, and since it is expected that GMFs are in general function of a number of parameters (including wind direction, azimuth angle, dielectric properties of the sea etc.), with a dominant dependence on wind speed and incidence angle. For this reason, it is always desirable to model and incorporate these dependencies in the GMF, rather than trying to remove them from the observables.

The steps taken to develop each retrieval algorithm are as follows:

1) All available DDM observables and the corresponding collocated ground truth winds and incidence angle are assembled;
2) An empirical 2D GMF, function of both wind speed and incidence angle, is constructed from the matchups using least squares regression analysis;
3) The GMF is used as the basis for the mapping from DDM observable to estimated wind speed;
4) The Root Mean Square (RMS) difference between the estimated wind and the NDBC ground truth wind speed is computed as the RMS error in the retrieval.

As already mentioned, this type of wind retrieval algorithm is often used in scatterometry and SAR, where empirical GMFs are derived from a large collocation study between observed measurements and in situ buoy and/or NWP model data. Similarly, CYGNSS will make use of a large dataset of collocated wind speed information from different sources (including buoys, model outputs, aircraft measurements and satellite cross-overs) that will help construct a robust empirical GMF model for each observable. In order to work with independent datasets for the inversion algorithm training and for the performance evaluation of the algorithm itself, we split the full 13-day Nature Run data into two subsets: one including all the data with an odd minute in their time stamp, and another with an even minute in their time stamp. The choice of separating the datasets based on their odd and even minutes is commonly adopted in these cases when two independent datasets need to be created. The odd minute dataset is used to train the Look-Up Table (LUT)

which is used to estimate the winds, and the even minute dataset is used to assess the performances of the algorithm, in terms of RMS error.

8. Generation of Geophysical Model Function

For each observable, the training data are formed as the L1b observable, computed over a delay range of [-0.25 0.25] chips and a Doppler range of [-1000 1000] Hz, to which the SPSA has been applied, and with odd minutes in their time stamp. Then, only the "clean" data with high enough RCG are selected for LUT generation, and in our case we have used a RCG threshold of 20. The derivation of a full 2D GMF, function of both wind speed and incidence angle, was in this case made difficult by the lack of enough available samples, particularly at certain ranges of incidence angles. An alternative approach was used here, which is described in the following steps:

1) The dependence of the observables on the incidence angle is pre-estimated, using observables computed from noise-free DDMs (the so-called L2a correction)
2) A unique GMF is derived for all the samples, to map the selected observables against their ground truth wind speed, which is true wind speed from the nature run, spatially averaged over an area of 25 km x25 km, centered at the specular point.
3) A number of different, small enough incidence angle intervals are considered, and the GMF obtained in step 2 is scaled according to the L2a correction developed in step 1, to obtain one GMF for each incidence angle interval. This represents our final 2D GMF, function of both wind speed and incidence angle.

This method, which represents a simplified but approximate way to derive a full 2D GMF, will then be replaced by the more appropriate approach of deriving the 2D GMF entirely from the samples, once a high enough number of simulated storms will be made available. The following subsections describe in detail each of the steps listed above.

8.1 L2a Correction

The L2a correction represents the correction for the dependence on the incidence angle θ_i, which is not eliminated by the L1b corrections. After the L1b corrections, a dependence on θ_i of the DDM of scattered power remains in the equation for the radar cross section, $\sigma_0(\boldsymbol{r})$ which according to the model by [Zavorotny and Voronovich, 2000], can be written as:

$$\sigma_0(\boldsymbol{r},\theta_i)=\frac{\pi\left|\Re(\theta_i)\right|^2 q^4(\boldsymbol{r},\theta_i)}{q_z^4(\boldsymbol{r},\theta_i)}f_{\boldsymbol{q}}(\boldsymbol{s}) \tag{8.1}$$

and it depends on the Fresnel reflection coefficient $\Re$, and on the Probability Density Function (PDF) $f_{\boldsymbol{q}}(...)$ of the sea surface slopes, the latter defined as $\boldsymbol{s}=-\boldsymbol{q}_\perp/q_z=\left[-q_x/q_z,-q_y/q_z\right]$. The vector $\boldsymbol{q}=\left[q_x,\ q_y,\ q_z\right]$ is known as the scattering vector, namely the bisector of the angle formed by the transmitter-point on surface and receiver-point on surface ranges. Assuming a Gaussian PDF for the sea surface slopes, it can be expressed as follows:

$$f_{\boldsymbol{q}}\left[\frac{-\boldsymbol{q}_{\perp}(\boldsymbol{r},\theta_i)}{q_z(\boldsymbol{r},\theta_i)}\right]=\frac{1}{2\pi\sqrt{\sigma_x^2\sigma_y^2}}\exp\left[-\frac{s_x^2}{2\sigma_x^2}-\frac{s_y^2}{2\sigma_y^2}\right] \tag{8.2}$$

where σ_x^2 and σ_y^2 are the mean square slopes along the *x* and *y* direction (for simplicity, here we are not accounting for any effect of the wave direction in the PDF). If we consider all the terms in equation (8.1) to be approximately constant within the scattering area contributing to the observables, and equal to the value at the specular point, then the x- and y- component of the scattering vector are null, and equation (8.1) becomes:

$$\sigma_0(\boldsymbol{r},\theta_i)=\frac{\pi\left|\Re(\theta_i)\right|^2 q_z^4(\boldsymbol{r},\theta_i)}{2\pi\sqrt{\sigma_x^2\sigma_y^2}\,q_z^4(\boldsymbol{r},\theta_i)}=\frac{\left|\Re(\theta_i)\right|^2}{2\sqrt{\sigma_x^2\sigma_y^2}} \tag{8.3}$$

in other words, the dependence on the incidence angle remains only contained in the Fresnel Reflection coefficient (equations 36-39 in [Zavorotny and Voronovich, 2000]), and thus it would be sufficient to correct for $\Re$ to eliminate such dependence. A plot of $\sigma_0(\boldsymbol{r})$ at the specular point versus incidence angle is shown in Figure 21.

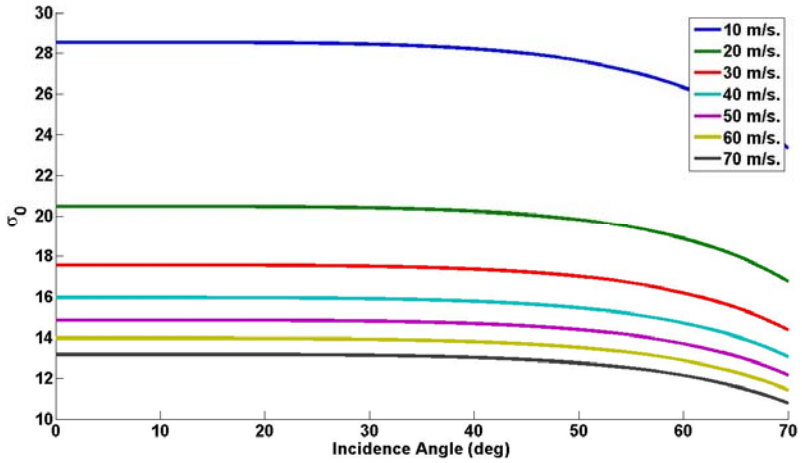


Figure 21. curves of $\sigma_0(\boldsymbol{r}_{SP})$ versus incidence angle, for different wind speed values.

Unfortunately, compensating for the effect of $\Re$ is not sufficient to fully eliminate the dependence of $\sigma_0(\boldsymbol{r})$ on θ_i. This is shown in Figure 22, where a set of DDMAs from noise-free simulated DDMs has been generated for a number of different realistic CYGNSS geometries, corresponding to different incidence angles. Both L1b DDMA and L2a DDMA are illustrated in Figure 22 versus the incidence angle, where L2a DDMA have been obtained by simply dividing the L1b DDMA value by the squared module of $\Re$ for the specific incidence angle.

