ASSIMILATION OF GNSS-R DELAY-DOPPLER MAPS INTO WEATHER MODELS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Feixiong Huang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. James L. Garrison, Chair

School of Aeronautics and Astronautics

Dr. Robin L. Tanamachi

Department of Earth, Atmospheric, and Planetary Sciences

Dr. Arthur E. Frazho

School of Aeronautics and Astronautics

Dr. Dengfeng Sun

School of Aeronautics and Astronautics

Approved by:

Dr. William A. Crossley

Head of the School Graduate Program

ACKNOWLEDGMENTS

First and foremost, I am deeply grateful to my major advisor, Prof. James Garrison, who is accomplished, open-minded and innovative, for the support and guidance during my PhD studies. I would particularly like to thank Prof. Garrison for giving me opportunities of joining the CYGNSS science team, attending various conferences/workshops and having internships. Those experiences gave me chances to communicate with scientists all over the world and helped me become an independent researcher.

I would like to acknowledge my committee members, Prof. Robin Tanamachi, Prof. Arthur Frazho and Prof. Dengfeng Sun for their patient instructions on meteorology and mathematics. Special thanks to Prof. Tanamachi for giving a lot of valuable suggestions on my dissertation.

I would like to express my sincere gratitude to another two of my advisors outside the university, Dr. Ross Hoffman from NOAA and Dr. Ad Stoffelen from KNMI. Dr. Hoffman is the co-advisor of my research project and has provided a great many important suggestions on my work. Dr. Stoffelen is my advisor when I was visiting KNMI. The visit at KNMI saved my PhD and will be an invaluable experience in my life.

I would also like to thank my collaborators and also friends, Mark Leidner, Bachir Annane and Giuseppe Grieco, for the kind help and instructions during my entire PhD, without those I would not have been able to complete this research. Special thanks to Bachir and Mark for helping producing results in Chapter 6.

Thanks to all my labmates and friends, Benjamin Nold, Han Zhang, Abi Komanduru, Soon Chye Ho, Priyankar Bhattacharjee, Seho Kim, Kevin Shi, Kaitie Schoenfeldt, Sidd Subramanyam, Archana Suhas Choudhari, Jared Covert and many

more, who provided academic suggestions, stimulating discussions, encouragement as well as happy distractions.

Lastly, I am deeply thankful to my parents for their unfailing love and support.

The work of this dissertation was supported in part by NASA under Grant NNX15AU18G, "Assimilation of GNSS-R Delay-Doppler Maps into Hurricane Models" and in part by the EUMETSAT OSI SAF Visiting Scientist Program.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

ABBREVIATIONS

BKE	Basic Kinematic Equation
BLUE	Best Linear Unbiased Estimator
BRCS	Bistatic Radar Cross-section
CYGNSS	Cyclone Global Navigation Satellite System
CDR	Climate Data Record
DA	Data Assimilation
DDM	Delay-Doppler Map
DDMI	Delay-Doppler Map Instrument
E2ES	End-to-End Simulator
ECEF	Earthcentered Earth-fixed
ECMWF	European Centre for Medium-Range Weather Forecasts
EIRP	Equivalent Isotropically Radiated Power
EKF	Extended Kalman Filter
FFT	Fast Fourier Transform
FSP	Frequency of Superior Performance
FSW	Flight Software
GFS	Global Forecast System
GMF	Geophysical Model Function
GNSS	Global Navigation Satellite System
GNSS-R	Global Navigation Satellite System Reflectometry
GPS	Global Positioning System
GSI	Gridpoint Statistical Interpolation
HWRF	Hurricane Weather Research and Forecasting
KA-GO	Kirchoff Approximation to Geometric Optics

KNMI	Royal Netherlands Meteorological Institute
LES	Leading Edge Slope
MSS	Mean Square Slope
MV	Minimum Variance
NASA	National Aeronautics and Space Administration
NBRCS	Normalized Nistatic Radar Cross-section
NCEP	National Centers for Environmental Prediction
NHC	National Hurricane Center
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OSE	Observing System Experiment
PDF	Probability Density Function
PRN	Pseudo-Random-Noise
QC	Quality Control
RMSE	Root-mean-square Error
SCAT	Scatterometer
SD	Standard Deviation
SNR	Signal-to-noise Ratio
SWH	Significant Wave Height
SOC	Science Operation Center
SSA	Small-slope Approximation
SV	Space Vehicle
TC	Tropical Cyclone
VAM	Variational Analysis Method
WRF	Weather Research and Forecasting
WVC	Wind Vector Cell

ABSTRACT

Huang, Feixiong Ph.D., Purdue University, December 2020. Assimilation of GNSS-R Delay-Doppler Maps into Weather Models. Major Professor: James L. Garrison.

Global Navigation Satellite System Reflectometry (GNSS-R) is a remote sensing technique that uses reflected satellite navigation signals from the Earth surface in a bistatic radar configuration. GNSS-R observations have been collected using receivers on stationary, airborne and spaceborne platforms. The delay-Doppler map (DDM) is the fundamental GNSS-R measurement from which ocean surface wind speed can be retrieved. GNSS-R observations can be assimilated into numerical weather prediction models to improve weather analyses and forecasts. The direct assimilation of DDM observations shows potential superiority over the assimilation of wind retrievals.

This dissertation demonstrates the direct assimilation of GNSS-R DDMs using a two-dimensional variational analysis method (VAM). First, the observation forward model and its Jacobian are developed. Then, the observation's bias correction, quality control, and error characterization are presented. The DDM assimilation was applied to a global and a regional case.

In the global case, DDM observations from the NASA Cyclone Global Navigation Satellite System (CYGNSS) mission are assimilated into global ocean surface wind analyses using the European Centre for Medium-Range Weather Forecasts (ECMWF) 10-meter winds as the background. The wind analyses are improved as a result of the DDM assimilation. VAM can also be used to derive a new type of wind vector observation from DDMs (VAM-DDM).

In the regional case, an observing system experiment (OSE) is used to quantify the impact of VAM-DDM wind vectors from CYGNSS on hurricane forecasts, in the case of Hurricane Michael (2018). It is found that the assimilation of VAM-DDM

wind vectors at the early stage of the hurricane improves the forecasted track and intensity.

The research of this dissertation implies potential benefits of DDM assimilation for future research and operational applications.

1. INTRODUCTION

1.1 GNSS-Reflectometry for Ocean Surface Winds

Ocean surface winds are important for the understanding of the dynamical ocean-atmosphere process, ocean navigation and also weather forecasting. As measuring winds by stations, ships, weather buoys and aircraft has very limited spatial sampling, remote sensing by spaceborne instruments can provide observations with global coverage.

Two primary remote sensors for ocean surface winds are the passive microwave radiometer and the scatterometer. Both operate at microwave frequencies, which can penetrate clouds and take measurements day and night. Passive microwave radiometers like SSM/I, TMI, AMSR, GMI, SMAP and WindSat measure the thermal emission of the ocean surface. Wind speeds can be retrieved by measuring the excess emission caused by surface roughness and sea foam which are related to the surface winds [3, 4]. Wind directions can be measured by a polarimetric radiometer like WindSat [5]. Scatterometers like NSCAT, QuikScat, ASCAT and OSCAT actively transmit a pulse signal and measure the backscatter from the ocean surface. The measured backscattered radar cross section is related to the surface roughness through the Bragg-resonant scattering [6]. Wind speeds and directions can be retrieved by special processing techniques [7, 8]. Radar altimeters and Synthetic Aperture Radar (SAR) also have the capability of measuring ocean surface winds, but with limited spatial coverage [9, 10].

Global Navigation Satellite System Reflectometry (GNSS-R) is a relatively new remote sensing technique that uses satellite navigation (GNSS) transmitters as non-cooperative sources in a bistatic radar configuration [11]. Ocean surface wind speed is one variable that can be estimated from GNSS-R observations. The Rayleigh criterion indicates that the ocean surface, under most conditions, will appear rough in the L-band wavelength ($\approx$ 20 cm) used by satellite navigation signals. GNSS signals are therefore scattered from a region on the rough ocean surface that much larger than the first Fresnel zone through the quasi-specular forward scattering. That region, centered at the specular reflection point (specular point), is called the glistening zone.

The concept of GNSS-R was first proposed by Martin-Neira regarding GPS satellites as sources of opportunity [12]. It was later validated by a number of stationary and airborne experiments [13–18]. Recently, a sequence of satellite technology demonstrations, including the UK Disaster Monitoring Constellation (UK-DMC) [19], UK TechDemoSat-1 (TDS-1) [20], NASA Cyclone Global Navigation Satellite System (CYGNSS) [21], Japan's WNISAT-1R [22] and China's BuFeng-1 A/B [23] has validated the capability of ocean wind speed retrieval from GNSS-R signals. The spaceborne GNSS-R missions provide a much broader spatial coverage of the wind observations using abundant signals from GNSS constellations, which have the potential to improve the weather forecasts.

1.2 GNSS-R Measurement and Wind Speed Retrieval Algorithm

The fundamental GNSS-R measurement is the delay-Doppler map (DDM), the cross correlation between a reflected GNSS signal and a locally generated model of the transmitted signal (code and carrier) over a range of delay, and Doppler frequencies.

In a GNSS-R receiver, the reflected signal, $u_r(t)$, is first cross-correlated with a local copy of transmitted signal over a range of delays, τ and Doppler frequencies, f

$$X(t,\tau,f) = \frac{1}{T_I}\int_{t}^{t+T_I} u_r(t')a(t'+\tau)e^{2\pi j(f_0+f)t'}dt'. \tag{1.1}$$

$X(t,\tau,f)$ is the complex voltage correlation result at time t, $a(t)$ is the baseband signal model (generated from a pseudorandom noise (PRN) code using Binary Phase Shift Keyed (BPSK) modulation) and T_I is the integration time. N sequential complex results are then incoherently averaged

$$Z(t,\tau,f) = \frac{1}{N}\sum_{m=1}^{N} |X(t+(m-1)T_I,\tau,f)|^2. \tag{1.2}$$

A measurement (in units of Watts) of the DDM power at the front-end of the receiver is produced by subtracting the noise floor, $Y_N(t)$, and calibration of the receiver gain, G_R [24, 25].

$$Y(t,\tau,f) = \frac{Z(t,\tau,f)}{G_R} - Y_N(t). \tag{1.3}$$

Numerous wind retrieval approaches can be derived from this DDM measurement. They can generally be grouped in two categories: 1) fitting a parameterized scattering model to the observed DDM, or some truncated or integrated form of it, and 2) derivation of an observable easily computed from samples of the DDM shown to have a strong, monotonic, dependence on the surface wind speed. Multiple examples of these two approaches can be found in the literature. Early experiments from airborne field campaigns (e.g. [15, 26]) demonstrated that the shape of the DDM is related to the roughness of the scattering surface. Empirical models have been developed from these airborne experiments, relating surface statistics to wind speed [17]. Examples of observables computed from a set of DDM samples at discrete delay-Doppler coordinates include: trailing or leading edge slopes, the average DDM power exceeding a threshold [18, 19, 27], and principal components analysis (PCA) [28]. Optimal combinations of observables, exploiting the partial decorrelation between residual errors

have been developed [29]. Some recent studies carried out new approaches based on statistical method [30, 31] or neural network [32–34].

Applying these methods to GNSS-R measurements from satellite orbit presents a challenge due to the large area of the ocean surface contributing to the DDM (glistening zone). Model-fitting approaches are not feasible, since a wind field cannot be assumed homogeneous across a large glistening zone observed from orbit. Resolution requirements (25 km in the present study) will also limit the usable range of delay and Doppler to only a few samples around the specular point. In this region, there is little sensitivity of the DDM shape to the surface winds. Consequently, observables defined on such a small region are essentially measurements of the average scattered power within the resolution cell.

1.3 The CYGNSS Mission

As all studies in this dissertation are based on the CYGNSS mission, this section gives a brief overview of the CYGNSS mission. Launched on December 15, 2016, CYGNSS is a NASA Earth Venture Mission consisting of an eight microsatellite constellation of GNSS-R observatories. The motivation for CYGNSS is to improve the forecast skill of tropical cyclone (TC) intensity through increased observation of the inner core [21,35]. GNSS transmissions lie in L-band (1-2 GHz), which has a lower rain attenuation than the higher frequencies typically used by scatterometers. This can increase the available observations within the inner core of a TC. Additionally, a constellation of small satellites (enabled by the order-of-magnitude lower size, weight and power of a reflectometry instrument vs. that of an active radar) at an inclination of 35°, provides better spatial and temporal sampling with up to 32 simultaneous measurements per second and an average revisit time of 7 hours across the Earth between -38° to 38° in latitude. This high rate of spatial and temporal sampling enables improved observation of the rapid intensification stage in TC development.

The constellation and low-inclination orbit of CYGNSS also fill the temporal and spatial gaps from conventional microwave instruments mostly in the polar orbit.

The Delay Doppler Mapping Instrument (DDMI) on each CYGNSS satellite can nominally track up to four GNSS reflections, providing one DDM per second at each reflection over a range of 128 delays and 20 Dopplers, sampled at 0.25 chip and 500 Hz, respectively [36]. The integration time T_i in equation (1.1) is 1 ms and the number of incoherent samples N in equation (1.2) is 1000. Communication bandwidth limitations for CYGNSS require that the DDM is truncated or "compressed" on board by the flight software, reducing it to a matrix of 17×11 pixels before transmission to the ground for processing. The CYGNSS Science Operation Center (SOC) generates several levels of data products from these DDMs. At Level 1a (L1a), the DDM is calibrated to absolute power in units of watts by equation (1.3). Figure 1.1 shows an example of the CYGNSS L1a DDM measurement. The "horseshoe" shape of the DDM represents power reflecting from a region on the ocean surface around the specular point, with a diameter ranging from 100 to 150 km, depending on the incidence angle and receiver altitude. Each sample of the DDM at a specific (τ, f) is sensitive to reflected power from points on the surface having a total path delay within one code chip and Doppler frequency within 1 kHz of (τ, f).

The Level 1b data product contains the bistatic radar cross-section (BRCS), generated from the DDM by removing the effects of satellite geometries, attitudes, direct power and antenna gain [24]. Level 2a data consists of the surface wind speed retrieval at the specular point with a resolution of 25 km. This retrieval uses the Average DDM (DDMA) and the Leading Edge Slope (LES) observables computed over a 3×5 window aligned with the specular point delay. Geophysical model functions (GMFs) were derived for these observables by fitting a model to training data [37]. Different sea state conditions (wave age and fetch length) near versus far from a hurricane can result in different sensitivities of GNSS-R observables to the wind speed. To resolve this issue, a GMF for fully developed seas (FDS) condition was developed based on matchups with global numerical model values at low-to-moderate wind speeds and

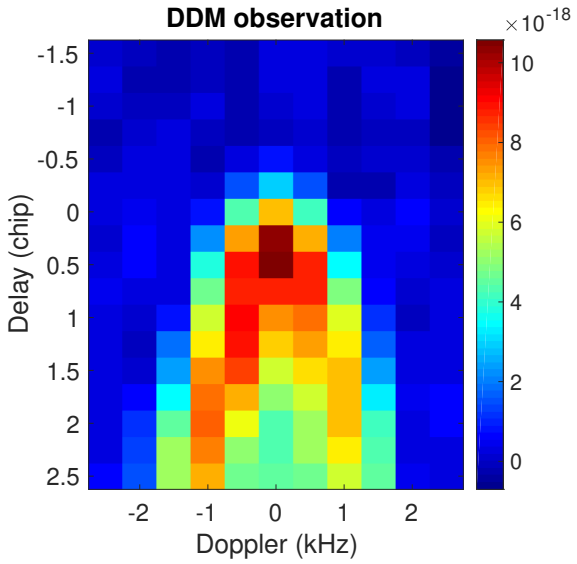


Figure 1.1.: An example of the CYGNSS L1a DDM. Units in Watts.

another GMF for young seas/limited fetch (YSLF) was developed based on matchups with aircraft measurements over hurricanes [38]. A minimum variance estimator is used to optimally combine the DDMA and LES retrievals [29]. The mean-square slope (MSS) of the ocean surface is also provided as a Level 2b data product, obtained directly from the BRCS [39]. In a comprehensive assessment study using analytical models of the measurement and matchups with numerical model values, the overall root mean square (RMS) uncertainty of the CYGNSS L2 wind speed retrievals is 1.4 m/s at wind speeds below 20 m/s and 17% at wind speeds above 20 m/s [40].

1.4 Numerical Weather Prediction and Data Assimilation

Numerical Weather Prediction (NWP) is the basis of almost all present-day weather forecasts. NWP models use mathematical and physical equations based on the dynamics of the atmosphere and related physical processes to predict atmosphere and ocean parameters such as pressure, temperature, humidity and wind. The atmospheric and ocean parameters are defined by a discretized geophysical field which is called the "state". Given the initial and boundary conditions of the state, those

dynamical equations (usually partial differential equations) governed by the physical processes can be run within a geographic area and predict the state at a future time.

NWP models are not perfect. They cannot precisely describe the complicated real motions of atmosphere and ocean but only approximately reflect the dynamical properties of atmosphere and ocean at a certain scale. The prediction accuracy of NWP models degrades remarkably with time. So observation data from traditional in situ methods (e.g., ships and bouys) and modern remote sensing methods (mainly from satellites) are required to correct the forecasts by NWP models. The primary goal of data assimilation (DA) is to find the best estimation of the state from an *a priori* estimate of the state and observations [41]. The secondary goal of DA is to quantify the uncertainty of the state estimation. Almost every NWP model is combined with a DA system. The *a priori* state is usually from the most recent model prediction and called the "background". After assimilating the observation data, the estimation of the state that optimally combines the background and observations is called the "analysis". Then the analysis is used as the initial condition of the NWP model to make a "forecast" by numerically integrating the dynamical equations in the model forward in time in discrete intervals. The backgrounds, analyses and forecasts constitute the cycles of a DA system. Figure 1.2 gives an illustration of the DA.