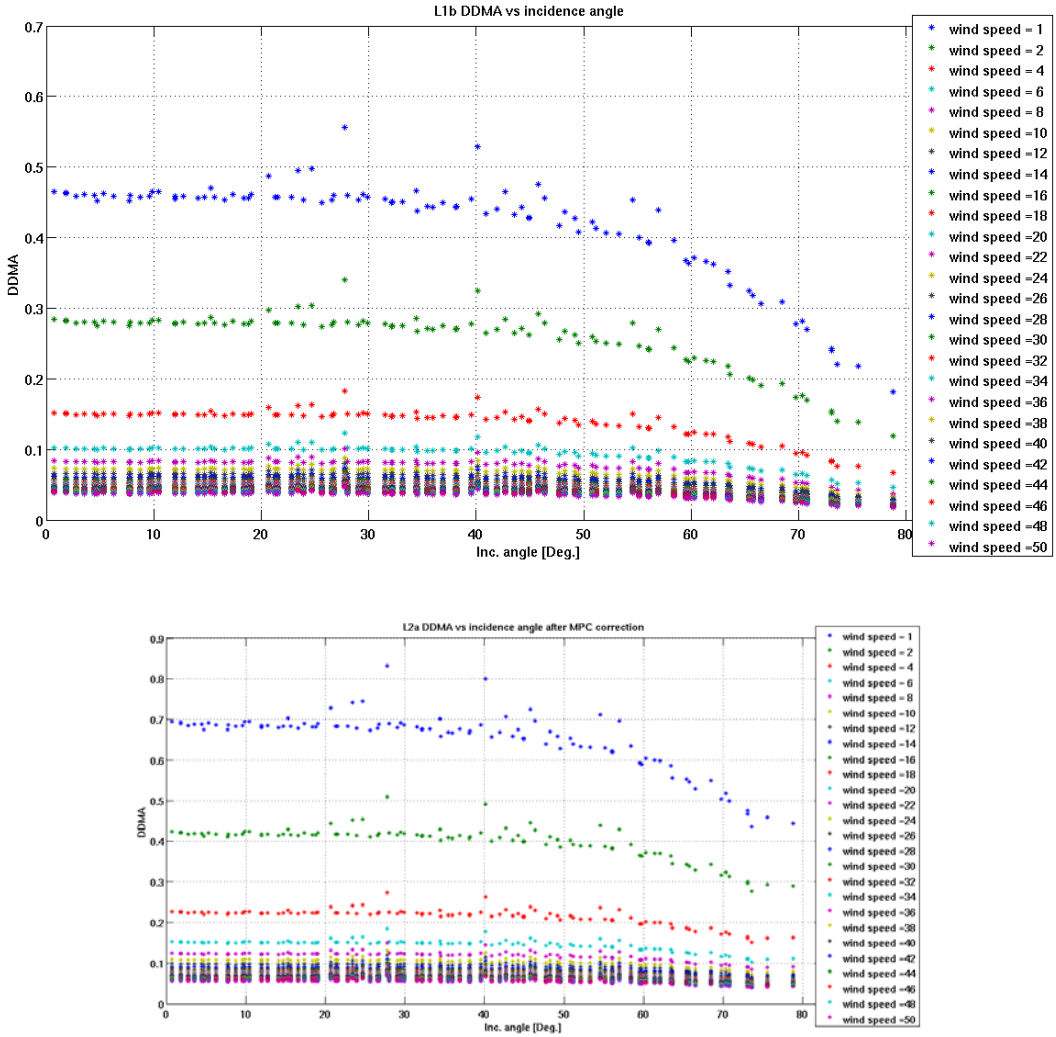


Figure 22. plot of noise-free L1b (top) and L2a (bottom) DDMA, versus incidence angle, obtained from noise-free DDMs simulated using the E2ES. The different colors refer to different wind speeds. In this case, the L2a correction has been implemented by simply compensating for the effect of the Fresnel Reflection coefficients (i.e. by considering the curves in Figure 21).

Figure 22 shows that a dependence on incidence angle remains after having removed the dependence on $\Re$, suggesting that the assumption of all the terms in equation (8.1) being constant and equal to their values at the specular point is not accurate enough. From the L1b DDMA in Figure 22, we have therefore developed for now an empirical correction, while we are investigating ways to come up with an analytical or semi-analytical correction for the incidence angle dependence. Having verified that the dependence of L1b DDMA on incidence angle does not vary with wind speed and having normalized each of the curves by their approximate value at θ_i=0°, we have selected a given curve corresponding to a given wind value, and found the coefficients *a*, *b* and *c* of a power function in the form of $y(\theta) = a\theta^b + c$, that best fits the data points of that curve. The normalized curves and the best-fit power curve are shown in Figure 23.

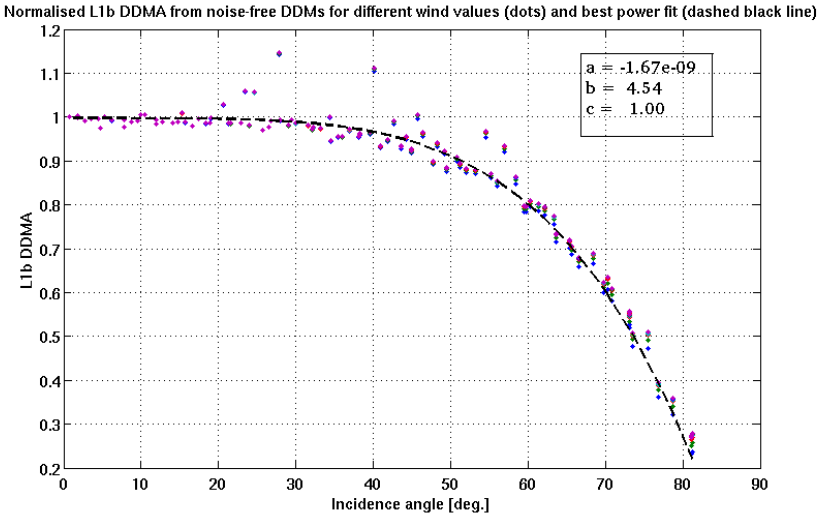


Figure 23. Normalized DDMA vs incidence angle curves, with superimposed best-fit power function (dashed black line)

This function has then been used to model the dependence of the DDMA on incidence angle. The same procedure has been applied to LES data, and the final coefficients for the power fit have been found to be the same as for the DDMA. The final expression to derive the L2a observables is therefore:

$$OBS(\theta)_{L2a} = \frac{OBS(\theta)_{L1b}}{y(\theta)} \tag{8.4}$$

Where OBS stands for the observable, and *y(θ)* is the power fit shown in Figure 23. Figure 24 shows the L2a DDMA, where the L2a correction has been now implemented through the use of the power fit mentioned above, and where the dependence on the incidence angle has been eliminated.

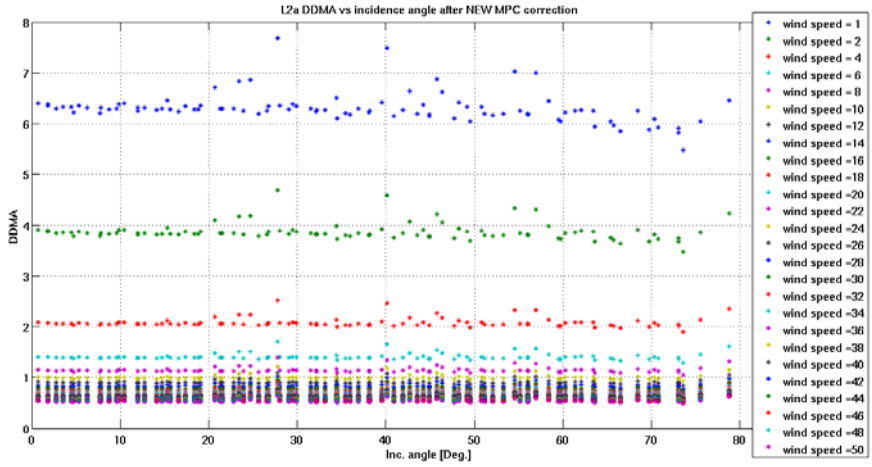


Figure 24. plot of noise-free L2a (bottom) DDMA, versus incidence angle, obtained from noise-free DDMs simulated using the E2ES. The different colors refer to different wind speeds. The L2a correction has been developed empirically using the curves in Figure 22 and Figure 23.

8.2 Derivation of GMF from all samples

A GMF is derived, to map the selected observables against their ground truth wind speed, which is true wind speed from the nature run, spatially averaged over an area of 25 km x25 km, centered at the specular point. The GMF is computed in the form of a Look-Up Table (LUT) of DDMA values corresponding to the 25 km x25 km spatially averaged wind values, and must be a monotonic smooth enough curve 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. A smoothing filter is finally applied to the LUT, to make sure it is a monotonic function.

The training data and LUT for both DDMA and LES samples are shown in Figure 25.