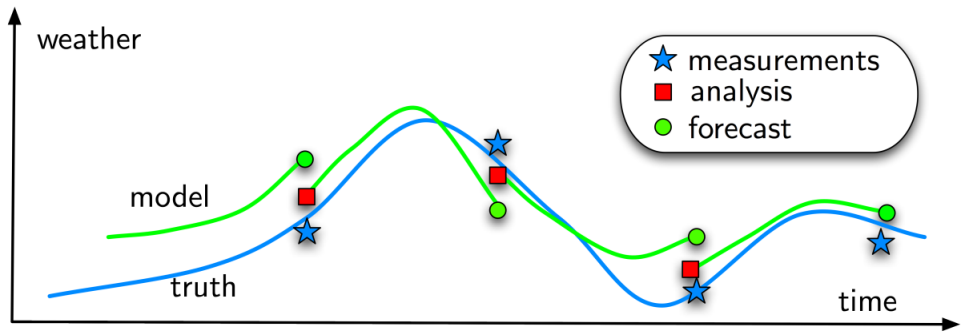


Figure 1.2.: Illustration of data assimilation.[1]

DA is basically an estimation problem. Both the observations and model forecasts exhibit errors resulting from the combined effects of observation (e.g., instrument

[1]Reference: http://pedagotech.inp-toulouse.fr/130202/co/1_5-chain.html

noise) and model errors (e.g., imperfect parameterization of sub-grid-scale processes). The observations are sparse both in time and space across the Earth. The observations and models can also have different spatial resolutions. To obtain the optimal estimate, which is also called Best Linear Unbias Estimator (BLUE), many approaches have been developed. In general, there are two types of DA approaches: stochastic methods and variational methods. Stochastic methods such as Ensemble Kalman filter (EnKF) [42] sequentially assimilate observations and directly compute analytical solutions of the BLUE. Variational methods such as 3D-Var [43, 44] and 4D-Var [45] compute the estimate through a minimization of the cost function. Stochastic methods compute the optimal solution and consider flow-dependent errors but are much more costly. 4D-Var is generally better than 3D-VAR as it considers errors' variation with time within a time window as opposed to only a single time.

The scale of motion in meteorology can be categorized into the global scale, synoptic scale, mesoscale and microscale, from the largest to smallest. Similarly, NWP models with DA systems can be classified into global and regional models. Global models produce forecasts for the entire world and regional models produce forecasts specific to a region at a higher resolution. Examples of global models are ECMWF (European Centre for Medium-Range Weather Forecasts), US GFS (Global Forecast System), ICON (Global German Standard), UKMO (United Kingdom Met Office) and US NOAA CFS (Climate Forecast System). Examples of regional models are WRF (Weather Research and Forecasting), NAM (North American Mesoscale), and HRRR (High Resolution Rapid Refresh). [2]

1.5 Motivation of This Research

Observations in DA systems can originate from a lot of platforms including ships, buoys, radiosondes, profilers, aircraft and satellites, etc. Some instruments (especially conventional ones) directly measure the NWP model state variables like temperature,

[2]Reference: https://windy.app/blog/what-is-a-weather-forecast-model-guide-on-forecast-models-all-around-the-world.html

humidity, water vapor and wind, while some other instruments (especially from satellites using remote sensing techniques) take measurements that are related to the state variables according to some physical processes (usually electromagnetic radiation). Those measurements can be, for example, atmospheric radiance from ocean color sensors, backscatter from radar or lidar, brightness temperature from radiometer and bending angle from GPS radio occultation receiver [46–49]. Model state variables can be computed from those measurements through algorithms with some physical or empirical models and are called "retrievals".

The assimilation of direct remote sensing measurements rather than the retrievals has been widely used. While a more sophisticated implementation is generally required, the assimilation of direct measurements can utilize more information from the raw data. Moreover, retrieval algorithms are usually based on some assumptions which are not always valid. Some retrieval algorithms use NWP data as reference, which can introduce correlated errors in DA [50].

The CYGNSS L2 retrieved wind speeds have been assimilated into NWP models in many studies. Recent results show that CYGNSS wind observations can improve the forecast of TCs in track, intensity and structure [51–55]. Another DA experiment based on multiscale tropical weather systems showed that simulated CYGNSS winds could improve the low-level wind and temperature [56]. A preliminary study of assimilating CYGNSS winds into global NWP models presented its advantage in offering more details in the global tropical surface winds [57]. While CYGNSS wind speed retrievals have shown its benefits in DA systems, the more fundamental measurement, CYGNSS Level 1 DDM power can also be assimilated. The advantages of assimilating DDMs versus wind speed retrievals are as follows:

1. The observables, NBRCS and LES, used for the CYGNSS wind retrieval are calculated by assuming the geometries and power parameters for DDM pixels in the 3×5 box are the same. The failure of this assumption can introduce non-geophysical dependence to the observables. Direct assimilation of the DDM can

account for all the non-geophysical factors by a physical-based forward operator or emulator which transforms the wind speed into a DDM.

2. As a lower-level measurement, the full DDM contains more information on the ocean reflections than the retrieved wind speed which is retrieved from only uses a few measurements around the specular point.

3. With a larger footprint (about 100 km) of the full DDM, the assimilation of DDMs can impact the analysis with a broader area.

4. The CYGNSS specular point moves in about 6 km/s on the Earth's surface, giving each point on the ocean surface along the track to be observed by more than 15 sequential DDMs. This feature provides a large number of "multi-look" observations and could reach better accuracy if the observations are characterized properly to avoid over-fitting.

The major challenge of the DDM assimilation is that a forward model that relates the DDM power to the surface wind speeds is required. Other than that, quality control, bias correction and error characterization of the DDM observations are also crucial. The research of this dissertation aims to address all related issues of DDM assimilation and apply it to both global and regional NWP models. In the global case, CYGNSS DDMs are assimilated into the ECMWF surface wind analyses by a Variational Analysis Method (VAM) and the results are assessed by collocated scatterometer winds. In the regional case, the VAM is used as a preprocess to produce wind vector observations from CYGNSS DDMs. The wind vector observations are then assimilated into the Hurricane Weather Research and Forecasting (HWRF) model in an observing system experiment (OSE). The impact of CYGNSS DDMs on the hurricane forecast will be assessed and compared to the use of other types of CYGNSS observations (e.g., CYGNSS L2 wind product).

1.6 Organization of This Dissertation

This dissertation is organized as follows:

- Chapter 2 presents the development of the GNSS-R DDM forward model. The forward model is assessed by comparing DDMs computed from HWRF model winds against measured DDMs from the CYGNSS Level 1 data.
- Chapter 3 presents the method for DDM assimilation. A variational method combined with the DDM forward model is summarized. The bias correction and quality controls of the DDM observations are also presented.
- Chapter 4 presents the error characterization of DDM observations. Two methods to compute the DDM error covariance matrix are described.
- Chapter 5 shows the results of assimilating CYGNSS DDMs into global ocean surface wind analyses. The background is from the ECMWF operational forecast. The results are assessed by collocated scatterometer winds (ASCAT, OSCAT). The computational efficiency is discussed.
- Chapter 6 presents the OSE for the impact of DDM assimilation on the hurricane forecast. Hurricane Michael in 2018 was picked up as the study case.
- Chapter 7 summarizes this research and gives recommendations for future work.

2. THE GNSS-R DDM FORWARD MODEL

2.1 Introduction

As stated previously, DDM is a more fundamental GNSS-R observable and incorporates information from a larger wind field than only that at the specular point. Direct assimilation of DDM power is, therefore, expected to better utilize more of the information in the GNSS-R measurement. In this chapter, an analytical GNSS-R forward model is developed to link the DDM to a gridded wind speed and enable assimilation of a calibrated 17×11 DDM (CYGNSS Level 1a data) into weather forecast models. Taking in a gridded wind field and GNSS-R metadata, the forward model produces a simulated DDM which will be evaluated through comparison with measured CYGNSS DDMs at a defined delay-Doppler coordinate. An analytical model for the Jacobian matrix, which describes the sensitivity of DDM power in each delay-Doppler bin to the gridded wind field, is also derived. These models meet all of the requirements for optimal assimilation of Level 1a DDM data. Furthermore, all of the code is written as a single callable function in the C programming language so that it can be easily integrated into any data assimilation system. The code of the forward model has been published on Code Ocean [58].

2.2 Model Description

Inputs to the forward model can separated into four data types: *Geometry*, including the positions and velocities of the transmitter, receiver and specular point; *Metadata*, including sample time, GPS PRN code, and specular bin index; *Power*, including the Effective Isotropic Radiated Power (EIRP) of the GPS transmitter and antenna parameters of the receiver; *Wind Field*, a gridded wind field in latitude/-

longitude coordinates. *Geometry, Metadata* and *Power* (except the receiver antenna patterns) are all from CYGNSS Level 1 data. The gridded wind field has a resolution of 0.125° in geodetic coordinates. The forward model has two outputs: a modeled DDM produced by the forward operator and a Jacobian matrix computed by analytically differentiating the forward operator. Specific variables names from the CYGNSS Level 1 data are listed in the Appendix. Fundamental code from the CYGNSS End-to-End Simulator (E2ES) [59] has been reused.

2.2.1 Forward Operator

The DDM observation $Y(\tau, f)$ is described in section 1.2. Assume that $Y(\tau, f)$ is an unbiased measurement,

$$Y(\tau, f) = h(\tau, f, \boldsymbol{x}) + \nu(\tau, f) \tag{2.1}$$

in which $h(\tau, f, \boldsymbol{x})$ is a model for the DDM. The vector $\boldsymbol{x}$ contains a set of wind speed values, ordered in geodetic latitude and longitude coordinates, which define the wind field on the ocean surface. The spatial resolution of the wind field used in this forward model is 0.125° in latitude and longitude. ν is a zero-mean random variable that represents the error of the measurement.

With knowledge of the wind field, satellite geometries, and power parameters, a widely used model for the DDM, derived from the Kirchoff Approximation to Geometric Optics (KA-GO), can be written as a surface integral [60],

$$h(\tau, f, \boldsymbol{x}) = \frac{\lambda^2 P_t G_t G_0}{(4\pi)^3} \iint \frac{G_r(\vec{\rho})}{R_t^2(\vec{\rho}) R_r^2(\vec{\rho})} \sigma^0\left(\vec{\rho}, x(\vec{\rho})\right) \times \chi^2\left(\tau - \tau_g(\vec{\rho}), f - f_g(\vec{\rho})\right) d\vec{\rho}. \tag{2.2}$$

The variable of integration, $\vec{\rho}$, is a two-dimensional position vector defining points on the ocean surface. λ is the wavelength of the GPS carrier. $P_t G_t$ is the GPS EIRP which is assumed constant over the entire scattering area. G_r is the receiver antenna gain. $R_t(\vec{\rho})$ and $R_r(\vec{\rho})$ are path distances from the mean ocean surface

at $\vec{\rho}$ to the transmitter and receiver, respectively. $x(\vec{\rho})$ is the wind speed at $\vec{\rho}$. $\chi^2\left(\tau - \tau_g(\vec{\rho}), f - f_g(\vec{\rho})\right)$ is the ambiguity function in which $\tau_g(\vec{\rho})$ is the "geometric" path delay through a point on the mean ocean surface at $\vec{\rho}$ and $f_g(\vec{\rho})$ is the corresponding Doppler frequency. An additional coefficient, the excess power gain, G_0, is included to account for uncalibrated biases. $G_0 = 1$ by default, but it can be adjusted to best fit the observed data.

σ^0 is the bistatic radar cross section (BRCS). In the KA-GO model, σ^0 depends on the probability density function of surface slopes, $p(\vec{s}, x)$,

$$\sigma^0(\vec{\rho}, x(\vec{\rho})) = \pi \left|\Re(\vec{\rho})\right|^2 \frac{q^4(\vec{\rho})}{q_z^4(\vec{\rho})} p(\vec{s}(\vec{\rho}), x(\vec{\rho})) \tag{2.3}$$

where $\Re$ is the Fresnel reflection coefficient. $\vec{q}(\vec{\rho})$ is the bisector,

$$\vec{q}(\vec{\rho}) = \frac{2\pi}{\lambda}(\hat{g}(\vec{\rho}) - \hat{n}(\vec{\rho})) = \vec{q}_\perp + q_z\hat{k} \tag{2.4}$$

computed from $\hat{g}$ and $\hat{n}$ which are unit vectors along the incident and reflected ray paths. This can be decomposed into $\vec{q}_\perp$ in the mean ocean surface and q_z perpendicular to it. Fig. 2.1 illustrates this geometry and the definition of relevant vectors. Note that $\vec{q}$ is in the scattering plane formed by $\hat{g}$ and $\hat{n}$.

The slope $\vec{s}$ can then be expressed as $\vec{s} = -\vec{q}_\perp/q_z$. Assuming an isotropic normal distribution for the surface slopes, $p(\vec{s}, x)$ is defined by a single parameter, the omnidirectional MSS, m,

$$p(\vec{s}, x) = \frac{1}{2\pi m(x)} \exp\left(-\frac{|\vec{s}|^2}{2m(x)}\right). \tag{2.5}$$

A common empirical model [17] is used to provide a monotonic relationship between MSS and wind speed.

$$m(x) = 0.225(0.003 + 0.00508 f(x)) \tag{2.6}$$

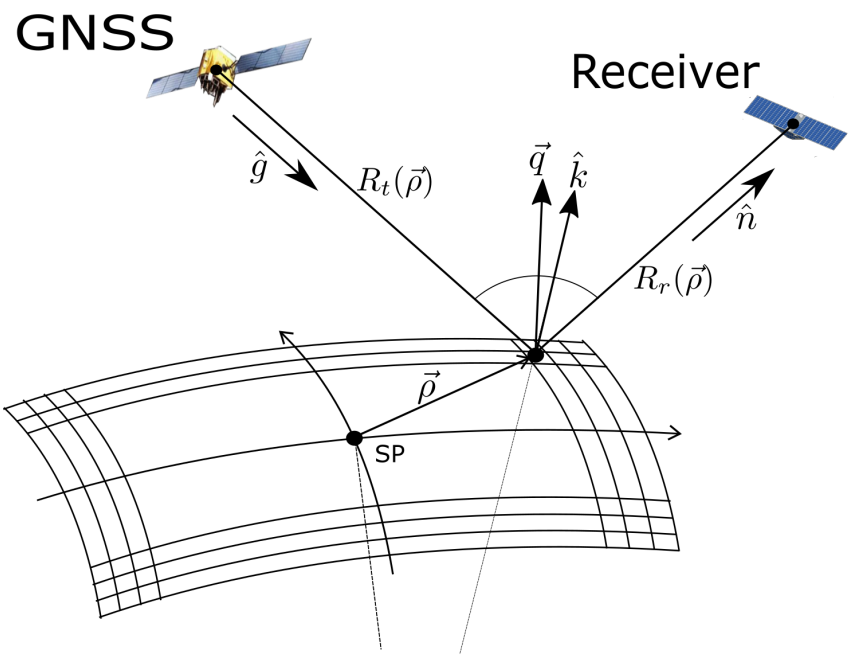


Figure 2.1.: Illustration of the geometry: $\vec{q}$ is defined as the bisector of the position vectors $\hat{g}$ from the transmitter to the specular point and $\hat{n}$ from the specular point to the receiver, with $R_t(\vec{\rho})$ and $R_r(\vec{\rho})$ are the corresponding distances; $\hat{k}$ is a local normal to the ocean surface at $\vec{\rho}$.

$$f(x) = \begin{cases} x & \text{if } 0 < x < 3.49 \\ 6\ln(x) - 4 & \text{if } 3.49 < x < 46 \\ 0.411x & \text{if } x > 46 \end{cases} \tag{2.7}$$

where the numerical MSS value representing the isotropic case, was obtained by averaging the up-wind and cross-wind slope variances, $m = (\sigma_c^2 + \sigma_u^2)/2$.

To accelerate computation, fast Fourier transform (FFT) methods can be applied to (2.2) formulated as a 2D convolution [61, 62]

$$h(\tau, f, \boldsymbol{x}) = Q(\tau, f, \boldsymbol{x}) * *\chi^2(\tau, f), \tag{2.8}$$

between the ambiguity function $\chi^2(\tau, f)$ and

$$Q(\tau, f, \boldsymbol{x}) = \iint B(\vec{\rho})\sigma^0(\vec{\rho}, x(\vec{\rho})) \times \delta(\tau - \tau_g(\vec{\rho}))\delta(f - f_g(\vec{\rho}))d\vec{\rho}. \tag{2.9}$$

$\delta()$ is the Dirac delta function and

$$B(\vec{\rho}) = \frac{\lambda^2 P_t G_t G_0}{(4\pi)^3} \frac{G_r(\vec{\rho})}{R_t^2(\vec{\rho}) R_r^2(\vec{\rho})}. \tag{2.10}$$

The delay-Doppler space needs to be discretized for computing (2.8). For application to data assimilation, the modeled DDM is also required to use the same delay-Doppler coordinates as the observed DDM, so as to enable calculation of difference between the observation and the forward model. Specular point position estimates generated by the DDMI on-board CYGNSS are very coarse and fixed to discrete values. Therefore, the true specular point is not generally located in the center of a single DDM bin [63]. Fractional precise specular bin indices, n_τ^s, n_f^s are generated in post-process and provided as part of the metadata listed in Table 2.1 [25]. n_τ^s, n_f^s are used to shift the delay-Doppler coordinates of the modeled DDM to align with those of the observed one.

Equation (2.8) is discretized and computed using a 2D FFT

$$h[n_\tau, n_f, \boldsymbol{x}] = Q[n_\tau, n_f, \boldsymbol{x}] * *\Xi[n_\tau, n_f] \tag{2.11}$$

in which

$$\Xi[n_\tau, n_f] = \chi^2(n_\tau \Delta\tau, n_f \Delta f). \tag{2.12}$$

$n_\tau = 1, 2, ..., N_\tau$ and $n_f = 1, 2, ..., N_f$ are indices in the discrete delay-Doppler space. $\Delta\tau$ and Δf are the delay and Doppler increments. For CYGNSS, $N_\tau = 17$, $N_f = 11$, $\Delta\tau = 0.25T_c$ where T_c is the period of a GPS C/A code chip in seconds and $\Delta f = 1/(2T_i) = 500$ Hz.