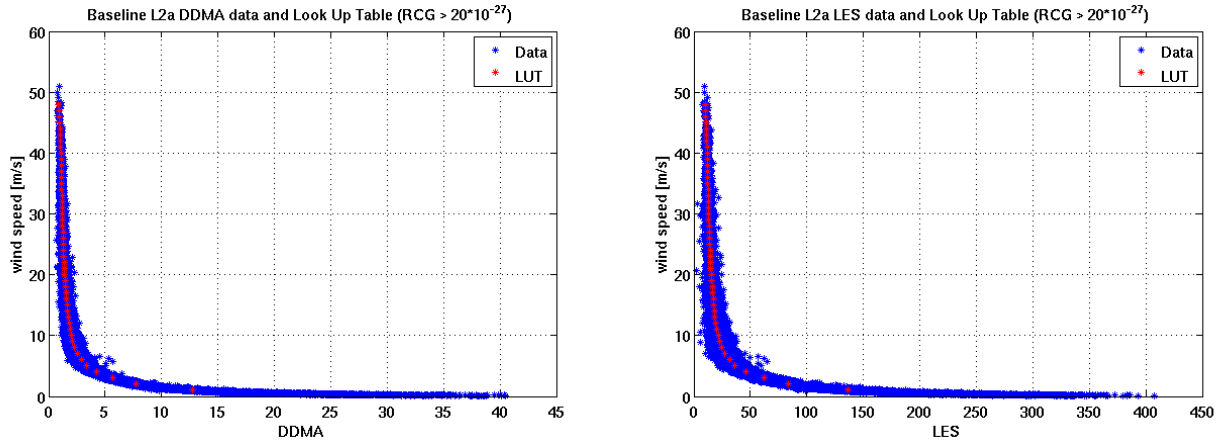


Figure 25. ground truth wind speeds versus L2a DDMA (a) and L2a LES (b) training data, shown as blue points, and selected with RCG>20. The Look Up Table derived from the data is shown as red points

8.3 GMF Scaling using L2a Correction

The dependence of the GMF on incidence angle is explicitly accounted for in the retrieval algorithm by adding a second dimension to the LUT that is indexed by incidence angle. We use the L2a correction to scale the GMF obtained in 8.2 to generate a different GMF for each incidence angle interval. We divide the total incidence angle range (from 0.1° to 80°) into 800 intervals of 0.1° each, and we derive a GMF for each interval by applying inversely the L2a correction as follows:

$$GMF(\theta) = GMF^{*} \cdot y(\theta) \tag{8.5}$$

where GMF^{*} is the total GMF derived in 8.2, and $y(\theta)$ is the L2a correction. The $GMF(\theta)$ resulting from equation (8.5) is shown for 6 incidence angle intervals together with the samples belonging to that interval in Figure 26. The 6 incidence angle intervals span the range of 0° to 55°, since this is the range that survives after the so-called Effective Field Of View (EFOV) filter (explained later on) is applied. The agreement between samples (blue) and GMFs (red) is good in all cases, expect for the last range of incidence angles (bottom right of Figure 26). This is only due to the fact that the strongest changes in the sample distribution with respect to incidence angle occur at the highest incidence angle values, hence at these incidence angles a GMF obtained from only the central 0.1° incidence angle interval is no longer well representative of all the samples within a 5° incidence angle interval. Note however that the wind retrieval process bins the samples into 800 different intervals (not just 6) and makes use of all the 800 GMFs.

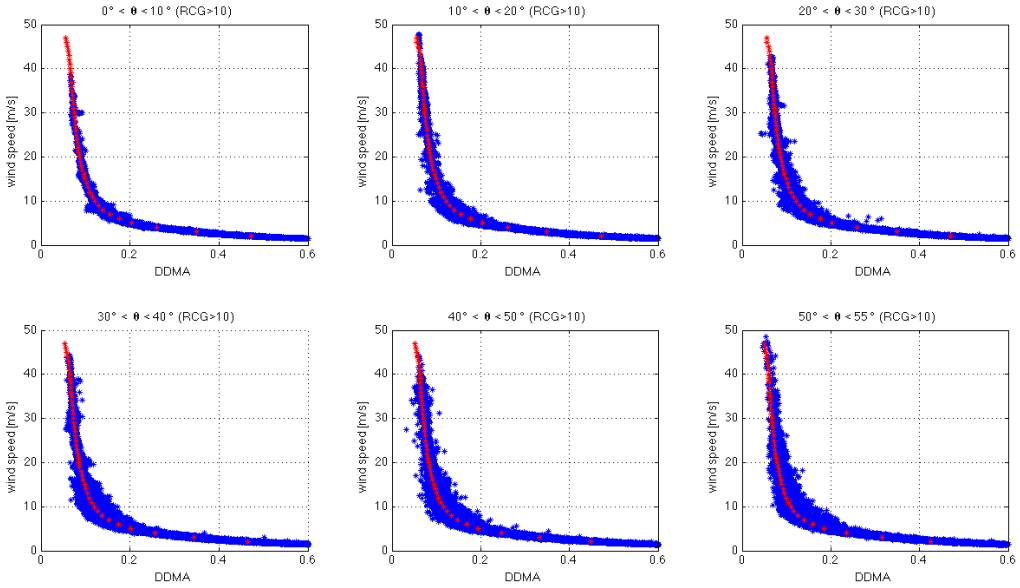


Figure 26. Scatterplot of DDMA samples versus wind speed (blue), for six different incidence angle intervals (left to right, top to bottom): [0° - 10°], [10* - 20°], [20° - 30°],[30° - 40°],[40° - 50°] and [50° - 55°]. The red stars show the GMF corresponding to the 0.1° incidence angle interval at the center of the interval considered (e.g. the GMF for the [4.95° - 5.05°] interval is shown for the samples belonging to the [0° - 10°] interval, and so on).

8.4 Wind Estimation using the LUT function

The winds are now estimated using the performance evaluation dataset, namely the subset of data from the 13-day Nature Run with even minutes in their time stamp. For each point of the performance evaluation dataset (either DDMA or LES), a wind speed is estimated using the LUT function for the incidence angle of the sample point. 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):

$$\hat{U}_{10}^{DATA} = U_{10}^{LUT} + \alpha\left(DDMA^{DATA} - DDMA\left(U_{10}^{LUT}\right)\right)$$
$$\alpha = \left(V_{10}^{LUT} - U_{10}^{LUT}\right) \Big/ \left(DDMA\left(V_{10}^{LUT}\right) - DDMA\left(U_{10}^{LUT}\right)\right) \quad (8.6)$$

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

$$\hat{U}_{10}^{EXT} = U_{10}^{i\ LUT} + \alpha^{EXT}\left(DDMA^{DATA} - DDMA\left(U_{10}^{iLUT}\right)\right) \quad i = 1, n \tag{8.7}$$

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 though linear fitting of the first 2 points (if *i=1*) or last 2 points (if *i=n*) of the LUT.

8.5 De-biasing of wind retrievals

At this point, we have for each observable a true wind and an estimated wind obtained following the method described the previous paragraph. One extra task to be accomplished is to check and remove any potential biases in the estimated winds. This is done by re-binning again the true winds with a given bin width (in our case we have chosen the bin width to be 3 m/s), and for each bin calculate the bias as follows:

$$bias\left(U_{10(c)}^{\ i}\right) = \frac{1}{N_k^i}\sum_{k=1}^{N_k^i}\left[\hat{u}^k - u^k\right] \tag{8.8}$$

where the hat indicates the estimated quantity (as opposed to the true quantity), $\hat{U}_{10(c)}^{i}$ represents the center of the i-th wind speed bin, and *k* is the index of all the samples falling within the i-th bin (from 1 to N_k^i). Once the bias is calculated for each bin, it is then removed (i.e. subtracted) from all the wind estimates that fall within that bin. The bias between estimations and ground truth turned out to be quite small so de-biasing is a useful operation but not crucial.

8.6 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]:

$$u_{MV} = \boldsymbol{m} \cdot \boldsymbol{u} \tag{8.9}$$

where $\boldsymbol{u}$ is the vector of individual estimates and $\boldsymbol{m}$ is the vector of coefficients. They 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 derivation of the coefficients for the estimator is illustrated in [Clarizia et al., 2013]. These coefficients are given by

$$\boldsymbol{m} = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i.j}^{-1} \right)^{-1} \boldsymbol{C}^{-1} \boldsymbol{1} \tag{8.10}$$

where $\boldsymbol{1}$ is a vector of ones, $\boldsymbol{C}^{-1}$ 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

$$\sigma_{MV}^2 = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} c_{i.j}^{-1} \right)^{-1} \tag{8.11}$$

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

$$\boldsymbol{C} = \boldsymbol{SRS} \tag{8.12}$$

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. 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 is estimated for different ranges of RCG. This happens because the correlation between retrievals from the two observables decreases for noisier data characterized by lower RCG and consequently lower SNR. This will then allow 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 covariance and correlation coefficient matrices for five different RCG intervals. It can be noticed that the correlation coefficients between DDMA and LES retrievals are the lowest for the lowest RCG range, and as expected they gradually increase as the RCG increases.

$C_1 = \begin{bmatrix} 108.1522 & 17.2250 \\ 17.2250 & 207.8095 \end{bmatrix}$	$R_1 = \begin{bmatrix} 1.0000 & 0.1149 \\ 0.1149 & 1.0000 \end{bmatrix}$	$3*10^{\wedge-27}$ <= RCG < $5*10^{\wedge-27}$
$C_1 = \begin{bmatrix} 37.5061 & 6.9362 \\ 6.9362 & 95.0033 \end{bmatrix}$	$R_1 = \begin{bmatrix} 1.0000 & 0.1162 \\ 0.1162 & 1.0000 \end{bmatrix}$	$5*10^{\wedge-27}$ <= RCG < $10*10^{\wedge-27}$
$C_1 = \begin{bmatrix} 5.9796 & 1.1326 \\ 1.1326 & 15.7289 \end{bmatrix}$	$R_1 = \begin{bmatrix} 1.0000 & 0.1168 \\ 0.1168 & 1.0000 \end{bmatrix}$	$10*10^{\wedge-27}$ <= RCG < $20*10^{\wedge-27}$
$C_1 = \begin{bmatrix} 0.7807 & 0.3144 \\ 0.3144 & 1.4212 \end{bmatrix}$	$R_1 = \begin{bmatrix} 1.0000 & 0.2985 \\ 0.2985 & 1.0000 \end{bmatrix}$	RCG >= $20*10^{\wedge-27}$

Table 1. Covariance matrices (C) and Correlation coefficient matrices (R) between winds retrieved from DDMA and from LES, for four different RCG intervals.

A scatter plot of true vs estimated MV winds for RCG>10 is shown in Figure 27. The MV estimated winds match quite well the true ones.

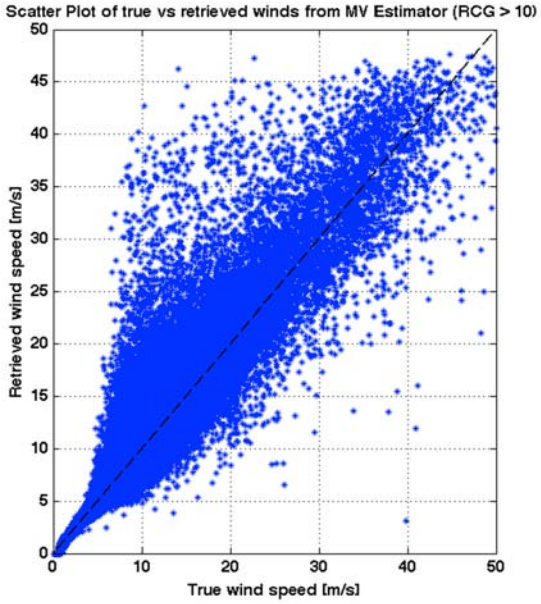


Figure 27. Scatter plot of retrieved wind speeds using the MV estimator, versus ground truth wind speeds, for those samples with RCG>10. These have been computed from the performance evaluation dataset.

8.7 *Performance Figure-of-Merit*

The performances of the MV estimator are compared to the DDMA and LES alone in a more quantitative way, in the form of wind speed Root Mean Square (RMS) error curves as a function of the wind speed bin. The RMS for the i-th wind speed bin is calculated as:

$$RMS\left(U_{10(c)}^{\ i}\right)=\sqrt{\frac{1}{N_k^i}\sum_{k=1}^{N_k^i}\left[\hat{u}_{(ub)}^k-u^k\right]^2} \tag{8.13}$$

where $\hat{u}_{(ub)}^k$ is the estimated unbiased wind speed. RMS error curves have been calculated over a wind speed range of 1 m/s to 48 m/s, considering a bin width of +- 10 m/s, and with a 1 m/s bin step. A plot of the RMS error values versus the central wind value of the wind speed bin is shown for the three cases in Figure 28. Here only the retrievals from samples with RCG>10 have been retained for the RMS error calculation. For convenience, the CYGNSS baseline 2 m/s / 10% requirement is also shown as a magenta dashed line. The significant improvement in the retrieval algorithm performances due to the MV estimator is clearly visible.

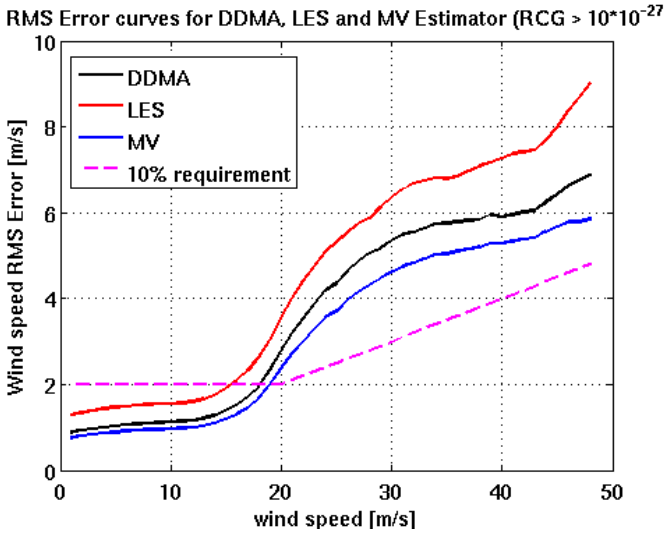


Figure 28. Root Mean Square (RMS) Error between true and retrieved wind speed, as a function of the true wind speed, for retrievals from DDMA, LES and MV estimator. The CYGNSS baseline 10% requirement is also shown for comparison.

8.8 *Time Averaging*

An extra processing step that contributes to improve the performances by reducing the RMS error, is to apply time averaging to the collected data. Looking again at Figure 17, it can be seen that for a range of incidence angles, 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 that can be averaged can be easily calculated, using some simplifications. Since the specular point is moving on the surface of about 6 km/s, an "Effective" Field of View (EFOV) can be defined for each sample as follows:

$$EFOV = \sqrt{\sqrt{IFOV} \cdot \left[\sqrt{IFOV} + (n-1) \cdot l\right]} \qquad l \approx 6\ km \tag{8.14}$$

where IFOV is the instantaneous FOV of the sample, l indicates the specular point displacement on the surface per second (6 km), and n is the total number of samples to average. Imposing an EFOV equal to 25 km, and solving for n, we obtain:

$$n = \frac{EFOV^2}{6\sqrt{IFOV}} - \frac{\sqrt{IFOV}}{6} + 1 \tag{8.15}$$

A plot of the values of n versus incidence angle, for two different delay ranges, and using an EFOV limit of 25 km, is shown in Figure 29. The value n has been rounded to the lowest integer value, in order to make sure that the 25km requirement is never violated.

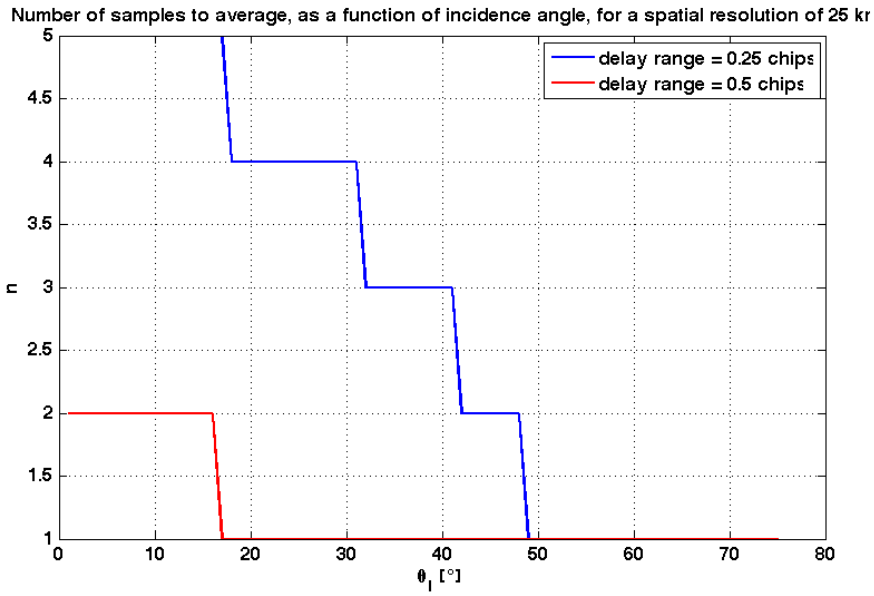


Figure 29. number of samples to average, as a function of the incidence angle θi, for two different delay ranges. The blue curve, corresponding to a delay range of [0 0.25] chips, is the one of interest.