$Q[n_\tau, n_f, \boldsymbol{x}]$ is the discrete form of the surface integral (2.9). When the Earth-centered Earth-fixed (ECEF) coordinates of the transmitter, receiver and specular

Table 2.1.: Parameters of CYGNSS Level 1 data used in the forward model (© 2020 IEEE)

Data Type	Variable Name	Type	Units	Description	Symbol
Geometry	tx_pos_x	int	meter	Transmitter posistion X	(none)
Geometry	tx_pos_y	int	meter	Transmitter posistion Y	(none)
Geometry	tx_pos_z	int	meter	Transmitter posistion Z	(none)
Geometry	tx_vel_x	int	m/s	Transmitter velocity X	(none)
Geometry	tx_vel_y	int	m/s	Transmitter velocity Y	(none)
Geometry	tx_vel_z	int	m/s	Transmitter velocity Z	(none)
Geometry	sc_pos_x	int	meter	Receiver posistion X	(none)
Geometry	sc_pos_y	int	meter	Receiver posistion Y	(none)
Geometry	sc_pos_z	int	meter	Receiver posistion Z	(none)
Geometry	sc_vel_x	int	m/s	Receiver velocity X	(none)
Geometry	sc_vel_y	int	m/s	Receiver velocity Y	(none)
Geometry	sc_vel_z	int	m/s	Receiver velocity Z	(none)
Geometry	sp_pos_x	int	meter	Specular point posistion X	(none)
Geometry	sp_pos_y	int	meter	Specular point posistion Y	(none)
Geometry	sp_pos_z	int	meter	Specular point posistion Z	(none)
Power	gps_eirp_watt	float	Watts	GPS EIRP	$P_t G_t$
Power	ddm_ant	int	(none)	The antenna that received signal	(none)
Metadata	prn_code	int	(none)	GPS PRN code	(none)
Metadata	ddm_ant	int	(none)	The antenna that received signal	(none)
Metadata	ddm_sp_delay_row	float	(none)	Specular delay bin index	n_τ^s
Metadata	ddm_sp_dopp_col	float	(none)	Specular Doppler bin index	n_f^s
Metadata	ddm_timestamp_utc	int	(none)	DDM sample time	(none)
Metadata	quality_flags	int	(none)	DDM quality flags	(none)

point are known, a curvilinear system, identified as the "surface frame" (SURF), is created and used to discretize the surface near the specular point into a $n_r \times n_v$ grid at a resolution of 1 km. The definition of the SURF frame is shown in Appendix A. The total area covered by this grid is 120 km × 120 km in order to encompass the expected glistening zone for all scenarios.

A bilinear interpolation is used to interpolate $\boldsymbol{x}$, provided at fixed points in geodetic coordinates, to 1 km grid points in the SURF frame, $\boldsymbol{x}'$.

$$\boldsymbol{x}' = \boldsymbol{M}\boldsymbol{x} \tag{2.13}$$

$\boldsymbol{x}'$ is a N' $\times$ 1 vector, $\boldsymbol{x}$ is a N $\times$ 1 vector and $\boldsymbol{M}$ is a N' $\times$ N transformation matrix. N depends on the geometry, $N' = 120 \times 120 = 14400$, and the computation of $\boldsymbol{M}$ can be referenced in Appendix A. $Q[n_\tau, n_f, \boldsymbol{x}]$ can then be computed numerically as

$$Q[n_\tau, n_f, \boldsymbol{x}] = \Delta A \sum_{i=1}^{N'} \left(B(\vec{\rho_i}) \sigma^0(\vec{\rho_i}, x_i') D[n_\tau, n_f, \vec{\rho_i}] \right) \tag{2.14}$$

in which $\Delta A = 1000^2$ is the discrete increment in surface area and

$$D[n_\tau, n_f, \vec{\rho_i}] = \delta \left[\left\lfloor n_\tau - n_\tau^s - \frac{\tau_g(\vec{\rho_i}) - \tau_g^s}{\Delta \tau} \right\rceil \right] \times \delta \left[\left\lfloor n_f - n_f^s - \frac{f_g(\vec{\rho_i}) - f_g^s}{\Delta f} \right\rceil \right]. \tag{2.15}$$

$\delta[\,]$ is the Kronecker delta function and $\lfloor \rceil$ stands for rounding to the nearest integer. τ_g^s and f_g^s are path delay and Doppler frequency at the specular point. τ_g and τ_g^s are in units of seconds and f_g and f_g^s are in units of Hz.

2.2.2 Jacobian

The Jacobian matrix is a partial derivative matrix of the forward operator, with respect to each wind speed in the input vector, $\boldsymbol{x}$.

$$\boldsymbol{H} = \frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial h[1,1,\boldsymbol{x}]}{\partial x_1} & \cdots & \frac{\partial h[1,1,\boldsymbol{x}]}{\partial x_N} \\ \vdots & \cdots & \vdots \\ \frac{\partial h[1,N_f,\boldsymbol{x}]}{\partial x_1} & \cdots & \frac{\partial h[1,N_f,\boldsymbol{x}]}{\partial x_N} \\ \frac{\partial h[2,1,\boldsymbol{x}]}{\partial x_1} & \cdots & \frac{\partial h[2,1,\boldsymbol{x}]}{\partial x_N} \\ \vdots & \cdots & \vdots \\ \frac{\partial h[N_\tau,N_f,\boldsymbol{x}]}{\partial x_1} & \cdots & \frac{\partial h[N_\tau,N_f,\boldsymbol{x}]}{\partial x_N} \end{bmatrix}. \tag{2.16}$$

To compute (2.16), equation (2.11) has to be linearized with respect to $\boldsymbol{x}'$. However, calculating the derivative with $\boldsymbol{x}'$, at the 1 km resolution, would be computationally expensive. To reduce the computational cost, the SURF frame is resampled at a 10 km resolution by selecting $K = N'/100 = 144$ points out of the N' points. The

resulting matrix of surface wind speeds is then "unrolled" into a $K \times 1$ vector, $\boldsymbol{X}$. Fig. 2.2 shows the selection of the $N'/100$ points in $\boldsymbol{X}$ and their spatial relation to $\boldsymbol{x}'$ and $\boldsymbol{x}$. Equation (2.11) then becomes

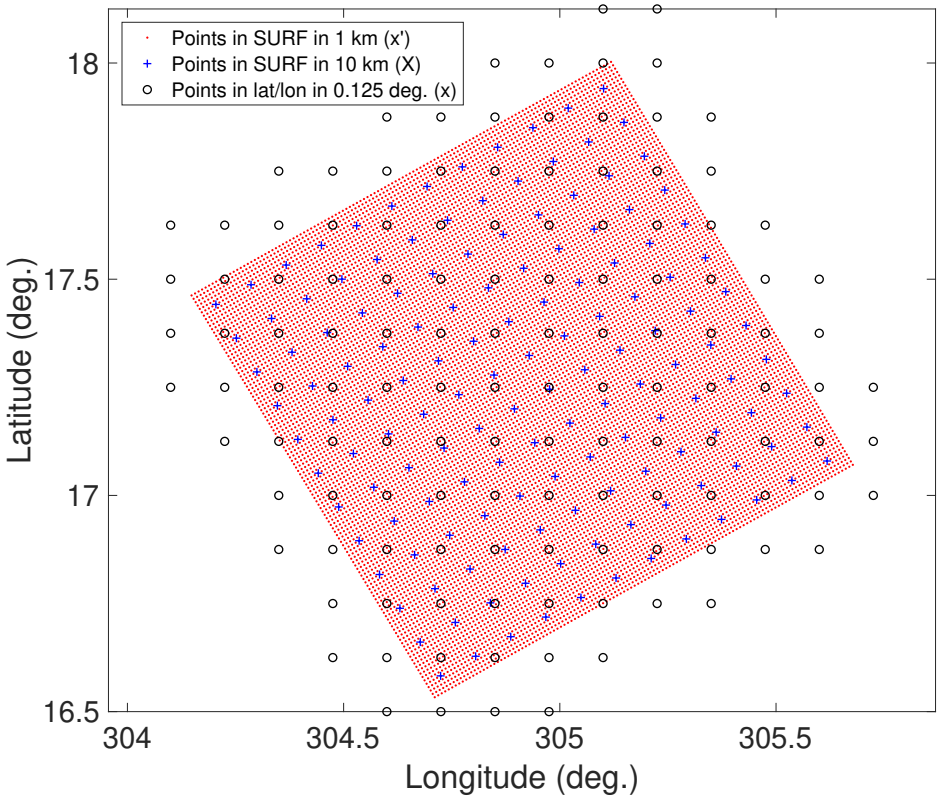


Figure 2.2.: Spatial relation of the grid points in three different resolutions: (1) grid points in SURF in 1km resolution ($\boldsymbol{x}'$); (2) grid points in SURF in 10km resolution ($\boldsymbol{X}$). (3) grid points in latitude/longitude coordinates in 0.125° resolution ($\boldsymbol{x}$). $\boldsymbol{X}$ are evenly distributed among $\boldsymbol{x}'$. © 2020 IEEE.

$$h[n_\tau, n_f, \boldsymbol{X}] = Q'[n_\tau, n_f, \boldsymbol{X}] * *\Xi[n_\tau, n_f] \tag{2.17}$$

in which

$$Q'[n_\tau, n_f, \boldsymbol{X}] = 100\Delta A \sum_{i=1}^{K} \left(B(\vec{\rho_i}) \sigma^0(\vec{\rho_i}, X_i) D[n_\tau, n_f, \vec{\rho_i}] \right) \tag{2.18}$$

and the discrete increment in surface area is $100\Delta A$.

Taking the derivative of (2.17) with respect to each X_i, we obtain

$$\frac{\partial h[n_\tau, n_f, \boldsymbol{X}]}{\partial X_i} = W_i[n_\tau, n_f, \boldsymbol{X}] * *\Xi[n_\tau, n_f] \tag{2.19}$$

in which

$$W_i[n_\tau, n_f, \boldsymbol{X}] = 100\Delta A\Big(B(\vec{\rho_i})\pi\,|\Re(\vec{\rho_i})|^2\,\frac{q^4(\vec{\rho_i})}{q_z^4(\vec{\rho_i})} \times \frac{\partial P(\vec{s}(\vec{\rho_i}), X_i)}{\partial X_i} D[n_\tau, n_f, \vec{\rho_i}]\Big). \tag{2.20}$$

Only the slope-PDF is dependent on the wind speed, so this is the only derivative which must be computed. Assuming a normal distribution allows a closed-form expression for this derivative,

$$\frac{\partial P(\vec{s}(\vec{\rho}), X)}{\partial X} = \frac{1}{2\pi m^2(X)}\left(\frac{|\vec{s}(\vec{\rho})|^2}{2m(X)} - 1\right) \times \exp\left(-\frac{|\vec{s}(\vec{\rho})|^2}{2m(X)}\right)\frac{dm(X)}{dX} \tag{2.21}$$

in which

$$\frac{dm(X)}{dX} = \begin{cases} 1.143 \times 10^{-3} & \text{if } 0 < X < 3.49 \\ \frac{6.858\times10^{-3}}{X} & \text{if } 3.49 < X < 46 \\ 4.69773 \times 10^{-4} & \text{if } X > 46. \end{cases} \tag{2.22}$$

A matrix, $\boldsymbol{H_0}$, is formed from $N_\tau N_f \times K$ (187×144 in the particular application to CYGNSS) elements of (2.19). The last step is to transform this matrix to geodetic coordinates. With the coordinate transformation in Appendix A, the latitude and longitude of each point in $\boldsymbol{X}$ can be calculated. The transformation matrix $\boldsymbol{T}$ from $\boldsymbol{x}$ (at 1 km) to $\boldsymbol{X}$ (at 10 km)

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{x} \tag{2.23}$$

can be obtained from $\boldsymbol{M}$ averaging blocks of 100 terms.

$$T_{ij} = \frac{1}{100}\sum_{i' \in I_i} M_{i'j} \tag{2.24}$$

I_i is the set of indices of the 100 points in $\boldsymbol{x'}$ that are nearest to $\boldsymbol{X}_i$. $M_{i'j}$ is the i'-th row, j-th column element in matrix $\boldsymbol{M}$, $i = 1, 2, ..., K$, $i' = 1, 2, ..., N'$, $j = 1, 2, ..., N$, and $\boldsymbol{T}$ is a matrix in dimension of $K \times N$. Applying the chain rule

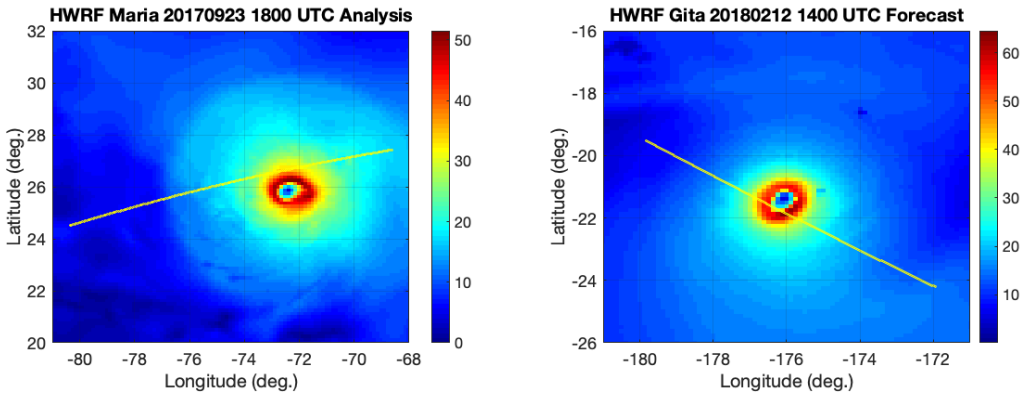
$$\frac{\partial h[n_\tau, n_f, \boldsymbol{x}]}{\partial x_i} = \frac{\partial h[n_\tau, n_f, \boldsymbol{X}]}{\partial X_1}\frac{dX_1}{dx_i} + \frac{\partial h[n_\tau, n_f, \boldsymbol{X}]}{\partial X_2}\frac{dX_2}{dx_i} + ... + \frac{\partial h[n_\tau, n_f, \boldsymbol{X}]}{\partial X_K}\frac{dX_K}{dx_i}, \tag{2.25}$$

the Jacobian matrix (2.16) can be obtained as

$$\boldsymbol{H} = \boldsymbol{H_0 T}. \tag{2.26}$$

2.3 Model Assessment

To assess the forward model and identify possible difficulties in assimilating DDMs, the model is applied to Hurricane Weather Research and Forecasting (HWRF) synoptic-scale (0.125° resolution) wind fields [64, 65] and the resulting DDMs are then compared against CYGNSS Level 1a data (version 2.1). CYGNSS measurements collected during overpasses of Hurricane Maria on September 23, 2017 and Cyclone Gita on February 12, 2018 are chosen for this comparison. One track in Maria covers the time from 18:03 to 18:08 UTC and the corresponding HWRF wind field is from the analysis at 18:00 UTC. For Cyclone Gita, the track covers the time from 14:03 to 14:05 UTC and is compared with winds from the HWRF 2-hour forecast at 14:00 UTC. The picked two tracks are the best tracks with good data quality that overpass the center of the tropical cyclones during the time. Since the time difference between the model and measurement is less than 10 minutes in this experiment, change of the wind field in such a short time period can be ignored. The CYGNSS specular tracks and wind fields for both cases are shown in Fig. 2.3. Measurements from GPS Block IIF satellites are removed from consideration because of the inaccurate estimation of the fluctuating transmitter power. The analysis wind field is reported with a resolution of 0.125°.



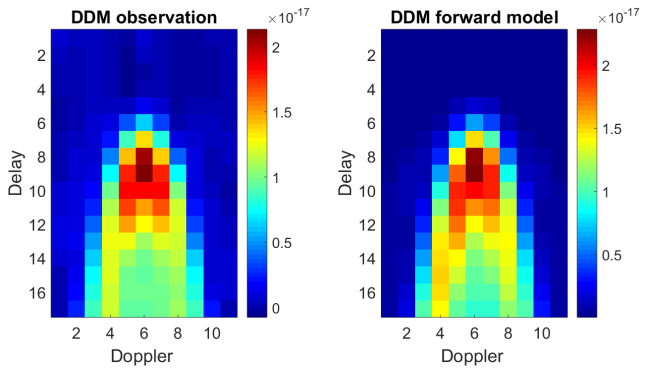
(a) (b)

Figure 2.3.: HWRF wind field and overpass of CYGNSS specular points (yellow lines): (a) Hurricane Maria on September 23, 2017 at 18:00 UTC and CYGNSS data on same day at around 18:06 UTC with CYGNSS SV 5 and GPS PRN 13. (b) Hurricane Gita on February 12, 2018 at 14:00 UTC and CYGNSS data on same day at around 14:04 UTC with CYGNSS SV 1 and GPS PRN 23. © 2020 IEEE.