This function can be then used to know, for each incidence angle, the number of samples that can be averaged, if a 25km requirement on the spatial resolution must be met. 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 (8.15), 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. Figure 17 also highlights that in principle two different approaches to time averaging are possible, and these are:

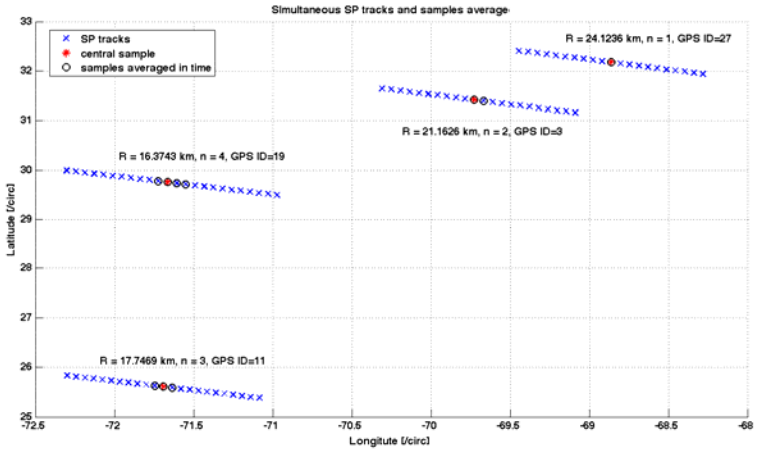
1. Average of the values of the observables
2. Average of the retrieved winds

Furthermore, it is possible to choose one of the following:

a. Use of shorter delay ranges (noisier observables) with a longer time averaging
b. Use of longer delay ranges (less noisy observables) with a shorter time averaging

Each of the four combinations has been tested, and the best performances have been obtained using 1) and using a. The choice of 1) rather than 2) can be explained by the change in slope of the LUTs shown in Figure 25. The slope of the LUTs changes more quickly as the wind increases, and therefore at high winds a small error in the observable would translate into a big error in the retrievals. Hence, averaging the observables helps reduce the noise in the observable itself, and translates into more accurate wind estimates.

The choice of a. instead of b. simply indicates that increasing the time averaging mitigates the noise in the observables more than computing the observables over a wider delay window. 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.



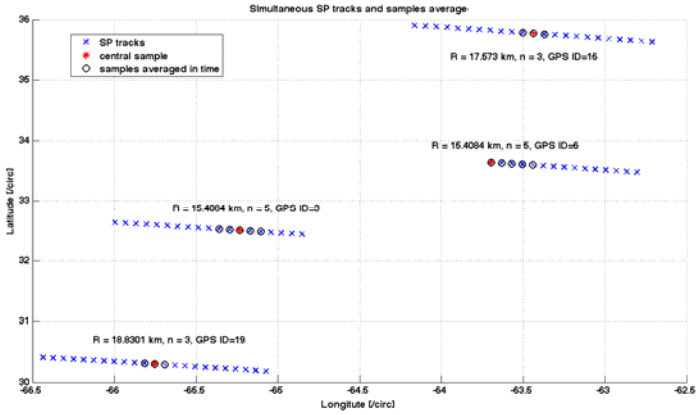


Figure 30. Two graphical illustrations of how the Time Averaging (TA) algorithm works, for four simultaneous SP tracks acquired by a single CYGNSS observatory.

The scatter plot of true vs retrieved winds using the MV estimator, before and after time averaging of the observables, is shown in Figure 31. Note that the LUTs used to retrieve winds from time-averaged DDMA and LES observables have not been re-computed, and are the same as those shown in Figure 25. A plot of the RMS error before and after time averaging is shown in Figure 32, which highlights the further improvement in the algorithm performances when time averaging is used.

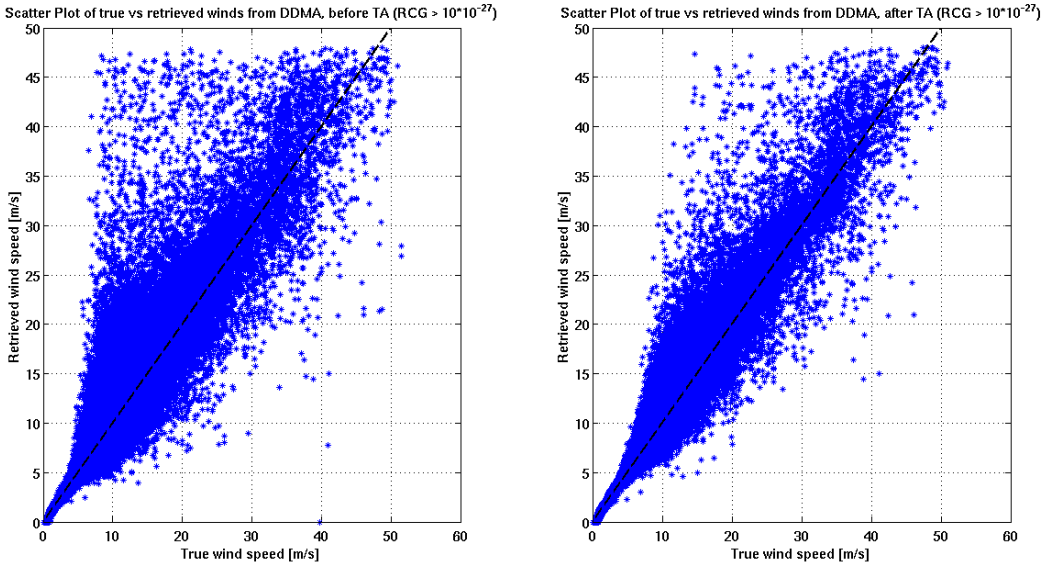


Figure 31. scatter plot of MV estimated vs true winds before (a) and after (b) time averaging, for samples with RCG>10.

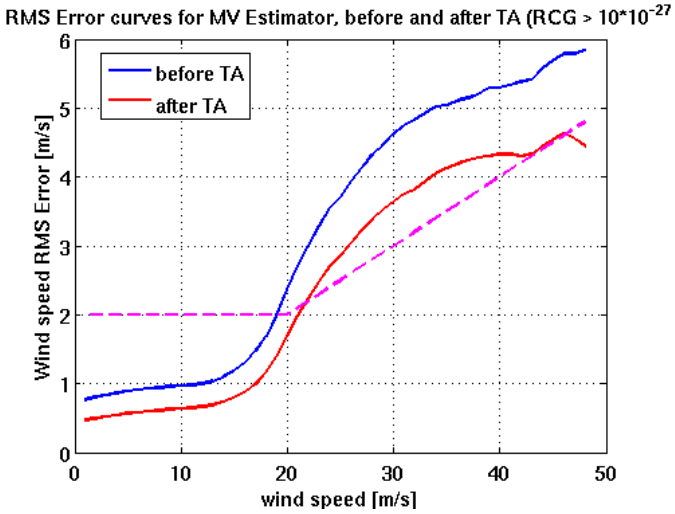


Figure 32. RMS error curves before and after TA, for RCG > 10.

8.9 IFOV Filter

A final processing step needed in the retrieval algorithm is to filter out all those retrievals whose IFOV or EFOV does not meet the requirements on spatial resolution. The 25 km spatial resolution requirement, shown as a dashed black line in figure 15, highlights that all the samples acquired with an incidence angle greater than ~54.5° do not fulfill the requirement, since their EFOV (or IFOV) is higher than 25 km, and therefore need to be discarded. This last filter applied to the retrievals also contributes to improve the final RMS error, as shown in Figure 33. The improvement is particularly strong when a low RCG threshold is used, suggesting that samples with high incidence angle also tend to have a low RCG, and discarding them improves the quality of the final set of observables and hence of the retrievals. However, filtering out these samples implies a reduction in the final coverage. The reduction in the total number of samples from the full 13-day Nature run expressed as percentage is shown with respect to different RCG thresholds in Table 2.

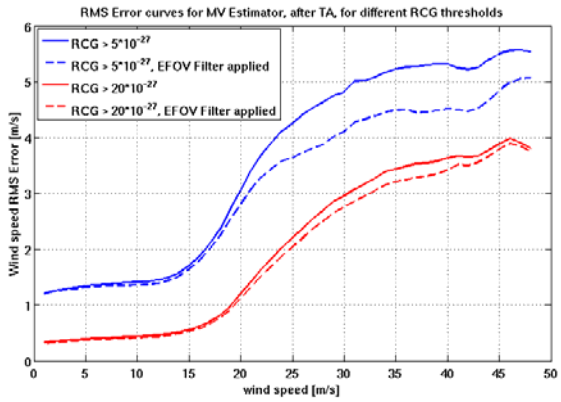


Figure 33. RMS error curves with (dashed line) and without (continuous line) EFOV filter, for two different RCG thresholds.

RCG Threshold	Number of samples	Number of samples with $\theta_i > 54.5°$	% of reduction on number of samples
RCG > 3	2029070	397193	29.6%
RCG > 5	1830639	297216	16.2%
RCG > 10	1418675	175384	12.4%
RCG > 20	1201712	73135	6.1%

Table 2. reduction in the number of samples when EFOV filter is applied, for different cases of different RCG thresholds.