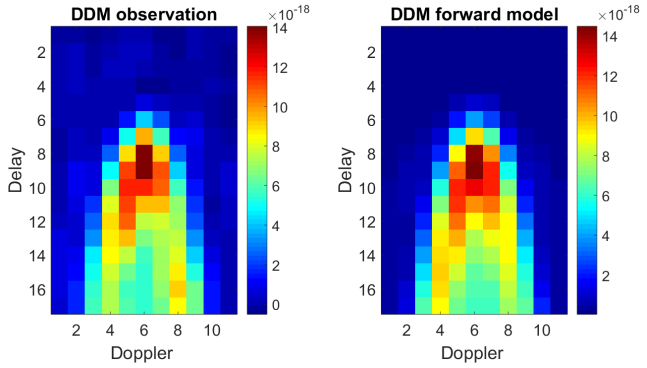
2.3.1 Assessment of the Forward Operator

Data for each variable in Table 2.1 were obtained from the CYGNSS L1a product at a 1 Hz rate. The receiver antenna patterns estimated by pre-launch measurements and on-orbit corrections were provided by the CYGNSS SOC [25]. At each time of observation, the modeled DDM produced by the forward operator was compared with that obtained from the L1a data. Fig. 2.4 shows examples of the forward operator for three different cases inside Cyclone Gita exhibiting average wind speeds in the glistening zone of 6 m/s (a), 15 m/s (b) and 30 m/s (c). It can be observed that for all cases, the modeled and observed DDMs have similar power values and asymmetric shapes. Two metrics were used to quantify the agreement between the forward operator and observations. To compare the absolute power between observed and modeled DDMs, the average relative difference of effective bins was computed,

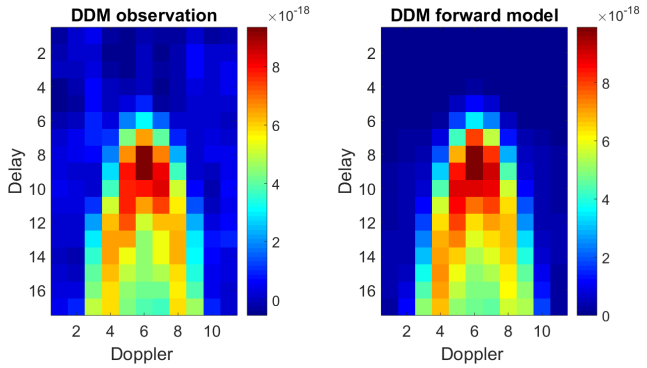
$$\epsilon_k = \frac{1}{N} \sum_{i,j} \frac{Y_k(\tau_i, f_j) - h_k(\tau_i, f_j)}{Y_k(\tau_i, f_j)}, \tag{2.27}$$



(a)



(b)



(c)

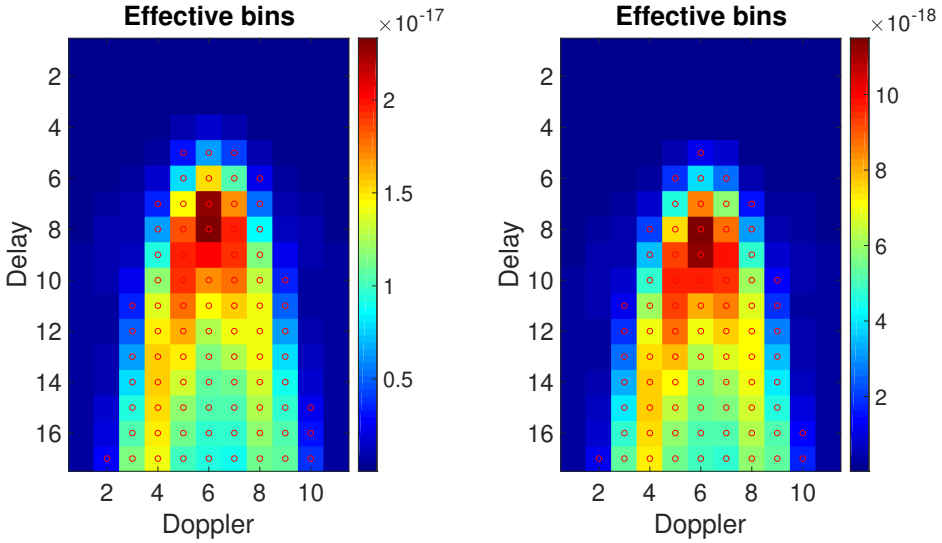
Figure 2.4.: Comparisons between observed DDM and modeled DDM at three different cases for Cyclone Gita on February 12, 2018: wind speed average over the glistening zone = 6 m/s (a), 15 m/s (b), and 35 m/s (c). DDM values are in units of Watts. © 2020 IEEE.

where k is the index of the DDM, Y_k and h_k are observed and modeled DDM, i and j are indices of effective bins and N is the number of effective bins. Effective bins

are defined as DDM samples having a power larger than 1/10 of the DDM peak. Fig. 2.5 shows the selection of DDM bins using this method for two different cases. To compare the shapes of the two DDMs, the correlation coefficient, ρ_k, between the effective bins of the two DDMs is computed from

$$\rho_k = \frac{\sum_{i,j}(Y_k(\tau_i, f_j) - \bar{Y})(h_k(\tau_i, f_j) - \bar{h}_k)}{\sqrt{(\sum_{i,j}(Y_k(\tau_i, f_j) - \bar{Y})^2)(\sum_{i,j}(h_k(\tau_i, f_j) - \bar{h}_k)^2)}} \tag{2.28}$$

where $\bar{Y}$ and $\bar{h}_k$ are mean values of all effective bins of the observed and modeled DDM [63]. When the shapes of the two DDMs are more similar, the coefficient should be closer to 1.

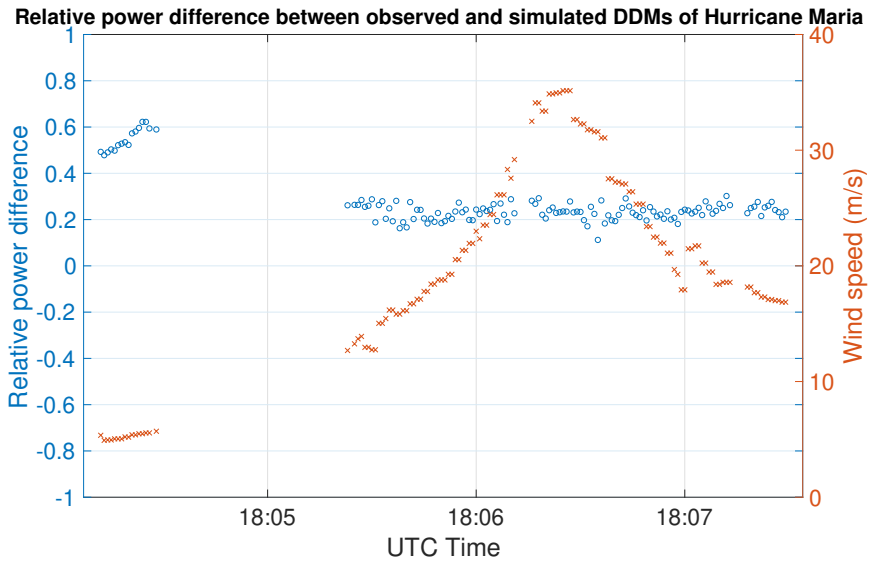


(a) (b)

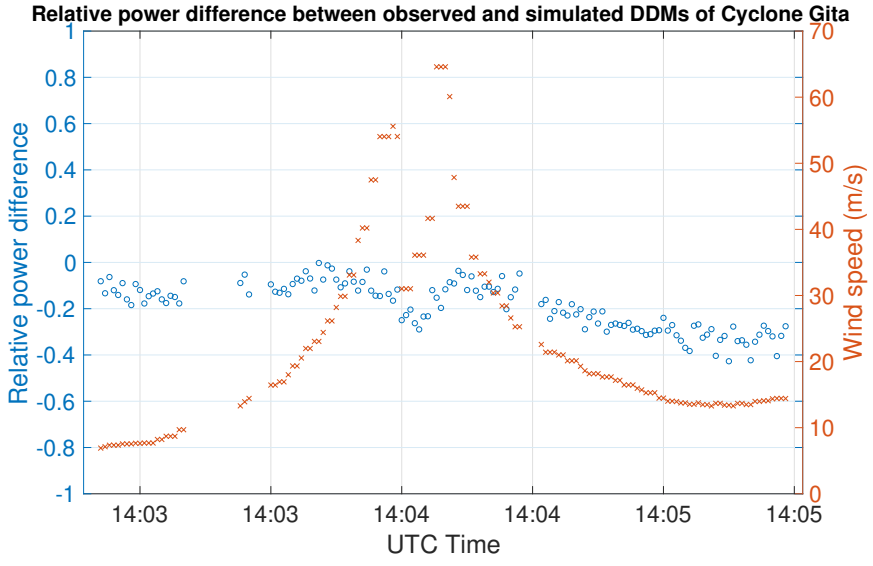
Figure 2.5.: Selection of effective DDM bins (red circles), using 1/10 of the peak power in two cases for Cyclone Gita on February 12, 2018: wind speed = 5 m/s (left) and 30 m/s (right). DDM values are in units of Watts. © 2020 IEEE.

Fig. 2.6 shows the average relative differences, ϵ_k, for the two tracks in Hurricane Maria and Cyclone Gita. Fig. 2.7 shows the correlation coefficients, ρ_k, for these

same tracks. Wind speeds at the specular point along these tracks are also plotted on these figures.

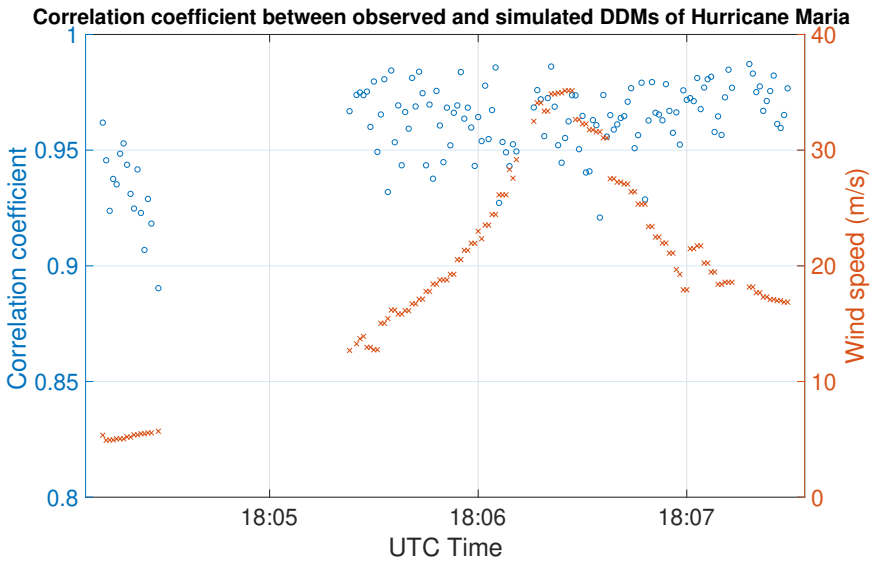


(a)

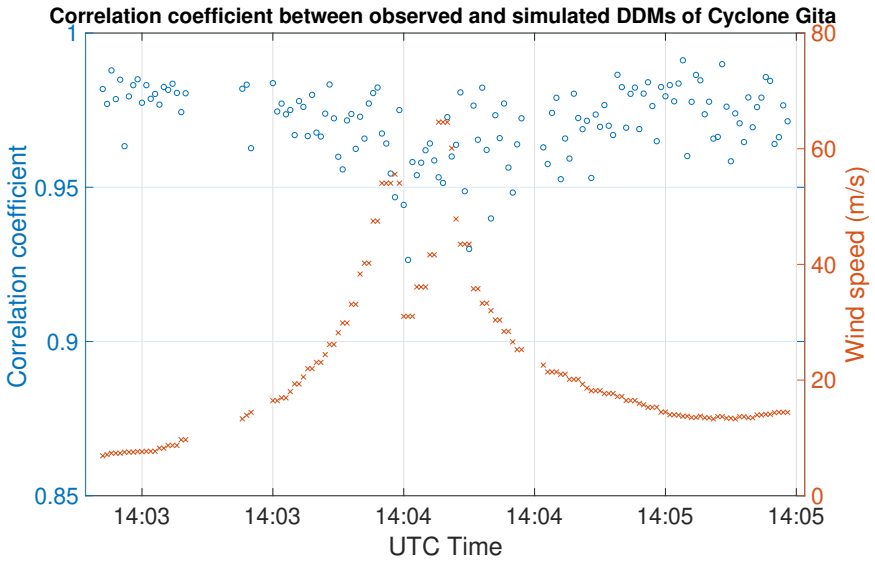


(b)

Figure 2.6.: Relative power differences between modeled DDMs and observed DDMs of the two tracks in (a) Hurricane Maria and (b) Cyclone Gita. © 2020 IEEE.



(a)



(b)

Figure 2.7.: Correlation coefficients between modeled DDMs and observed DDMs of the two tracks in (a) Hurricane Maria and (b) Cyclone Gita. © 2020 IEEE.

In Fig. 2.6, it can be seen that the relative difference mostly remains below 30%, except for some low wind speed cases for Hurricane Maria. The correlation coefficient plotted in Fig. 2.7 is mostly larger than 0.9, while there is a decrease in high wind

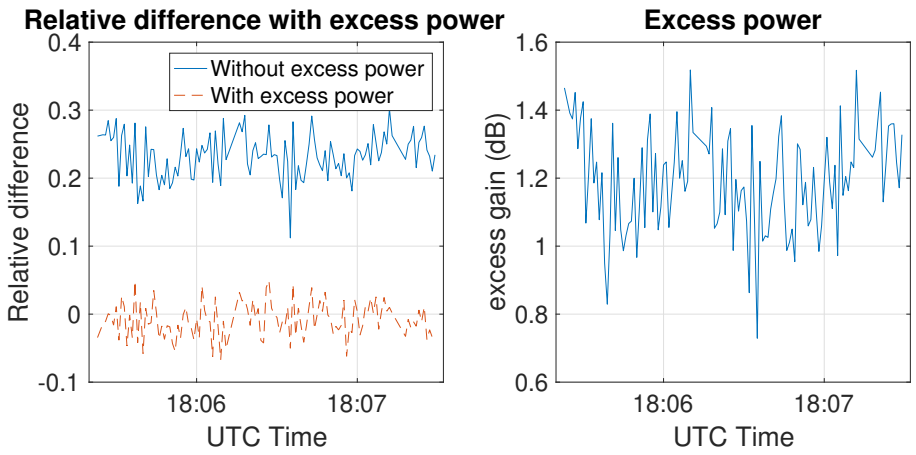
speed cases. Both figures show that the DDMs calculated by the forward operator are generally in agreement with CYGNSS measured ones.

Under the ideal case, the only cause of the difference between observation and model is the difference between the model wind field and the real winds. Adjustment of the model winds, to minimize this difference, is the essential principle of data assimilation. However, there are several other uncertainties that influence the performance of the forward model. These must be understood and accounted for in any practical data assimilation scheme. In our comparison, we selected cases in which the time difference between observation and model (less than 10 minutes) are small enough such that temporal change of wind field can be ignored. The HWRF analysis (Maria) and 2-hour short forecast (Gita) were found to be highly consistent with the best track data according to a verification report [66]. Using these examples, we evaluated four (4) sources of uncertainty:

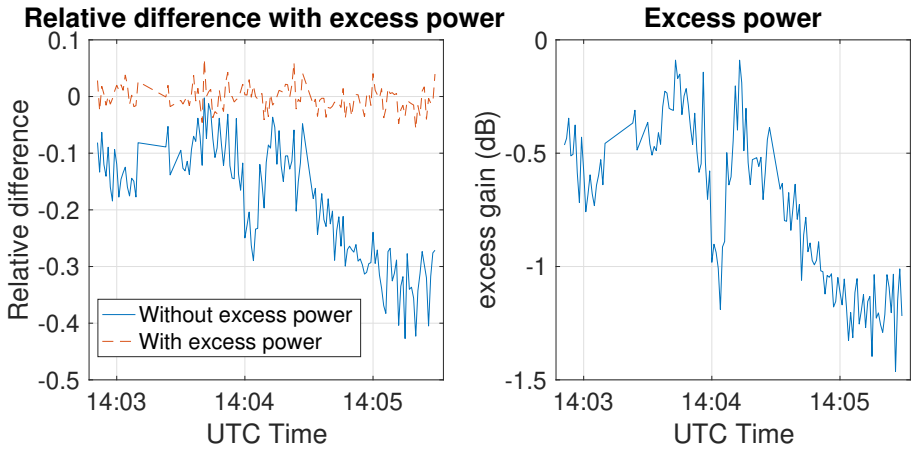
1. *Low wind speed (swell and specular reflection)* Ocean surface roughness is not only sensitive to short waves driven by the local wind but also to long waves forced by non-local swell [67], [68]. At medium or high wind speed cases, the local wind plays the dominant role and the effect of swell can be ignored. However, at very low wind speed cases ($\lessapprox 5$ m/s), the local wind is weak and the swell may be evident. In the forward model, MSS is computed as an empirical function of only the local wind speed [17], incorporating no information about swell. Ocean swell can be characterized by the significant wave height (SWH) or MSS from a wave model (i.e. WAVEWATCH III [69] or WAM [70]). Bayesian estimation has also been applied to correct for the effect of long waves using SWH [30]. Those methods need auxiliary data of the sea state in addition to the wind speed, which could complicate the forward model. At the present time, data assimilation is parameterized in terms of wind speed (or wind vector) only. Furthermore, at low wind speed, the ocean surface may be smooth enough such that coherent scattering exists and the KA-GO model would no longer be applicable [71]. These un-modeled effects are possible causes of the high relative

differences observed for low wind speeds as shown in Fig. 2.6. To avoid the effect of swell and coherent scattering, quality controls on the background wind speed (i.e., a threshold on the minimum wind speed) or the relative difference in equation (2.27) can be used to filter out the cases which may exhibit those effects.

2. *DDM power bias:* A bias is observed in both cases in Fig. 2.6, in which the relative power differences are all above zero for Maria and all below zero for Gita. One of the major sources for this bias is the inaccurate estimated GPS transmitter EIRP. For CYGNSS, the GPS EIRP is estimated using measurements from the zenith navigation antenna on each CYGNSS satellite and a ground-based GPS power monitor [40, 72]. Accuracy of the EIRP estimate can be influenced by variation in the transmitter power and calibration of the zenith antenna pattern. To evaluate this influence, the excess power gain G_0 in equation (2.11) was estimated independently. For each modeled DDM, the excess power is adjusted by fitting all effective bins in the modeled DDM against the observed DDM using a least square method. Fig. 2.8 shows the relative power differences with and without adjusting G_0 and the excess power along the track for Hurricane Maria and Cyclone Gita. Samples for wind speed at the specular point under 5 m/s have already been removed from this comparison, due to aforementioned concerns about swell and coherent scattering. The relative differences become much closer to zero and the bias is removed after adding the excess power for both cases. It shows that the excess power parameter in the forward model can be used to effectively correct the bias. Errors in the receiver antenna pattern, resulting from changes from exposure to the space environment, for example, would also introduce different power error of each DDM bin. These can be assumed to remain constant through the mission and can thus their effect on the bias can also be calibrated. In the initial studies in this paper, a maximum threshold on the relative power difference in equation (2.27) will be set for quality control.



(a)



(b)

Figure 2.8.: Relative power differences with and without adding the excess power in each modeled DDM and the excess power along the track: (a) Maria (b) Gita. © 2020 IEEE.