9. Quality Control Flags in the Retrieval Algorithm

Some flags for data quality control have been set-up for the L2 wind speed retrieval algorithm. At present, the retrievals are flagged as "bad", and therefore simply removed, in the following cases:

1) When the retrieved wind speed is lower than zero
2) When the retrieved wind speed is higher than 70 m/s, since it starts to hit unrealistic values for wind speed.

These flags are at the moment all applied at the end of the retrieval algorithm processing steps, namely after the implementation of the MV estimator, the time averaging and the EFOV filter. Alternatively, these flags could be applied before, to the DDMA and LES retrievals, and then again to the MV estimator, since time averaging and EFOV filter are by design not going to generate invalid retrievals if the input retrievals are valid.
Another aspect worth considering is whether or not the observables with a value lower than zero should be still taken into account in the retrieval, or whether they should also be flagged as bad data and removed from the observable dataset. While for LES this might be appropriate (i.e. it is expected that the leading edge of the integrated delay waveform is always positive), this strategy is more controversial as far as DDMA is concerned. Since the noise floor is removed from the DDM samples prior to computing the DDMA, a negative DDMA is not necessarily wrong, but it simply indicates that the noise floor is stronger than the signal itself, therefore causing a negative DDMA value. In some cases, it is preferred to try and retrieve winds from all the possible observable values, even for those cases where the signal is very weak. However, the typical shape of the LUT shown in Figure 25 is such that a negative DDMA value will certainly translate into a very high retrieved winds, which is in any case likely to go above the 70m/s threshold set for realistic wind speeds.

10. Results and Performances for 25km Spatial Resolution

Here we present an overview of the algorithm performances when the CYGNSS baseline 25km spatial resolution requirement is considered.

Wind speed RMS error curves as a function of the true wind speed, for different RCG thresholds, are shown in Figure 34, along with the baseline requirement on the RMS error. The different RCG thresholds are used to determine spatial coverage statistics. From Figure 34, we can conclude that while requirements are reasonably met in all cases for low wind speeds (i.e. lower than ~15 m/s), an RCG higher than 10 is the minimum one that meets the requirements at wind speeds higher than ~15 m/s.

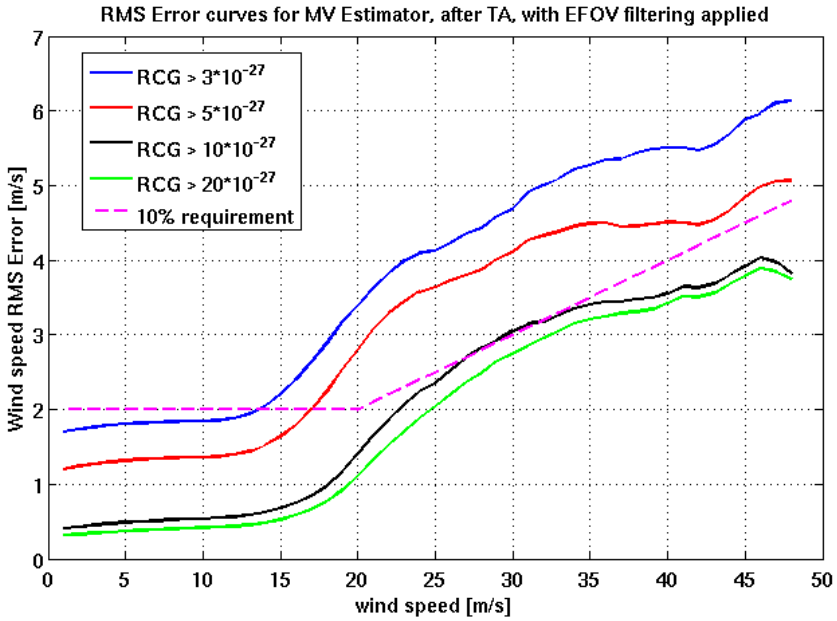


Figure 34. Final RMS Error curves versus true wind speed, for four different RCG thresholds.

A slightly different way to look at the RMS error is presented in Figure 35, where the curves represent this time the relative RMS error, i.e. the RMS error divided by the true wind speed at the center of each bin.

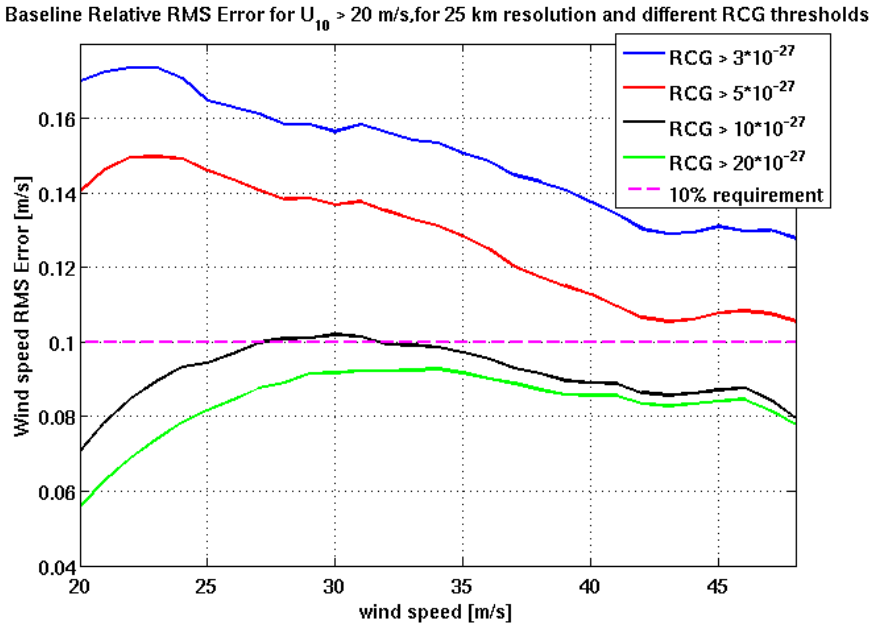


Figure 35. Relative RMS error curves versus true wind speed, for four different RCG thresholds.

The RMS error performances are also shown with respect to different RCG intervals, in Figure 36. The different RCG intervals are of interest since they have been used to fine tune the retrieval covariance, which is ultimately used for assimilating CYGNSS data into the forecast models. In this case, the RMS error is much worse for the first three cases, since each interval no longer contains clean samples with high RCG values (as in the case of RCG thresholds), which were highly contributing to the performance improvement. The last case (RCG>20), shown as green line, is identical to that shown in Figure 34.

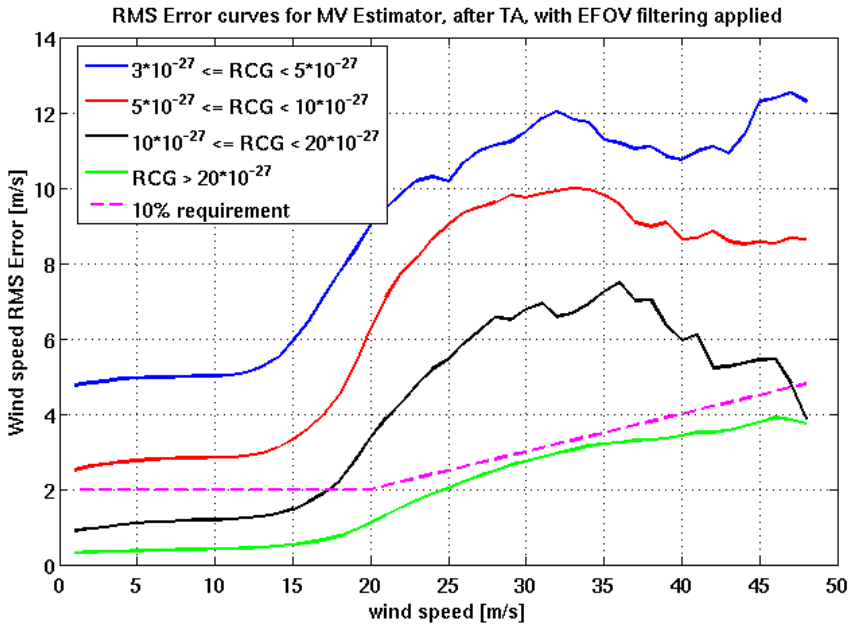


Figure 36. Final RMS Error curves versus true wind speed, for four different RCG intervals.

Table 3 and Table 4 illustrate the absolute average RMS error for all cases with true wind speeds lower than 20 m/s, respectively for different RCG thresholds, and different RCG intervals. In Table 3, we notice that the requirements of an RMS error below 2 m/s is well satisfied for a RCG threshold as low as 5, whereas in Table 4 requirements are only satisfied when 10 < RCG < 20.