3. *Inaccurate specular point location:* The specular point delay and Doppler, reported by the DDMI, are very coarse, due to the simplified geometry model used for on-board open-loop tracking and the limited precision of discrete steps (0.25 chip, 500 Hz). Specular point delay and Doppler are re-computed on the ground by the CYGNSS SOC using a more accurate method of specular point calculation and a higher fidelity mean sea surface model [25]. These higher precision specular bin indices, n_τ^s, n_f^s, are obtained from the CYGNSS L1a metadata and

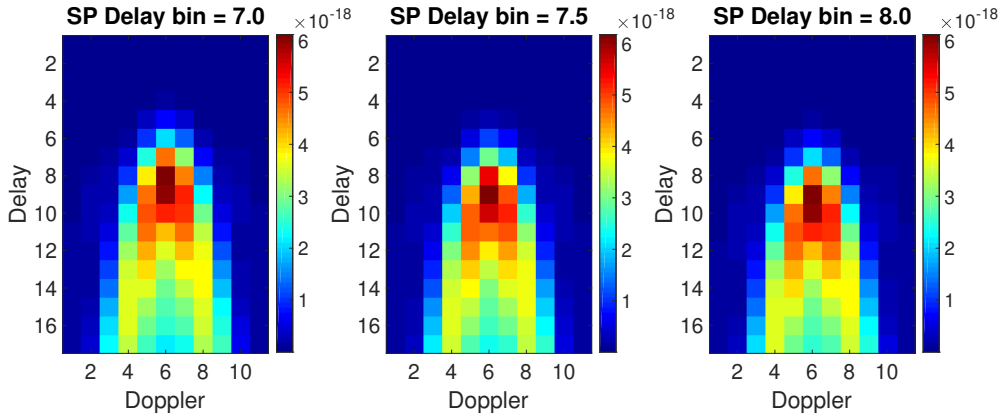


Figure 2.9.: DDMs computed by the forward model using specular delay bins of 7.0, 7.5, 8.0 (from left to right). One example of the Hurricane Maria case was selected with the specular point at 26.5°N, 72.8°W. © 2020 IEEE.

used in the forward model (2.11). However, some errors still remain as a result of errors in the sea surface height model, local height variation (e.g., increased wave height introduced by storm surge), and C/A code tracking error, resulting in some misalignment between the observed and modeled DDMs. Fig. 2.9 illustrates this effect and shows change in the modeled DDMs with displacement of the specular point delay in the step of 0.5 bin. Fig. 2.10 shows the correlation coefficients between the observed DDM and modeled DDM using different specular delay bin index. Correlation coefficient between the observed and modeled DDM can be used as quality control to filter out cases with large error in the specular point position, as this metric represents the shape similarity between the two.

4. *High wind speed:* The empirical relationship between wind speed and MSS in equation (2.6) was derived by fitting a large collection of airborne GNSS-R observations from high-elevation satellites against wind speed measurements [17]. In the very high wind speed region near the hurricane eyewall, the model should be used cautiously because of the complicated sea state, breaking of the waves or unavailability of the sea surface slope information [73]. The decrease

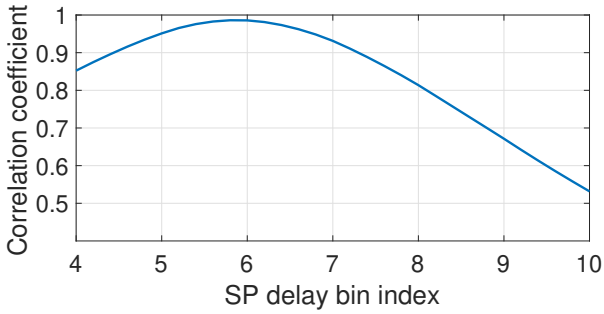


Figure 2.10.: Correlation coefficients between the observed DDM and modeled DDM using different specular delay bin index with a delay step of 0.2 bin. One example of the Hurricane Maria case was selected with the specular point at 26.5°N, 72.8°W. © 2020 IEEE.

in correlation coefficient in Fig. 2.7 when the wind speed is high could be possibly caused by this model representation error. Discarding observations with background wind speed exceeding a threshold ($\gtrapprox$ 35 m/s) can be used as an additional quality control filter.

The discussion above shows that, while the forward model accurately represents the relationship between the scattered and surface conditions over the full extent of the DDM in most cases, there are some important discrepancies which must be understood and accounted for. We propose that these effects can be mitigated through limiting the usable range of wind speeds, estimating the excess power gain, G_0, and defining quality control tests on the relative power difference and correlation coefficient between the observed and modeled DDM.

2.3.2 Validation of the Jacobian Matrix

The Jacobian matrix represents the sensitivity of each sample of the DDM to the wind speed at each surface grid point. An example of the matrix, shown in Fig. 2.11, is computed for Hurricane Maria at a time when the specular point is near the hurricane eyewall. The two dimensions of this matrix are the DDM values (organized with delay and Doppler unrolled into a vector) and surface wind speeds (organized

with latitude and longitude unrolled into a vector). Some rows and columns in the matrix are zero because they correspond to delays earlier than the specular point or later than the largest iso-delay ellipse. The scattered power in some is so small that they are dominated by noise and thus the Jacobian is effectively zero. High absolute values in Fig. 2.11 matrix represent the sensitivity of the DDM bins to wind speeds at specific locations falling within the corresponding delay and Doppler range.

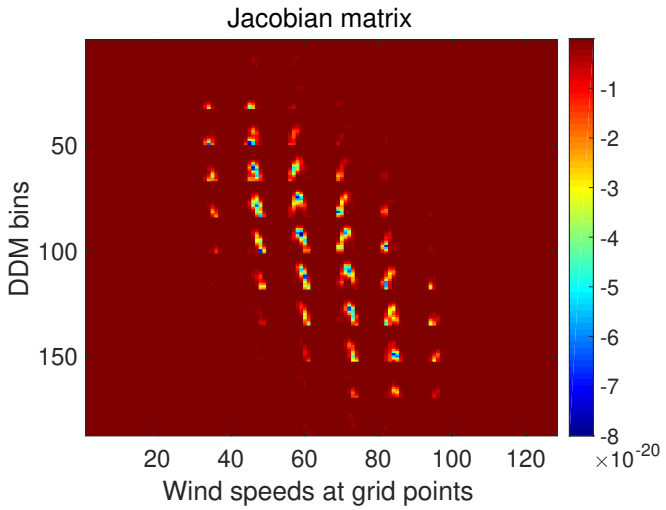
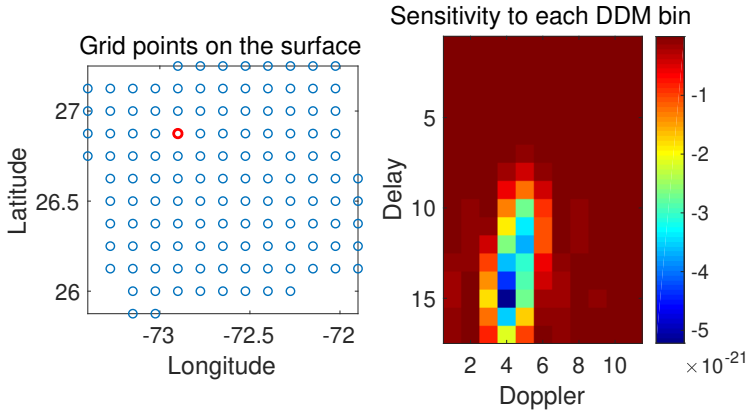
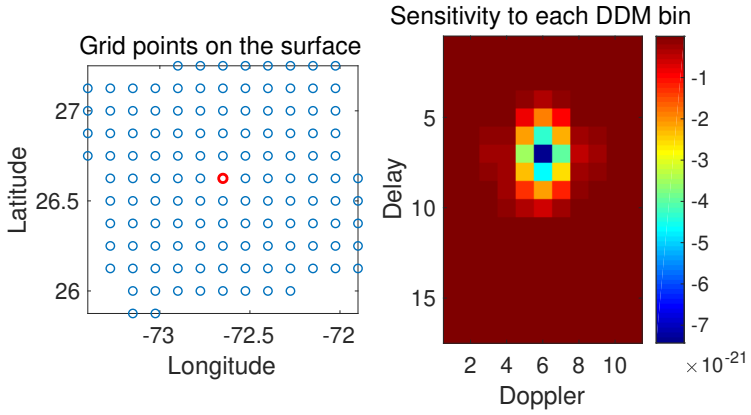


Figure 2.11.: An example of the Jacobian matrix computed from the Hurricane Maria case when the specular point is near the hurricane eyewall. The specular point is at 26.5°N, 72.8°W. © 2020 IEEE.

To visually present the matrix, a given column representing the sensitivity of all DDM samples to the surface wind speed at a specific surface point, can be expanded into a matrix in delay and Doppler coordinates. Two examples are shown in Fig. 2.12. This shows that information about one surface point is present in multiple DDM samples with some degree of independence. Similarly, a given row representing the sensitivity of a specific sample of the DDM to a region of wind speeds on the surface grid is shown in Fig. 2.13. Note that the delay-Doppler ambiguity can be observed in Fig. 2.13(b) in which two geographically separated regions on the surface both contribute to the observation made at one delay and Doppler.



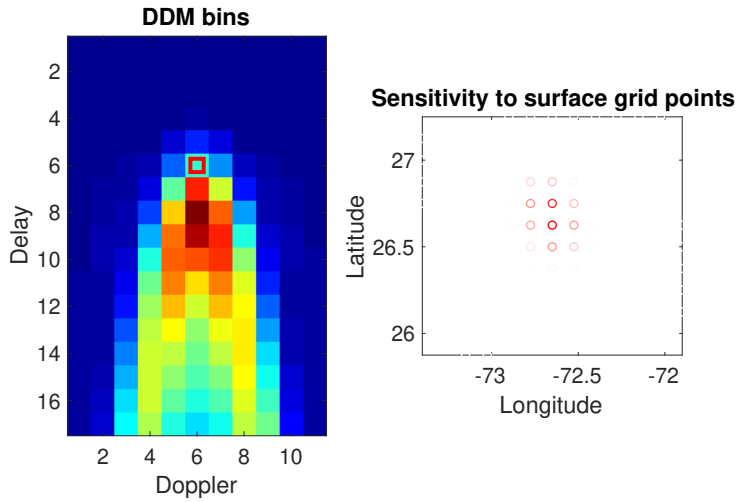
(a)



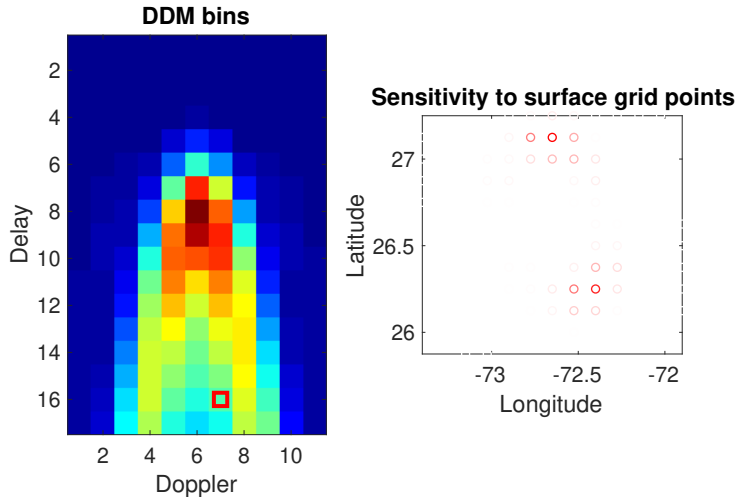
(b)

Figure 2.12.: Two cases for sensitivity of wind speed in a grid point to all DDM bins computed from the Jacobian matrix in Fig. 2.11. In each case, a grid point (red circle) of the wind field is selected in the left plot. The sensitivity of the power in all DDM bins with respect to wind speed of the selected grid point is shown in the right plot. © 2020 IEEE.

The Jacobian matrix computed by equation (2.26) was derived in a closed form by differentiating the forward operator, assuming all geometries, power and wind speed are homogenous within each 10 km cell, rather than the 1 km grid used in the forward operator. This linearization and the 10-km approximation may induce errors in the computation of the Jacobian matrix. To validate the Jacobian matrix, the matrix is computed by the finite difference of the forward model (at 1 km resolution)



(a)



(b)

Figure 2.13.: Two cases for sensitivity of a DDM bin power to wind speeds on all grid points computed from the Jacobian matrix in Fig. 2.11. In each case, a DDM bin (red square) is selected in the left plot; The sensitivity of wind speed in the wind field grid with respect to the power of the selected DDM bin is shown in the right plot, where the saturation of the red color represents the absolute value of the sensitivity. © 2020 IEEE.

using a step of 0.0001 m/s. Fig. 2.14 shows a comparison of Jacobian matrices by finite difference vs. that from the analytical form. It can be seen that they have similar patterns. The relative difference is shown in Fig. 2.14(c). After discarding

the meaningless zero values in the matrix, the relative error is about 0.19 in average and the correlation coefficient between the two matrices is 0.92. In the same way, Jacobian matrices of the track of 100 DDMs for Hurricane Maria were validated by the finite difference with an average relative error is 0.20 and average correlation coefficient is 0.90 for the track. Although the analytical form may exhibit some error due to the larger grid size, it is much efficient with approximately one-tenth of the computation time of the finite difference method.

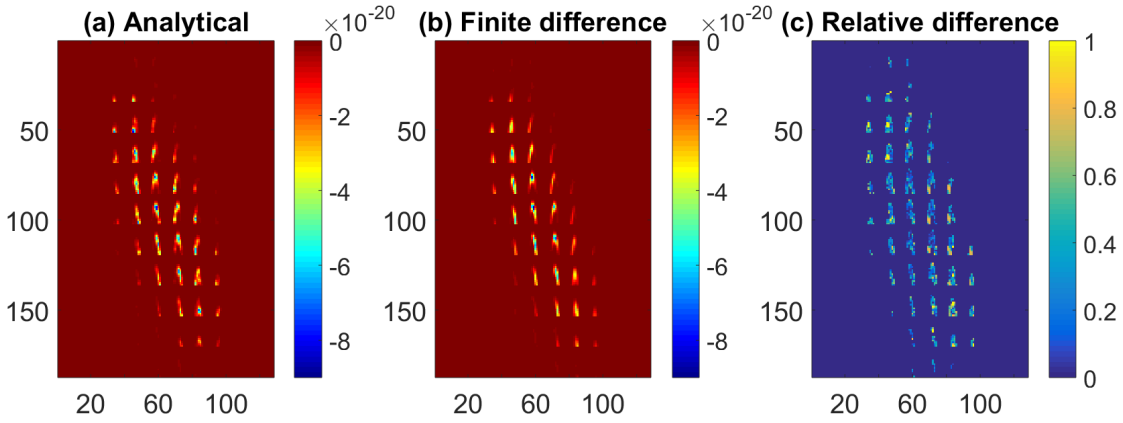


Figure 2.14.: Comparison of the Jacobian matrices computed by finite difference and analytical form for the Jacobian matrix in Fig. 2.11. The three plots are the analytical version (a), finite difference version (b) and the relative difference between the two (c). © 2020 IEEE.

2.4 Impact of the Specular Point Position Error

An inaccurate specular point position in the observed DDM can cause a shift in the DDM and affect the accuracy of wind speed retrieval algorithm [63]. In this section, the impact of the specular position error on the DDM assimilation is studied.

The CYGNSS onboard Delay Doppler Mapping Instrument (DDMI) open-loop tracking uses a coarse targeting algorithm to estimate a specular point with over 3 chips of error and produces a 128×20 raw DDM. The resolution is 0.25 chip in delay and 500 Hz in Doppler. The center bin of the 128×20 DDM coincides with the

DDMI-estimated specular point delay and Doppler. Then the flight software (FSW) compression algorithm uses a low-pass-filter to compress (truncate) the raw DDM into a 17×11 DDM to meet the downlink requirement [24]. The science operation center (SOC) also calculates a precise specular point (truth) based on the geoid and provides the precise specular bin index in decimal in the Level 1 data. The details in the delay dimension are shown below. It goes in the same way for the Doppler dimension.

The DDM is a correlation between the received signal and the GPS transmitter signal over a range of delays and Dopplers as shown in equations (1.1)–(1.3). From the downlinked 17×11 DDM, the FSW provides a coarse specular delay τ_{n_0} (time delay from transmitter to specular point plus from specular point to receiver) and a delay index n_0 (integer). Let the 17 delay coordinates in the DDM are

$$\tau = [\tau_0, \tau_1, ..., \tau_{n_0}, ..., \tau_{16}]. \tag{2.29}$$

The SOC also calculates a precise specular point delay τ_s using the precise specular point location. The precise specular bin index n_s is calculate as

$$n_s = n_0 + (\tau_s - \tau_{n_0})/(0.25\tau_c) \tag{2.30}$$

where τ_c is the time duration of the GPS C/A code chip (977 ns). Note that n_s s a decimal, not an integer. n_s is stored in the CYGNSS L1 data.

In the data assimilation, the forward model tries to compute a modeled DDM in the same delay coordinates as equation 2.29. The computation of DDM in the forward model is described below.

The forward model first uses the precise positions and velocities of the transmitter, receiver and specular point in the CYGNSS Level 1 data to calculate the specular delay τ_s. Then a DDM with 81 delay bins at resolution of 0.05 chip is calculated with the specular point at bin $\lfloor n_s \times 5 \rceil$, where $\lfloor \rceil$ stands for rounding the number to the nearest integer. The 17 delays in the 17×11 DDM is calculated by selecting

bins at $5n$ where $n = 0, 1, ..., 16$. For example, if $n_s = 5.43$, the specular bin is at bin $\lfloor 5.43 \times 5 \rceil = 27$; The 17 delays are at bin 0, 5, ..., 25, 30, ... ,75, 80.

Now it is possible to calculate the delay shift between the observed DDM and the simulated DDM. Since the delay resolution is constant, the shift at any delay bin is the shift of the whole DDM. So at the delay bin n_0 according to equation 2.30, the delay in the observed DDM, τ_{n_0} is

$$\tau_{n_0} = \tau_s + 0.25\tau_c(n_0 - n_s). \tag{2.31}$$

In the modeled DDM, the delay at bin n_0, τ'_{n_0} is

$$\tau'_{n_0} = \tau_s + 0.05\tau_c(5n_0 - \lfloor 5n_s \rceil). \tag{2.32}$$

From equation (2.31)(2.32), the shift in delay between the observed DDM and modeled DDM, $\Delta\tau$ is

$$\Delta_\tau = \tau_{n_0} - \tau'_{n_0} = 0.05\tau_c(\lfloor 5n_s \rceil - 5n_s). \tag{2.33}$$

From equation (2.33), the relationship between the delay shift $\Delta\tau$ and n_s is shown in Figure 2.15.

It shows that the magnitude of the delay shift is less than 0.025 chip or 0.1 delay bin, which is negligible. In the same way, since the sampling resolution of the Doppler frequency in the forward model is 25 Hz, the shift in Doppler frequency Δf is

$$\Delta f = 25(\lfloor 20n_s^f \rceil - 20n_s^f) \tag{2.34}$$

which has a magnitude of less than 12.5 Hz.

The fact is that although the observed DDM is shifted by the inaccurate specular position estimated by the DDMI, when the forward model uses the precise specular position, decimal specular bin index and high sampling resolutions in delay and Doppler, it can also compute a shifted DDM in the similar delay/Doppler coordinates with errors less than 0.025 chip in delay and 25 Hz in Doppler, which can be ignored.

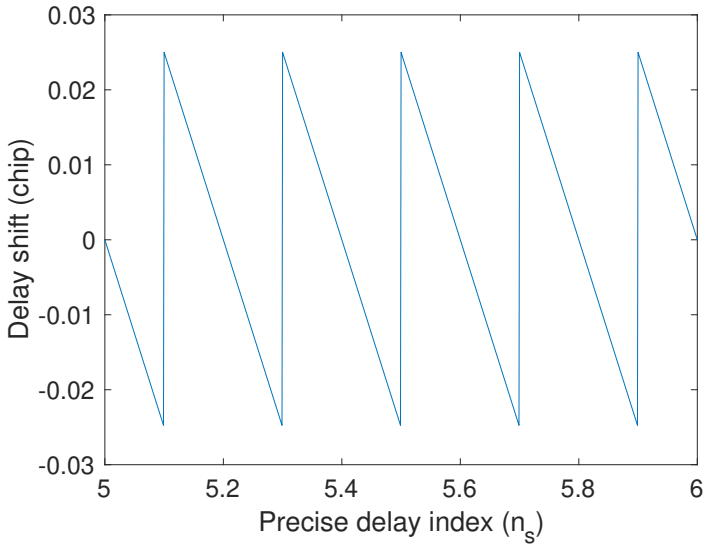


Figure 2.15.: The shift in delay between the observed DDM and modeled DDM versus the precise delay bin index from equation (2.33). The precise delay bin index is selected to between two adjacent integers.

2.5 The Extended Kalman Filter Wind Speed Retrieval Algorithm

As another application of the forward model, a wind speed retrieval algorithm is briefly summarized in this section. The details can be found in [1].

As described in section 1.3, the CYGNSS baseline wind retrieval algorithm uses DDM pixels in only a 3×5 box from the measured 17×11 DDM due to the 25-km resolution limitation, discarding more than 90% of the data. The motivation for this new approach is to utilize the additional information in the remaining DDM pixels outside the region around the specular point, while continuing to meet the 25-km resolution requirement. The retrieval of wind speeds away from the specular point, within a swath determined by the maximum delay of the DDM will also be demonstrated.

This algorithm makes use of the receiver motion to provide multiple looks at each point on the ocean surface through the overlapping areas of subsequent DDMs. Previously published methods incorporating multiple looks at the same ocean surface

have included a stare processing method [74], [75] and multiple-antenna techniques [76], [77]. In the new results presented here, an extended Kalman filter (EKF) will be applied in a very general algorithm to combine sequential DDMs.

Several textbooks [78], [79] provide derivations of the classical EKF equations. Implementing the EKF generally requires a state vector, an observation vector, an observation model, an initial state estimate, a state transition model, observation covariance matrices and a covariance matrix for the initial state estimate. In this approach, the observation vector is the CYGNSS L1 DDM power and a batch of 7 DDMs was found to work best. The state vector is a set of 10-km gridded wind speeds on an area of the ocean surface defined by the largest iso-range ellipse of the 4th DDM in the sequence. The observation model is the forward model described in section 2.2 except the state is modified to be 10 km grid on the SURF frame (vs. latitude and longitude coordinates). The initial state estimate can be the CYGNSS L2 wind speeds or computed by any single-point observable. No state transition model is used as we assume the observed wind field does not change within 7 consecutive DDMs. Covariance matrices of the initial state and observations are all set empirically. Figure 2.16 shows the overlapping of the iso-range ellipses for seven consecutive DDM measurements and the state vector (10-km grid points on the SURF frame of the 4th DDM).

Detailed implementations and results on synthetic CYGNSS data can be found in [1]. Although specifications of the CYGNSS mission will be used in this section, the EKF algorithm should be easily adaptable to other GNSS-R satellite missions that provide a calibrated DDM along with associated meta-data such as direct GNSS signal power, orbit geometry, and antenna gain patterns.

2.6 Summary

A generalized forward model for assimilating GNSS-R DDMs into numerical weather models has been presented. Assimilation of full DDMs could incorporate more in-

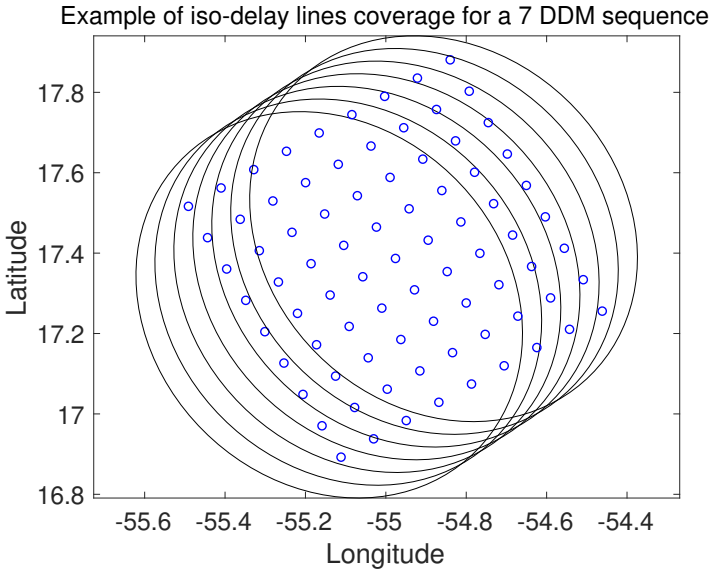


Figure 2.16.: Ocean surface coverage showing overlapping of the iso-range ellipses for 7 consecutive measurements with incidence angle of 35.5°. The blue circles are the 10-km grid points on the SURF frame of the 4th DDM.

formation into forecast models than what is provided by baseline CYGNSS Level 2 wind products. The forward model takes in a small set of satellite parameters (e.g. CYGNSS Level 1a products and antenna patterns) and a gridded wind field. Its output is the modeled DDM and Jacobian matrix.

Performance of the forward model was assessed on two tracks of CYGNSS data from Hurricane Maria and Cyclone Gita by comparing observed and computed DDMs. Results of this comparison show agreement with relative differences under 30% and correlation coefficients larger than 90%. Limitations of the forward model have been presented, including effects of swell, observation bias, specular positioning error and limitations of the KA-GO theory especially at high wind speed cases. Quality controls on wind speed, relative power difference and correlation coefficient between the observed and modeled DDM were defined based upon these deficiencies. An analytical Jacobian matrix was validated by comparison with one computed by finite differences. The average relative error between the analytical version and a Jacobian

computed by finite difference is 20%. The average correlation coefficient between the two is 0.90.

Although the model was developed for CYGNSS, it can be applied to any other spaceborne GNSS-R missions which provide DDM measurements. As long as the satellite geometries, power parameters and metadata listed in section 2.1 are also provided, the DDMs can be assimilated into NWP models by the forward model.

An EKF wind speed retrieval algorithm is developed based on the forward model. This method utilizes the multi-look feature of DDM observations and has the potential to perform better than the CYGNSS baseline GMF-based wind speed retrieval algorithm.

3. DATA ASSIMILATION METHOD

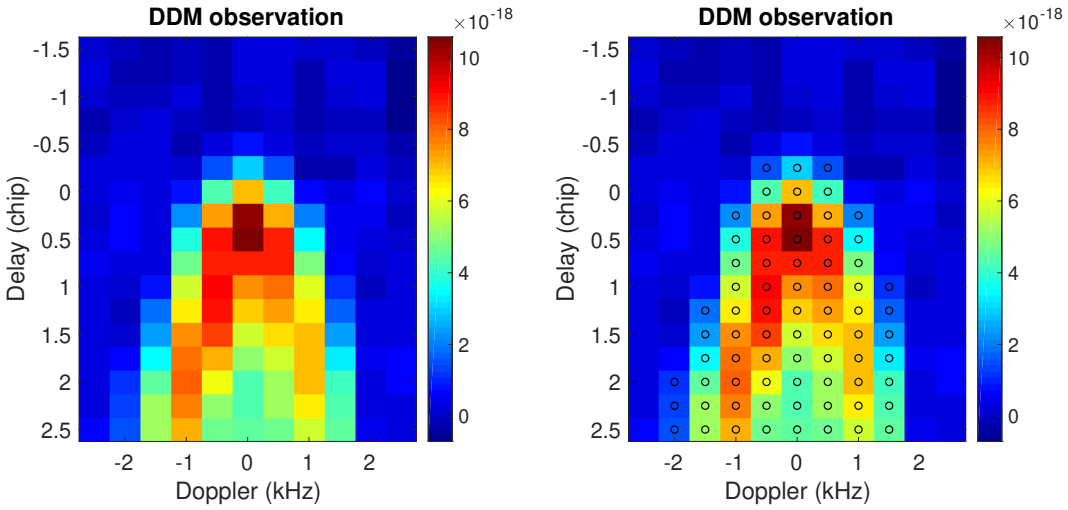
3.1 DDM Observation

As described in section 1.2, the GNSS-R DDM observation is the average power of the cross-correlation between the reflected signal and a replica of the transmitted signal over a range of delays, τ, and Doppler frequencies, f. The CYGNSS L1 DDM power is provided at 17 discrete delays at increments of 0.25 GPS C/A code chip (244 ns) and 11 discrete Doppler frequencies at increments of 500 Hz. In the remaining part of the paper, the word "DDM" will be referring to the DDM power (not DDM BRCS) unless specified otherwise.

Due to the geometry and delay/Doppler range selected by the receiver, some samples of the DDM contain little or no information about the surface wind speed. Those observations are not useful for DA and need to be discarded. In an empirical method to select informative DDM samples, samples with power magnitude larger than 10% of the peak DDM power are selected for use in DA. Figure 3.1(a) shows an example of the CYGNSS L1 DDM. The informative samples of the same DDM are shown in Figure 3.1(b). All K of the informative samples of the DDM at one time, t, are grouped into a vector

$$\boldsymbol{Y}(t) = \begin{bmatrix} Y(t, \tau_1, f_1) \\ Y(t, \tau_2, f_2) \\ \vdots \\ Y(t, \tau_K, f_K) \end{bmatrix} \tag{3.1}$$

which will be used as the observation in DA.



(a) (b)

Figure 3.1.: An example of the CYGNSS Level 1 17×11 DDM measurement (a) and DDM informative samples used in DA shown as black circles (b). Units in Watts.

3.2 The Variational Analysis Method

This study uses a two-dimensional VAM, based on the surface wind vector field, to assimilate DDMs. This approach was first introduced in [80–82] to resolve scatterometer wind ambiguities and then applied to assimilate satellite wind observations from a large-scale dataset in [83]. It was used to add wind direction information to the CYGNSS retrieved wind speed in an observing system simulation experiment (OSSE) [51]. It was applied to demonstrate DDM assimilation using a few examples in [2].

VAM finds the minimum of the objective cost function J,

$$J(\boldsymbol{\beta}) = J_b(\boldsymbol{\beta}) + J_{ddm}(\boldsymbol{\beta}) + J_c(\boldsymbol{\beta}), \tag{3.2}$$

where $\boldsymbol{\beta} = [u_1, u_2, ..., u_N, v_1, v_2, ..., v_N]$ is a vector containing the analysis wind field in u (East) and v (North) components. N is the number of grid points in the wind

field. $J_b(\boldsymbol{\beta})$ is the wind background term presenting the misfit between the analysis and the background,

$$J_b(\boldsymbol{\beta}) = \lambda_b \frac{1}{\sigma_\beta^2} \left(\boldsymbol{\beta} - \boldsymbol{\beta}^b\right)^T \left(\boldsymbol{\beta} - \boldsymbol{\beta}^b\right) \tag{3.3}$$

where $\boldsymbol{\beta}^b$ is the background wind field vector and σ_β^2 is the variance of each wind vector component in the background. $J_{ddm}(\boldsymbol{\beta})$ is the DDM observation term presenting the misfit between the analysis and the observation,

$$J_{ddm}(\boldsymbol{\beta}) = \lambda_{ddm}(\boldsymbol{h}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) - \boldsymbol{Y})^T \boldsymbol{R}^{-1}(\boldsymbol{h}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) - \boldsymbol{Y}) \tag{3.4}$$

in which $\boldsymbol{Y}$ is the DDM observation and $\boldsymbol{R}$ is the DDM observation error covariance matrix. $\boldsymbol{h}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \boldsymbol{h}(\boldsymbol{x})$ is the vector of DDM samples computed by the forward operator in section 2.2.1 using the analysis wind speed field,

$$\boldsymbol{x} = \left[\sqrt{u_1^2 + v_1^2}, \sqrt{u_2^2 + v_2^2}, ..., \sqrt{u_N^2 + v_N^2}\right]^T. \tag{3.5}$$

$J_c(\boldsymbol{\beta})$ is the smoothness and constraint term including Laplacian, divergence, and vorticity of the difference between the background and analysis,

$$J_c(\boldsymbol{\beta}) = \lambda_{lap} J_{lap}(\boldsymbol{\beta}) + \lambda_{div} J_{div}(\boldsymbol{\beta}) + \lambda_{vor} J_{vor}(\boldsymbol{\beta}) \tag{3.6}$$

where $J_{lap}(\boldsymbol{\beta})$, $J_{div}(\boldsymbol{\beta})$, and $J_{vor}(\boldsymbol{\beta})$ are computed as described in [82]. λ_b, λ_{ddm}, λ_{lap}, λ_{div} and λ_{vor} are weighting factors of each term.

In the term J_b, the background error is characterized by a single constant value, σ_β^2. The background error correlations are characterized by the constraint term J_c, which is derived from the Navier-Stokes equations for viscous fluid motion. The combination of J_b and J_c acts the same as a background covariance matrix in the traditional variational DA algorithm, as explained in [82]. Rather than assimilating all DDMs in one cost function, the DDMs on each CYGNSS specular point track are assimilated sequentially to reduce the computation and memory cost. One DDM will

be assimilated at a time and the analysis wind field will be updated after processing each DDM until all observations within the DA cycle have been assimilated. The observation error covariance matrix, $\boldsymbol{R}$, presents the errors and correlations of all DDM samples at the same time. Characterization of the matrix $\boldsymbol{R}$ will be given in Chapter 4. Tuning the background and observation weights λ_b and λ_{ddm} can correct the estimated errors of the background and observation in the VAM as the estimated errors are usually not accurate due to lack of enough information. The constraint weights λ_{lap}, λ_{div} and λ_{vor} should be large enough to correctly shape the error correlations of the background wind field. They are set to ensure the influence of the observations spreading out to the same scale of the background effective model resolution. Specification of all the weight values in the experiment will be given in section 5.3.

To minimize the cost function $J(\boldsymbol{\beta})$, the gradient has to be calculated,

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial J_b(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \frac{\partial J_{ddm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \frac{\partial J_c(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \tag{3.7}$$

specially,

$$\frac{\partial J_{ddm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2\lambda_{ddm}\boldsymbol{H}_{\beta}^{T}\boldsymbol{R}^{-1}(\boldsymbol{h}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) - \boldsymbol{y}) \tag{3.8}$$

where $\boldsymbol{H}_{\beta}$ is the partial derivative matrix respect to u, v components and can be related to the Jacobian matrix $\boldsymbol{H}$ computed in section 2.2.2 by

$$\boldsymbol{H}_{\beta} = \begin{bmatrix} \boldsymbol{H}_{1,*} \odot L_u & \boldsymbol{H}_{1,*} \odot L_v \\ \boldsymbol{H}_{2,*} \odot L_u & \boldsymbol{H}_{2,*} \odot L_v \\ \vdots & \vdots \\ \boldsymbol{H}_{P,*} \odot L_u & \boldsymbol{H}_{P,*} \odot L_v \end{bmatrix} \tag{3.9}$$

$\boldsymbol{H}_{i,*}$ is the i-th row of the $\boldsymbol{H}$ matrix. Row vectors

$$L_u = \left[\frac{u_1}{x_1}, \frac{u_2}{x_2}, ..., \frac{u_N}{x_N}\right] \tag{3.10}$$

and

$$L_v = \left[\frac{v_1}{x_1}, \frac{v_2}{x_2}, ..., \frac{v_N}{x_N}\right] \tag{3.11}$$

are partial derivatives of $\boldsymbol{x}$ respect to the u and v components. The operator $\odot$ represents the element-wise multiplication between two vectors and $P = N_\tau N_f$ is the number of DDM bins.

The initial value of the analysis $\boldsymbol{\beta}$ is set to equal to the background. The minimization problem is solved by a Quasi-Newton algorithm [84].

3.3 Bias Correction

It is crucial to have unbiased observations in order to obtain the Best Linear Unbiased Estimator (BLUE) in DA [41]. Bias can arise in the measurement or the forward operator and should be removed before assimilating the observations. The DDM forward operator requires an estimate of the transmitter EIRP and the receiver antenna patterns for each CYGNSS satellite. The CYGNSS mission uses a ground-based power monitor to estimate the EIRP, which is provided in the Level 1 data [85]. Receiver antenna patterns are estimated by pre-launch measurements and on-orbit corrections [25]. These patterns were made available to us by the CYGNSS project and are distributed as part of the forward model code [58]. Previous studies have found bias in the CYGNSS observations which largely resides in the estimated transmitter EIRP with some contribution from the receiver antenna patterns [40, 86]. In order to remove this bias, we assume that the GPS transmitter EIRP remains constant for all observations along the same CYGNSS specular point track. This is a reasonable assumption, given that the duration of a track is generally less than 20 minutes. This suggests a "track-wise" DDM bias correction scheme, similar to that used by [87] for correcting bias on the retrieved wind speed. In our DA approach, however, a bias correction will be applied to the DDM power.

Our basic assumption is that the background wind field from a global NWP model (e.g., ECMWF) is globally unbiased [88]. Thus, comparing the average of a large

sample of measurements against model predictions from a background reference can be used to correct the observation bias. In this scheme, DDMs on a continuous specular point track formed by one specific pair of GPS transmitter and CYGNSS receiver are first identified. Both the transmitter EIRP and uncertainty in the receiver antenna gain patterns would be multiplicative error sources. Therefore, a scaling term is computed as the mean proportion between the M measured DDMs and the corresponding modeled DDM computed from the background along the specular point track.

$$\Phi = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{K_m}\sum_{i=1}^{K_m}\frac{Y_i(t_m)}{h_i(\boldsymbol{x}, t_m)} \tag{3.12}$$

where t_m is the time of the m-th DDM; K_m is the number of informative samples of the m-th DDM; $h_i(\boldsymbol{x}, t_m)$ is the power of the i-th modeled DDM sample at time t_m, computed from the background wind field using the forward operator.