Res = 25 km U_{10} < 20 m/s	**RCG > 3 * 10^{-27} Absolute RMS Err [m/s]**	**RCG > 5 * 10^{-27} Absolute RMS Err [m/s]**	**RCG > 10 * 10^{-27} Absolute RMS Err [m/s]**
Average RMS Error	1.8977	1.4041	0.5730

Table 3. Summary of performances for winds lower than 20 m/s, for different RCG thresholds.

Res = 25 km U_{10} < 20 m/s	**3 * 10^{-27} <= RCG < 5 * 10^{-27} Absolute RMS Err [m/s]**	**5 * 10^{-27} <= RCG < 10 * 10^{-27} Absolute RMS Err [m/s]**	**10 * 10^{-27} <= RCG < 20 * 10^{-27} Absolute RMS Err [m/s]**
Average RMS Error	5.3467	2.9983	1.2184

Table 4. Summary of performances for winds lower than 20 m/s, for different RCG intervals.

Table 5 and Table 6 show instead the relative average RMS error for true wind speeds higher than 20 m/s, and again respectively for different RCG thresholds, and different RCG intervals. The relative average RMS error is calculated as illustrated in the following two equations:

$$RMS_{REL} = \int_{v>20} \overline{RMS}(v) f_v(v \mid H) dv \tag{8.16.a}$$

$$\overline{RMS}_{REL}\left(U^i_{10(c)}\right) = RMS_{REL}\left(U^i_{10(c)}\right) \Big/ U^i_{10(c)} \tag{8.16.b}$$

where $f_v(v \mid H)$ is the conditional PDF of the true wind speeds, with H representing the condition of wind higher than a given threshold (in this case, higher than 20 m/s).

Res = 25 km U_{10} > 20 m/s	**RCG > 5 * 10^{-27} Rel. RMS Err [%]**	**RCG > 10 * 10^{-27} Rel. RMS Err [%]**	**RCG > 20 * 10^{-27} Rel. RMS Err [%]**
Average RMS Error	14.0%	**9.2%**	**7.9%**
Av. RMS Error for 20 m/s <= U10 < 30 m/s	16.3%	11.9%	10.4%
Av. RMS Error for 30 m/s <= U10 < 40 m/s	12.2%	9.7%	9.3%
Av. RMS Error for U10 >= 40 m/s	8.6%	6.4%	6.2%

Table 5. Summary of performances for winds higher than 20 m/s, for different RCG thresholds and different wind speed ranges.

Res = 25 km U_{10} > 20 m/s	**3 * 10^{-27} <= RCG < 5 * 10^{-27} Rel. RMS Err [%]**	**5 * 10^{-27} <= RCG < 10 * 10^{-27} Rel. RMS Err [%]**	**10 * 10^{-27} <= RCG < 20 * 10^{-27} Rel. RMS Err [%]**
Average RMS Error	40.1%	32.7%	20.1%
Av. RMS Error for 20 m/s <= U10 < 30 m/s	42.8%	38.2%	26.9%
Av. RMS Error for 30 m/s <= U10 < 40 m/s	45.7%	29.5%	22.1%
Av. RMS Error for U10 >= 40 m/s	47.3%	20.7%	9.3%

Table 6. Summary of performances for winds higher than 20 m/s, for different RCG intervals and different wind speed ranges.

The RMS error in Table 5 and Table 6 is shown for both a single wind speed threshold (i.e. all winds higher than 20 m/s), and for different wind speed intervals. Table 5 highlights that the minimum RCG threshold required to meet the baseline requirements for winds higher than 20 m/s is 10, consistently with what has been shown in Figure 34 and Figure 35. It also shows that the RMS error decreases with increasing wind speed, probably because the wind at the center of each bin (i.e. the denominator in equation (8.16b)) increases more rapidly than the increase in the absolute RMS error (i.e. the numerator in equation (8.16b)).

10.1 Some examples of Retrieved vs True Winds

In this last section, we present some comparisons of true and retrieved winds, for five cases characterized by different RCG values. A plot of the true wind speeds versus RCG, for all the 13-day Nature Run data, is shown in Figure 37.

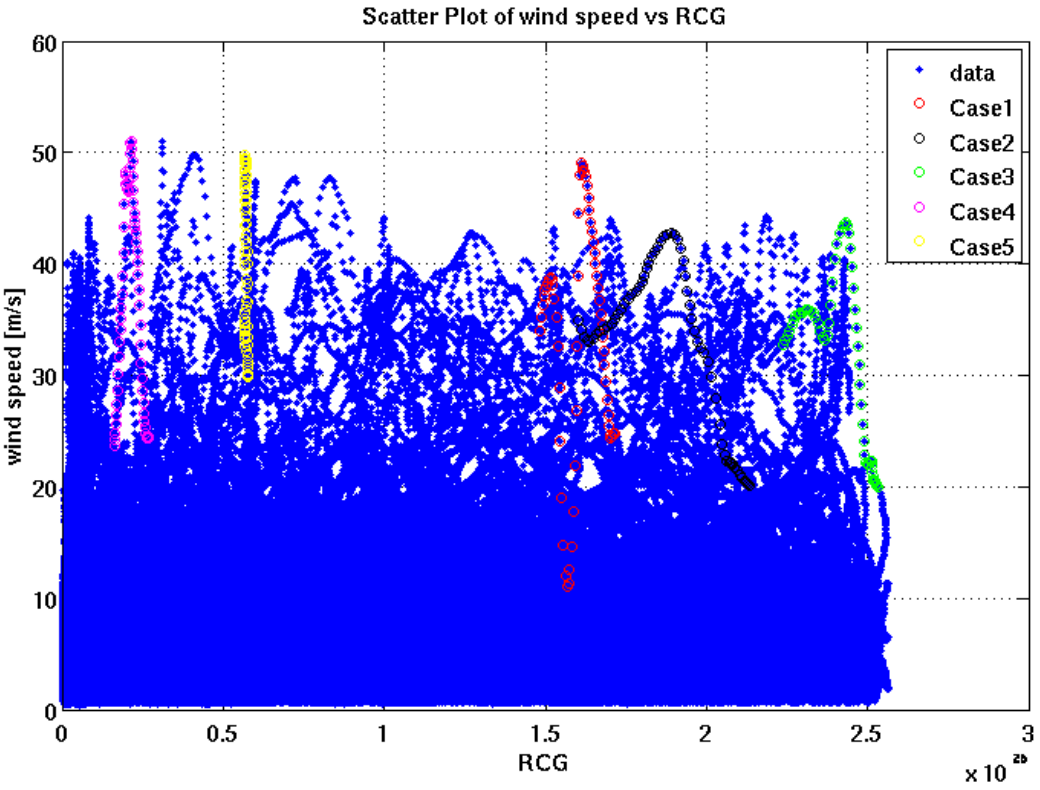
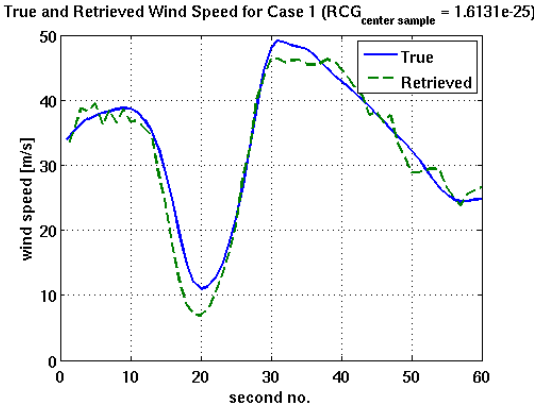


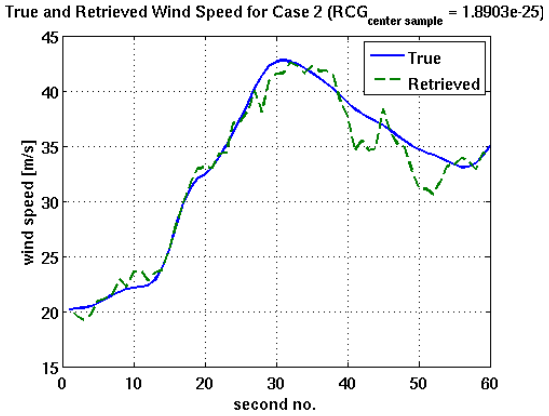
Figure 37. plot of true wind speed versus RCG, for all the data from the 13-day Nature Run simulation.