When assimilating DDMs on the track, each modeled DDM from the forward model is multiplied by the scaling term Φ, such that the cost function (3.6) becomes

$$J_o(\boldsymbol{x}) = \lambda_{ddm}(\Phi\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{Y})^T\boldsymbol{R}^{-1}(\Phi\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{Y}) \tag{3.13}$$

This method requires less memory and computation compared to computing the bias using all DDMs on the track at a time by the least square method and should give similar results.

3.4 DDM Quality Control

The following quality control (QC) tests are applied to filter CYGNSS Level 1 DDM observations before DA.

1) The netCDF variable "quality_flags" values in the CYGNSS L1 data are required to be zero. This discards cases in which the observation is over or close to land, the spacecraft has attitude rotation larger than 1°, the transmitter power has a high uncertainty or there are some calibration issues.

2) All data with signal-to-noise ratio (SNR) less than 3 dB are discarded. Small SNR indicates high noise power, making it difficult to extract informative DDM samples.

3) All data with incidence angle larger than 60° are discarded. DDMs observed under large incidence angle can have a glistening zone larger than 120 km × 120 km, which cannot be modeled accurately by the forward operator.

4) All data with background wind speed at the specular point less than 2 m/s or larger than 35 m/s are discarded. The swell at very low wind speed cases and the complicated sea state at very high wind speed cases cannot be modeled well by the forward operator described in section 2.3.1. The reduced sensitivity of the DDM observable to high wind speed is also well known.

5) Relative power difference and correlation coefficient between the observed DDM and modeled DDM from the background are used to identify additional observation data quality issues and reduce model representation errors. They are discussed in detail in section 2.3.1. Data with relative power difference larger than 100% and correlation coefficient less than 0.9 are discarded.

The QC tests are summarized in Table 3.1.

Table 3.1.: QC tests for the assimilation of CYGNSS DDMs.

Control Item	Value
CYGNSS L1 "quality_flags" variable	0
SNR	> 3 dB
Incidence angle	$< 60°$
Wind speed at specular points	2–35 m/s
Relative power difference	$< 100\%$
Correlation coefficient	> 0.9

3.5 Impact of the Delay-Doppler Ambiguity

It is well known that each iso-delay ellipse and iso-Doppler line has two intersections, which is called delay-Doppler ambiguity. Figure 3.2 and 3.3 show two cases of iso-delay and iso-Doppler lines in the 120 km × 120 km specular frame with incidence angle of 35° and 61°. A larger incidence angle can result in a larger spreading of the iso-delay and iso-Doppler lines.

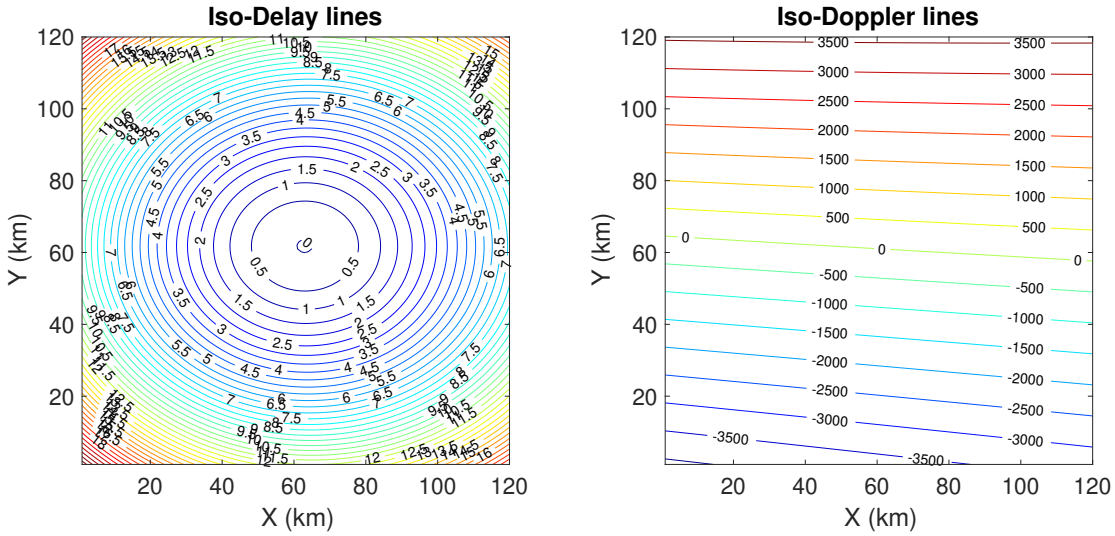


Figure 3.2.: Iso-delay and iso-Doppler lines in the 120 km × 120 km specular frame with incidence angle of 35°. The number on each line stands for the delay (chip) and Doppler (Hz).

In the CYGNSS DDM, the resolution is 0.25 chip in delay and 500 Hz in Doppler. Each DDM bin (τ, f) contains the total scattering power of the area with delay $[\tau - 0.125, \tau + 0.125]$ chips and Doppler $[f - 250, f + 250Hz]$. Figure 3.4 and 3.5 show the physical scattering area of one DDM bin in two cases. Note that the effective scattering area is larger than this because of the power spreading effect from the Woodward ambiguity function [89].

To simplify the problem, we just use the physical scattering area. We can quantify the ambiguity by defining the "ambiguity distance" which stands for the largest distance between two points in the scattering area of one bin. The ambiguity distance

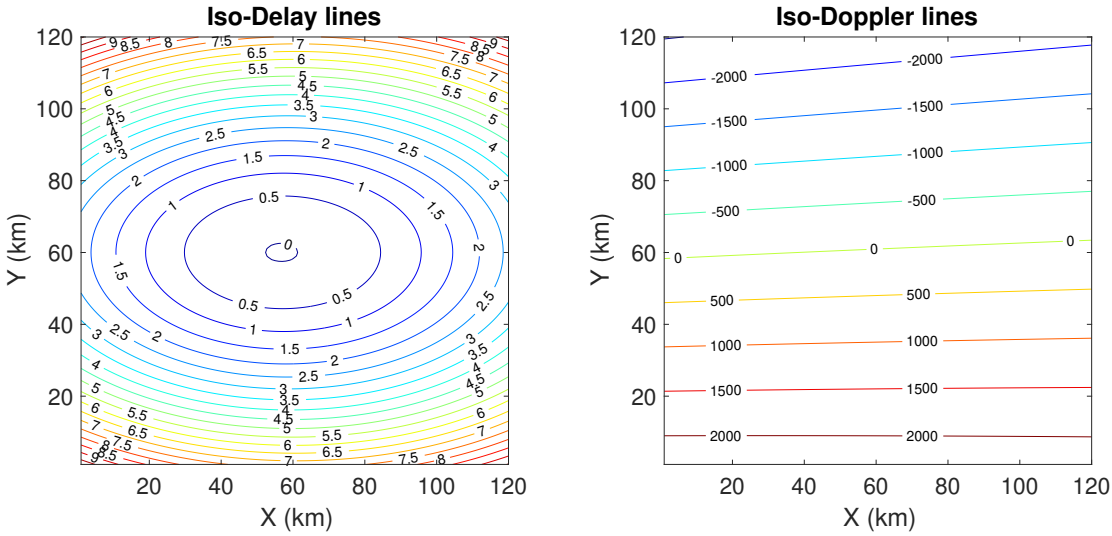


Figure 3.3.: Iso-delay and iso-Doppler lines in the 120 km × 120 km specular frame with incidence angle of 61°. The number on each line stands for the delay (chip) and Doppler (Hz).

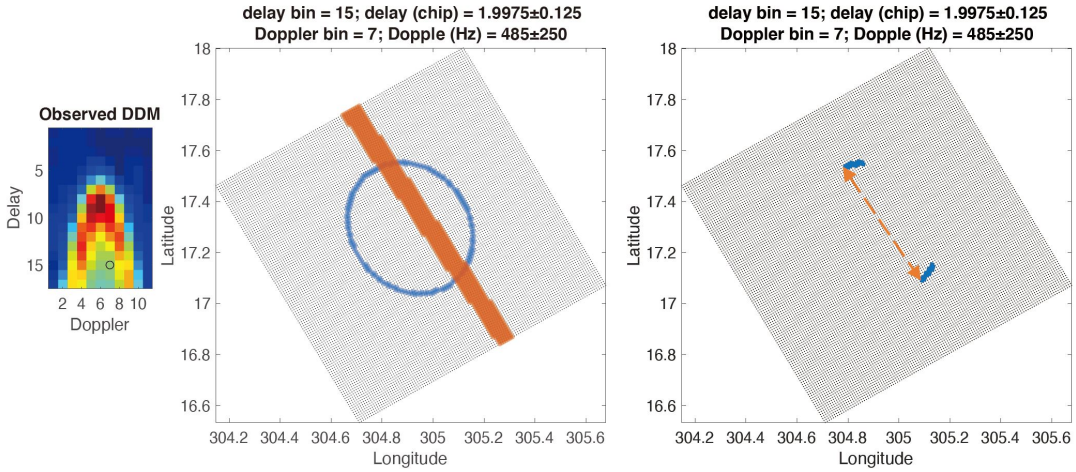


Figure 3.4.: Illustration of the scattering area of one DDM bin with incidence angle of 35°: DDM bin (left), scattering area of the corresponding delay and Doppler (middle), scattering area of the DDM bin (right). Black dots are grid points on the specular frame. The distance of the orange dash line stands for the ambiguous distance.

is shown in the right subfigure in Figure 3.4 and 3.5 in orange lines. The largest ambiguity distance can be 120 km when the incidence angle is less than 60°. In the

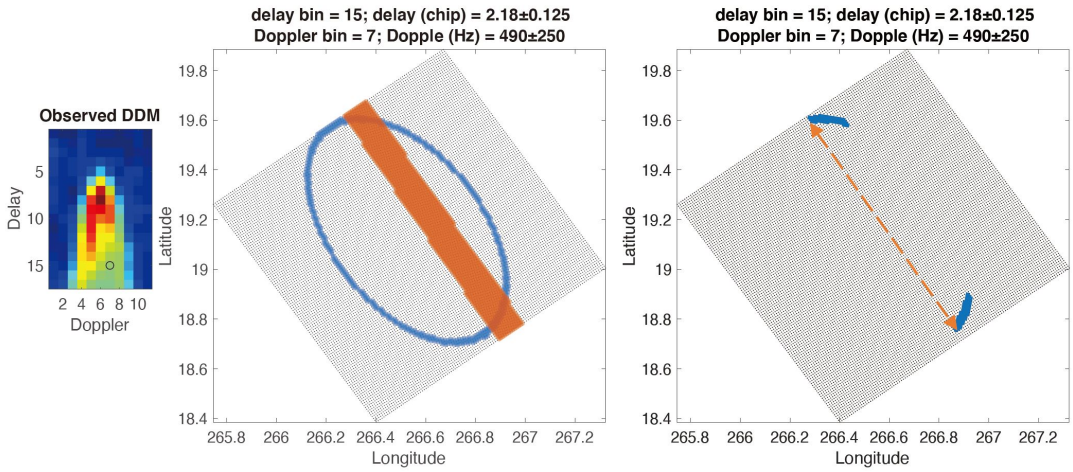
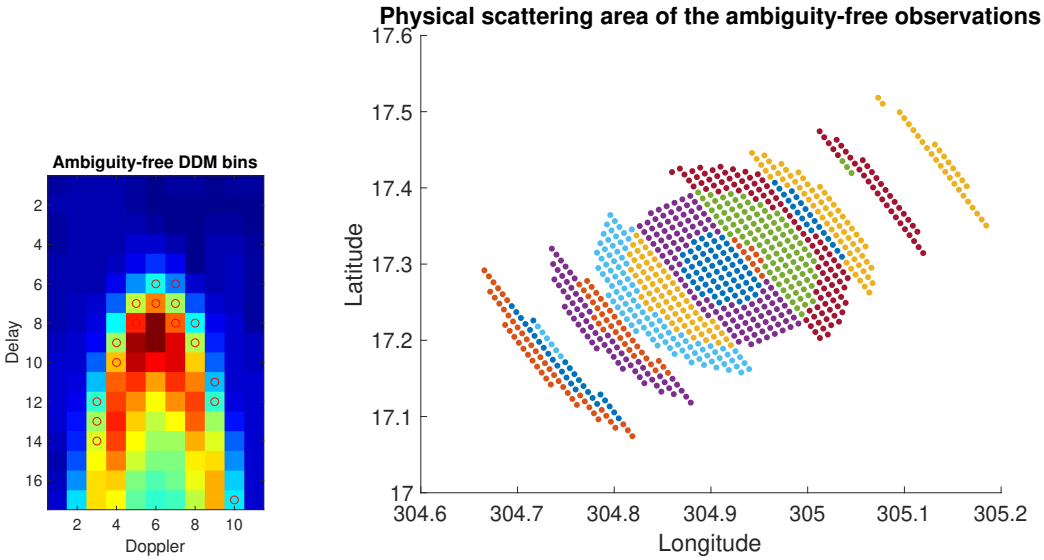


Figure 3.5.: Illustration of the scattering area of one DDM bin with incidence angle of 61°: DDM bin (left), scattering area of the corresponding delay and Doppler (middle), scattering area of the DDM bin (right). Black dots are grid points on the specular frame. The distance of the orange dash line stands for the ambiguous distance.

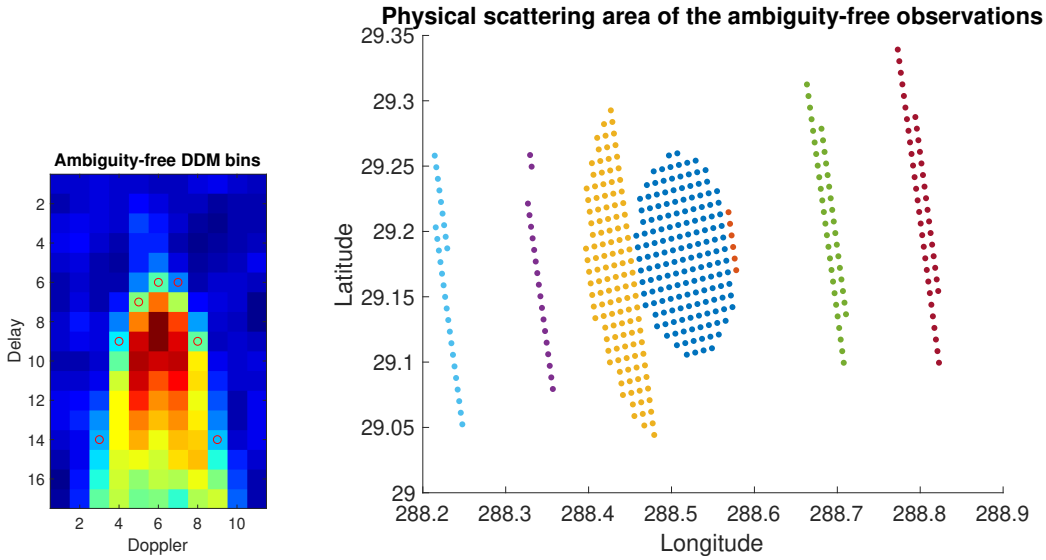
assimilation of DDMs, a threshold on the ambiguity distance can be set to select "ambiguity-free" bins depending on the wind variability scale of the background. In the assimilation of the global weather model, this is usually not necessary since the effective resolution of the background (e.g., 150 km for ECMWF) is large enough that the constraint terms in the VAM should smooth the result over the largest ambiguous distance. In the case of assimilation of in a regional or hurricane model [90], especially in the case of larger wind variability (e.g., near the hurricane eye), this method has the potential to select the ambiguity-free observations to retrieve small-scale wind patterns.

Figure 3.6 and 3.7 show the ambiguity-free bins and corresponding physical scattering area in two cases with different incidence angles (35° and 61°) using 30 km as the criterion of the ambiguous distance. It can be seen that under the same criterion, the number of ambiguity-free DDM observations at high incidence angle is less than the one at low incidence angle.



(a) (b)

Figure 3.6.: Ambiguity-free DDM bins (left) and the corresponding physical scattering area (right) using 30 km as the ambiguity distance. The incidence angle of this case is 35°. Different color stands for the scattering area of different DDM bin.



(a) (b)

Figure 3.7.: Ambiguity-free DDM bins (left) and the corresponding physical scattering area (right) using 30 km as the ambiguity distance. The incidence angle of this case is 61°. Different color stands for the scattering area of different DDM bin.

4. ERROR CHARACTERIZATION OF DDM OBSERVATIONS

4.1 Introduction

An accurate DDM observation error covariance matrix, $\boldsymbol{R}$, is required for optimal estimation in data assimilation. Observation errors usually include measurement error (error related to the instrument and measurement technique) and representation error (error related to the forward operator, unresolved scales and pre-processing) [50]. This section will only focus on the statistics of the measurement error as the representation error (especially in the estimated transmitter EIRP and receiver antenna gain) can be largely alleviated by the bias correction and QC in Chapter 3. The weights λ_{ddm}, λ_{lap}, λ_{div} and λ_{vor} in (3.3) to (3.6) can additionally be varied to adjust the relative importance of new observations vs. the background wind field and constraints.

As stated earlier, the VAM assimilates one DDM each time and the observation error covariance matrix, $\boldsymbol{R}$, represents the errors and correlations of all K informative samples in one measured DDM

$$\boldsymbol{R} = \mathrm{E}\left\{ \left(\boldsymbol{Y} - \mathrm{E}\{\boldsymbol{Y}\}\right) \left(\boldsymbol{Y} - \mathrm{E}\{\boldsymbol{Y}\}\right)^T \right\} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1K} \\ \vdots & \ddots & \vdots \\ \sigma_{K1} & \cdots & \sigma_K^2 \end{bmatrix}. \tag{4.1}$$

The observation $\boldsymbol{Y}$ is a vector assumed to follow a Gaussian distribution by the central limit theorem as it is an average value over a large number, N, by equation (1.1) and (1.3).

Measurement error is assumed to come from both background noise and speckle noise. Background noise includes thermal emission from the ocean, correlation of the signal with that from other GNSS transmitters, and receiver thermal noise [91].