From this plot, we have selected five different specular point tracks of 60 seconds each, that can be easily identified in the plot, which reach high enough wind speeds. The five cases are also highlighted in Figure 37, and it can be noticed that while the first three are in the high RCG range, the last two cases correspond instead to medium RCG values. Plots of the true and retrieved wind speeds versus second number, for cases from 1 to 5, are shown in Figure 38. In some cases (i.e. case 1) the retrievals match the true winds very well, and this is generally true for the first three cases of high RCG values. The latter two cases show instead that the error between the two is much higher.

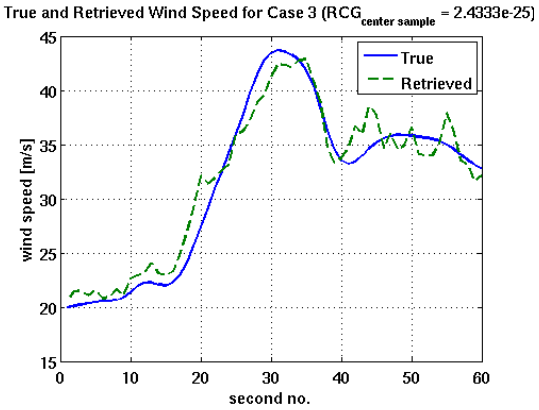
CASE 1



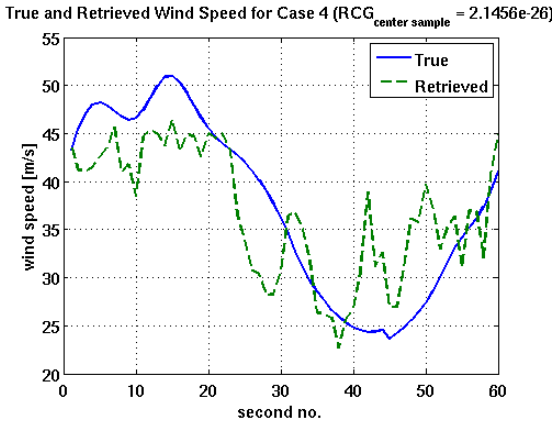
CASE 2



CASE 3



CASE 4



CASE 5

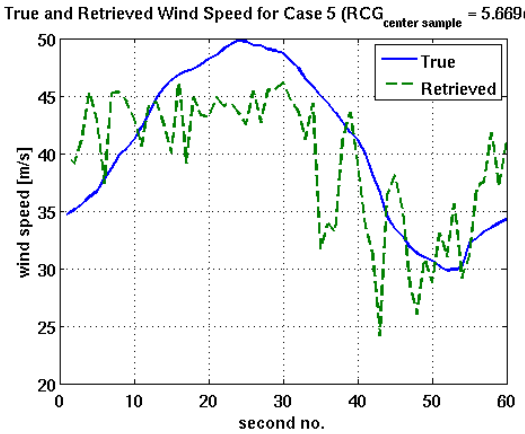


Figure 38. illustration of true and retrieved wind speeds, versus second no., for 5 cases of 60-second acquisition each.

11. Summary and Conclusions

Here we summarize the main changes and additions of the L2 wind speed retrieval algorithm, compared to the algorithm outlined in [Clarizia et al., 2013c], as well as the main conclusions.

The main characteristics and additional processing steps of the present algorithm are:

- Observable DD-range selected based on spatial resolution
- L2a corrections
- Use of 13-day Nature Run with 50/50 split of population
- Derivation of GMF from training samples, and use of L2a correction to derive a GMF for every incidence angle interval
- Wind speed MV Estimator using RCG-dependent covariance matrix
- Time Averaging
- EFOV filter

The general performances of the retrieval algorithm for the CYGNSS baseline spatial resolution requirement of 25 km x 25 km have been presented in the previous paragraphs, and characterized using RCG thresholds and RCG intervals, and for the two distinct cases of wind speed lower than 20 m/s and higher than 20 m/s. The main conclusions are summarized in Table 7, where the final retrieval error is illustrated for the two RCG thresholds that allow meeting the requirements, for low and high wind speed. The 24hr spatial coverage corresponding to that RCG threshold selection is also reported.

Res = 25 km	**RCG Threshold**	**Avg RMS Retrieval Error**	**24 hr Spatial Coverage**
U_{10} < 20 m/s	> 5e-27	1.40 m/s	79.6% (8 s/c) 75.0% (7 s/c)
U_{10} > 20 m/s	> 10e-27	9.2%	69.6% (8 s/c) 64.8% (7 s/c)

Table 7. summary of performances for the CYGNSS retrieval algorithm.

12. References

[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 et al., 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., Ruf, C. S., Jales P. and Gommenginger C., "Spaceborne GNSS-R Minimum Variance Wind Speed Estimator", *IEEE Transactions on Geoscience and Remote Sensing*, in review.

[Clarizia et al., 2013b] 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., 2013c] Clarizia M.P., Chang P., Gleason S., Jelenak Z., Johnson J., O'Brien A., Ruf C., Yi Y., Zavorotny V., "Algorithm Theoretical Basis Document: Level 2 Wind Speed Retrieval (Version 1)", Pre-Release Draft, June 2013.

[Clarizia et al., 2013d] Clarizia M.P., Chang P., Gleason S., Jelenak Z., Johnson J., O'Brien A., Ruf C., Yi Y. and Zavorotny V., "Algorithm Theoretical Basis Document: End-to-End Simulator Technical Memo ", Pre-Release Draft, May 2013.

[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.

[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, 2014] Gleason, S. T., "Algorithm Theoretical Basis Document: Level 1b DDM Calibration", December 2014.

[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.

[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.

[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 et al., 2013] Ruf C. and M.P. Clarizia, "Derivation of a Minimum Variance Unbiased Estimator from Delay-Doppler Map Average and Delay Waveform Slope", CYGNSS Technical Note, June 2013.

[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 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.

13. Appendix: L1b Corrections for Leading Edge Slope

We have shown in Section 5.4 that the L1b correction for DDMA is given by:

$$DDMA_{L1b} = C_F \cdot DDMA_{L1a}$$
$$C_F = \frac{\left(R_0^{SP} R^{SP}\right)^2}{G_{RX}^{SP} A_{DDMA}} \tag{A.1}$$

The following demonstrates that the same correction applies to the LES, namely:

$$LES_{L1b} = C_F \cdot LES_{L1a}$$
$$C_F = \frac{\left(R_0^{SP} R^{SP}\right)^2}{G_{RX}^{SP} A_{DDMA}} \tag{A.2}$$

The LES from L1a DDMs of scattered power is obtained through linear regression of the rising edge of the Integrated Delay Waveforms, and the equation for the slope from the simple linear regression method is:

$$LES_{L1a}\left(\Delta\tau, \Delta f\right) = \frac{\sum_{m=1}^{M} \tau_m y_m - \frac{1}{M}\sum_{m=1}^{M} \tau_m \sum_{m=1}^{M} y_m}{\sum_{m=1}^{M} \tau_m^2 - \frac{1}{M}\left(\sum_{m=1}^{M} \tau_m\right)^2} \tag{A.3}$$

Where τ_m is the delay and the x-value of the IDW, and y_m is the y-value of the IDW derived from the L1a DDM, which can be expressed as:

$$y_m = \sum_{n=1}^{N} \overline{Y}\left(\tau_m, f_n\right) \tag{A.4}$$

In principle, the L1b LES should be computed from DDMs whose pixels are L1b corrected individually, i.e.

$$LES_{L1b}\left(\Delta\tau, \Delta f\right) = \frac{\sum_{m=1}^{M} \tau_m \hat{y}_m - \frac{1}{M}\sum_{m=1}^{M} \tau_m \sum_{m=1}^{M} \hat{y}_m}{\sum_{m=1}^{M} \tau_m^2 - \frac{1}{M}\left(\sum_{m=1}^{M} \tau_m\right)^2} \tag{A.5}$$

$$\hat{y}_m = \sum_{n=1}^{N} CF\left(\tau_m, f_n\right) * \overline{Y}\left(\tau_m, f_n\right) \tag{A.6}$$

Where $\hat{y}_m$ is the y-value of the IDW from the L1b DDM. If CF is assumed constant as for the DDMA (and for the reasons explained in) it can be factored out of the summation as follows:

$$\hat{y}_m = CF * \sum_{n=1}^{N} \bar{Y}\left(\tau_m, f_n\right) = CF * y_m \tag{A.7}$$

and the final equation for the L1b LES becomes:

$$LES_{L1b}\left(\Delta\tau, \Delta f\right) = CF * \frac{\sum_{m=1}^{M} \tau_m y_m - \frac{1}{M}\sum_{m=1}^{M} \tau_m \sum_{m=1}^{M} y_m}{\sum_{m=1}^{M} \tau_m^2 - \frac{1}{M}\left(\sum_{m=1}^{M} \tau_m\right)^2} = CF * LES_{L1a}\left(\Delta\tau, \Delta f\right) \tag{A.8}$$