In this study, the background noise is assumed to be stationary white Gaussian, as the impact of the correlation from ambient signals is negligible, as discussed in [91]. Speckle is the result of distructive and constructive interference of random scattered signals during the coherent integration time. The background noise is additive while the speckle noise is multiplicative [92]. In previous studies, analytical models for second order statistics of the DDM complex voltage signal in the delay dimension, $X(t, \tau, 0)$, were derived by considering both thermal noise and speckle [93–96]. A detailed analytical model of the covariance matrix of the averaged DDM power in the delay dimension was derived and validated using actual data [97]. Analytical models, however, have practical limitations for direct use in DDM assimilation. First, those models require knowledge of the thermal noise statistics (equivalent thermal noise temperature) which is not estimated accurately for the CYGNSS mission. Second, present models only consider the correlations between measurements at different delays, while the correlations in the Doppler dimension and between the delay and Doppler are not characterized. Finally, analytical models require computation of a surface integral and convolution with the Woodward ambiguity function, which is computationally expensive and thus not practical for large scale DA. Another approach often used in NWP applications is to compute the error covariance directly from a large number of observation samples [98–100]. This method has a very low computational cost at the expense of requiring a large ensemble of observations with the same error statistics. In the spaceborne GNSS-R application, however, the relatively low sampling frequency (1 Hz for CYGNSS) and high receiver speed (resulting in fast changes in the geometry, antenna gain and observed wind field), limits the set of observations with similar statistics to a number too small to give a good estimation of the covariance matrix.

In this section, two methods to compute the DDM error covariance matrix are proposed. One method assumes it to be a diagonal matrix with error proportional to the observation and another method uses an empirical model considering the error correlations.

4.2 Scale Method

In the NWP data assimilation, it is common to use a diagonal observation error covariance matrix as the error correlations are generally difficult to estimate. The use of a diagonal matrix $\boldsymbol{R}$ has simple implementation and low computational cost but may lose information from the observation error correlations. Using a diagonal error covariance matrix usually results in a larger weight of the observation in the analysis compared with using a non-diagonal matrix.

[24] estimated the error in CYGNSS Level 1 DDM power to be 0.50 dB (12%) and 0.23 dB (5%) for wind speed below and above 20 m/s, respectively, by analyzing each error source in the calibration (Table II in [24]). With this in mind, we simply model the error as proportional to the observation magnitude. We used a constant of proportionality of 10% (in between the two values in [24]) and modeled the covariance matrix as

$$\boldsymbol{R} = \mathrm{diag}((0.1\boldsymbol{Y})^2) \tag{4.2}$$

where diag indicates a diagonal matrix generated from a vector.

4.3 Empirical Model

In this section, a parametric model for the DDM error covariance matrix, incorporating off-diagonal elements, is empirically developed from a large set of CYGNSS Level 1 observations. We will show that this model provides a good representation of the DDM error statistics with a low computational cost.

4.3.1 Model Analysis

The received signal at the GNSS-R receiver front end at time t can be written as

$$x_r(t) = x_s(t) + x_n(t) \tag{4.3}$$

where $x_s(t)$ is the complex scattered signal containing information of the ocean surface roughness and the speckle (due to the summation of signals with incoherent random phase). $x_n(t)$ is the background noise. Both $x_s(t)$ and $x_n(t)$ are assumed to be zero-mean complex Gaussian processes [6]. The received signal is cross-correlated with a local transmitted signal replica at a given delay-Doppler coordinate, $(\tau, f)_i$, over the coherent integration time, T_I (1 ms), to produce the complex voltage correlation signal every 1 ms, $X_i(t)$

$$X_i(t) = \frac{1}{T_I} \int_t^{t+T_I} x_r(t') a(t' + \tau_i) e^{2\pi j (f_0 + f_i) t'} dt'. \tag{4.4}$$

Decompose $X_i(t)$ into a signal term and a noise term

$$X_i(t) = X_{i,s}(t) + X_n(t) \tag{4.5}$$

where $X_{i,s}(t)$ is the correlation for the scattered signal $x_s(t)$. $X_n(t)$ is the correlation for the background noise $x_n(t)$, independent from the delay-Doppler coordinate. $X_{i,s}(t)$ and $X_n(t)$ are still zero-mean complex Gaussian. Assume the variance for the real and imaginary parts of the noise term is constant with value of σ_n^2.

The correlation signal is then squared to generate the power signal at a single coherent look every 1 ms as

$$y_i(t) = |X_i(t)|^2 = |X_{i,s}(t)|^2 + X_{i,s}(t) X_n^*(t) + X_{i,s}^*(t) X_n(t) + |X_n(t)|^2. \tag{4.6}$$

It has four terms including the signal term, noise term and two cross terms. The signal term $|X_{i,s}(t)|^2$ can be expressed as

$$|X_{i,s}(t)|^2 = y_{i,0}(t) \epsilon_i(t) \tag{4.7}$$

where $y_{i,0}(t)$ is the theoretical signal expectation and $\epsilon_i(t)$ is the speckle noise. $\epsilon_i(t)$ is shown to be closely in unit mean exponential distribution [92]. The noise term

$|X_n(t)|^2$ is also in exponential distribution with variance of $4\sigma_n^4$. The cross term $X_{i,s}(t)X_n^*(t) + X_{i,s}^*(t)X_n(t)$ has a zero mean and variance depending on the signal expectation.

The actual measurement is the DDM average power at every 1 s, computed by taking the average over consecutive N=1000 samples of the single coherent look $y_i(t)$, and subtracting the noise floor, $Y_N(t)$, giving

$$Y_i(t) = \overline{|X_{i,s}(t)|^2} + \overline{X_{i,s}(t)X_n^*(t)} + \overline{X_{i,s}^*(t)X_n(t)} + \overline{|X_n(t)|^2} - Y_N(t). \tag{4.8}$$

The noise floor $Y_N(t)$ is computed by taking the average value of M DDM power samples with rows preceding the specular point where the signal is not present. In CYGNSS, DDM samples in the first 45 rows in the raw 128×20 DDM are averaged to compute the noise floor, giving $M = 45 \times 20 = 900$ [24]. $Y_N(t)$ and $\overline{|X_n(t)|^2}$ should have the same expectation. The variances of $Y_N(t)$ and $\overline{|X_n(t)|^2}$ are $\frac{4\sigma_n^4}{MN}$ and $\frac{4\sigma_n^4}{N}$, respectively.

In equation (4.8), all average terms will approach the normal distribution according to the Central Limit Theorem. If the N samples are independent and have the same statistics, we can derive that the standard deviation (SD) of the first term in (4.8), $\overline{|X_{i,s}(t)|^2}$, is $1/\sqrt{N}y_{i,0}(t)$. However, in the spaceborne GNSS-R mission, the coherence time of $y_i(t)$ is usually a few milliseconds [97], which is larger than the integration time (1 ms), giving the fact that the N samples of the single coherent look are not independent [18, 101]. A nonlinear model for the SD of the speckle noise as a function of signal expectation and correlation time is given in [102]. The coherence time depends on the geometry and delay, Doppler of the corresponding DDM sample [103], inducing that the speckle noise can be different at different delay-Doppler coordinates. Especially, longer delays should give shorter correlation time [104], thus more independent samples within the incoherent averaging and smaller variance proportional to the signal expectation.

From equation (4.8), if we again decompose the average power, $Y_i(t)$, into a signal term and a noise term,

$$Y_i(t) = Y_{i,s}(t) + Y_n(t) \tag{4.9}$$

where

$$Y_{i,s}(t) = \overline{|X_{i,s}(t)|^2} + \overline{X_{i,s}(t)X_n^*(t)} + \overline{X_{i,s}^*(t)X_n(t)} \tag{4.10}$$

and

$$Y_n(t) = \overline{|X_n(t)|^2} - Y_N(t). \tag{4.11}$$

The signal term containing the speckle, $Y_{i,s}(t)$, is unbiased with SD related to the signal expectation. The relationship between the SD and signal expectation also depends on the delay-Doppler coordinate. The noise term $Y_n(t)$ is only related to the background noise and independent from the delay-Doppler coordinate. It has a zero mean and variance of

$$\sigma_N^2 = \frac{4(M+1)}{MN}\sigma_n^4. \tag{4.12}$$

In this model, the diagonal elements (variance) and the off-diagonal elements (covariance) of the matrix will be modeled empirically by developing parametric functions using the sample covariance matrices computed from a large set of DDM observations.

4.3.2 Data Collection

The first step is to collect a large set of DDM observations to compute enough sample observation variances/covariances under different scenarios as the reference to develop the empirical models. The DDM power in the CYGNSS Level 1 v2.1 product in June 2017 is used for this study. To compute a sample covariance matrix, each DDM sample $Y_i(t)$ is assumed to be a stationary random process within a short period of time. The statistics of a DDM depend on the observed wind speed, satellite geometries and power parameters [97]. The wind speed reference used here is the 10-meter ocean surface winds provided by the ECMWF ERA5 reanalysis in 0.25°

latitude-longitude grid [105]. The ECMWF ERA5 reanalysis winds are interpolated linearly in time and space to the specular point of each DDM as the reference wind speed. Given the approximate velocity of a CYGNSS specular point on the earth surface of around 6 km/s [39], and approximating the DDM covariance matrix as constant over scales equal to the effective ECMWF model resolution (150 km, [90]), batches of 25 sequential DDMs were used to compute the sample covariance. The satellite geometries, transmitter power, and antenna patterns were also assumed to remain constant within the corresponding 25 second time period. Such small batches of data will result in a large uncertainty in the individual covariance estimates. However, combining a large number of these batches together to estimate a small number of parameters defining the empirical model is expected to average out the uncertainty in the individual sample covariances.

Considering the assumptions above are not strong enough, some additional quality control (QC) tests were applied to the data:

- The "quality_flags" variable in the CYGNSS Level 1 data for each DDM is zero.
- The signal-to-noise ratio (SNR) for each DDM is larger than 3 dB.
- The minimum of wind speeds for each batch is larger than 3 m/s. This is to avoid the impact of the swell and coherent scattering.
- The range of wind speeds for each batch is less than 10% of the average wind speed for the batch. This is to confirm that the wind speed almost remains the same during the time of a batch, in case there is a high variational wind condition.

By the assumptions and QC tests mentioned above, 119193 DDM batches were collected in June 2017. Then a sample DDM covariance matrix can be computed from each batch.

4.3.3 Model for the Variance

In this section, an empirical model will be developed for the variance of the DDM observation (diagonal elements of the matrix (4.1)). From equations (4.9)–(4.12), the variance of a DDM sample at a delay-Doppler coordinate $(\tau, f)_i$ is

$$\sigma_i^2 = \sigma_{i,s}^2 + \sigma_N^2 \tag{4.13}$$

where $\sigma_{i,s}^2$ is the variance of signal term, $Y_{i,s}(t)$, containing the speckle. σ_N^2 is assumed constant in time and independent of the delay-Doppler coordinate. The sample variance $\hat{\sigma}_N^2$ for each batch is computed as an average of the sample variances of all DDM samples in the first two rows, where only the background noise is present. By taking the average over the sample variances from all batches, σ_N^2 is computed to be 9.576×10^{-38} W^2.

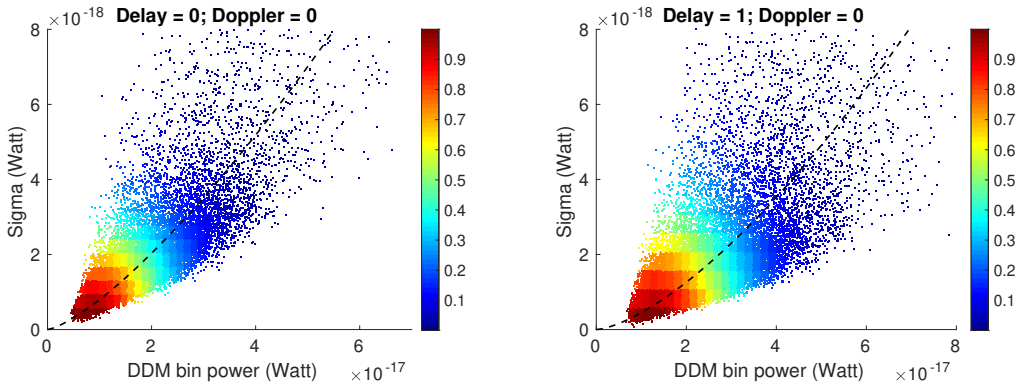
The sample variance of the DDM measurement $\hat{\sigma}_i^2$ for each batch is computed as well. Then the sample SD of the speckle noise for each batch is calculated by

$$\hat{\sigma}_{i,s} = \sqrt{\hat{\sigma}_i^2 - \sigma_N^2}. \tag{4.14}$$

As is stated in section 4.3.1, the SD of the speckle noise depends on the expectation of the signal. As we have no information about the expectation, we assume that the SD depends on the magnitude of the measurement. In each batch, the measurement magnitude is computed by the average value of all the samples in the batch. Figure 4.1 shows the scatterplots for the SD of the speckle noise and DDM power magnitude from all batches at two different DDM samples, in which Figure 4.1(a) is for the specular point. Scattering of the values on both figures is due to the limited number of samples in each batch.

An empirical parametric model is assumed to fit the data by the least square method as

$$\sigma_{i,s} = p[i] Y_i^{q[i]} \tag{4.15}$$



(a) (b)

Figure 4.1.: Scatterplots for the SD of the speckle noise and DDM power magnitude at two different DDM samples colored by the density of the points. In the titles of the two figures, "Delay" and "Doppler" stands for the delay and Dopper of the DDM sample shifted from the specular bin in the unit of sample, where positive shift in delay and Doppler stands for larger delay and Doppler. The fitting curves are shown in black dash lines on both figures.

where Y_i is the power magnitude at a given DDM sample, a and b are independent parameters of the model. The fitting curves are shown on Figure 4.1. As the indices of the specular bin in the DDM are provided in the CYGNSS L1 data, this model fitting was applied to all DDM samples at delays [-1,10] and Dopplers [-3,3] shifted from the specular bin in the unit of sample, where positive shift in delay and Doppler stands for larger delay and Doppler. This provides matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ in the dimension of 12×7 containing all values of p and q for DDM samples in the range of delay and Doppler. Figure 4.2 shows the fitting curves for SD of the speckle noise versus power magnitude for DDM samples at different delays and Dopplers.

4.3.4 Model for the Covariance

In this section, an empirical model for the correlation coefficient of two DDM samples will be developed. The covariance of two DDM samples (off-diagonal elements in equation (4.1)) can be calculated using the correlation coefficient and SD of the two

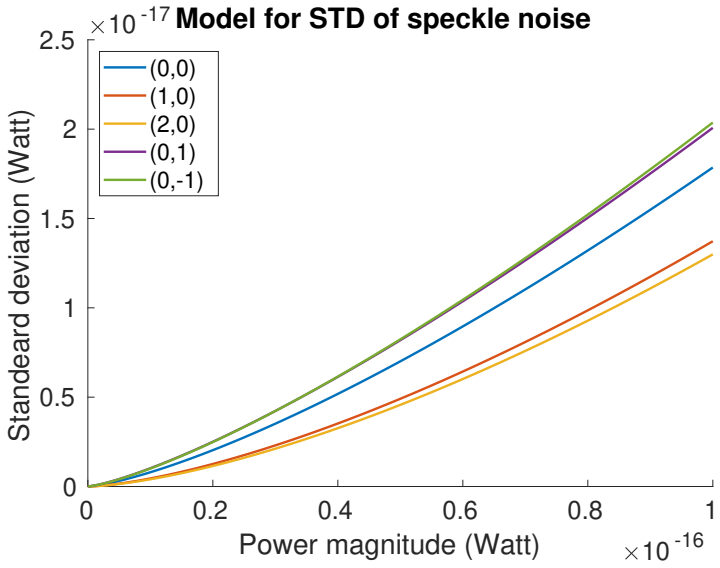
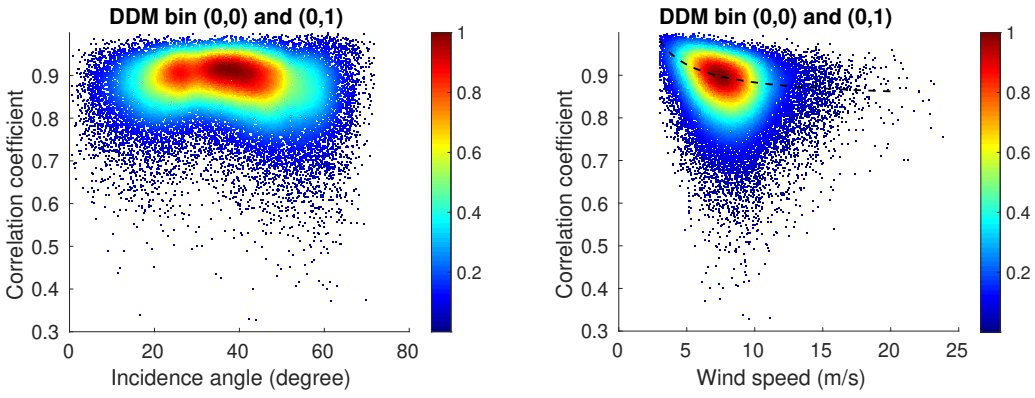


Figure 4.2.: Model for SD of speckle noise versus power magnitude for DDM samples at different delays and Dopplers. In the legend, (0,0) stands for the specular bin and other sets of numbers stand for DDM samples at delay and doppler shifted from the specular bin.

DDM samples computed as in section 4.3.3. The sample correlation coefficient at two different delay-Doppler coordinates, $(\tau, f)_i$ and $(\tau, f)_j$, for each batch is computed as

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j} \tag{4.16}$$

where $\hat{\sigma}_{ij}$ is the sample covariance of two DDM samples, i and j; $\hat{\sigma}_i$ and $\hat{\sigma}_j$ are sample SD of the two DDM samples. As no theoretical model for the correlation coefficient exists, the empirical model will be developed by observing the data. The covariance model in [97] gives a hint that the correlation coefficient might depend on the geometry, wind speed and parameters in the forward operator. Figure 4.3 shows the scatterplots of the correlation coefficient between DDM sample (0,0) and (1,0) versus incidence angle and wind speed at the specular point, where the incidence angle and wind speed are the average values of all DDMs in each batch. It can be observed that the correlation coefficient has no strong dependence on the incidence



(a) (b)

Figure 4.3.: Scatterplots of the correlation coefficient versus incidence angle and wind speed at the specular point between DDM sample (0,0) and (1,0) colored by the density of the points. The fitting curve for the right figure is shown in black dash line.

angle, but an evident dependence on the wind speed. Patterns from similar figures like Figure 4.3(a) can also prove that the correlation coefficient does not strongly depend on SNR, DDM power magnitude, transmitter EIRP and receiver antenna gain (not shown). So, the correlation coefficient between two DDM samples is assumed to be monotonically dependent on the wind speed at the specular point. An empirical parametric model was found to fit the data well by the least square method in the form

$$\rho_{ij} = a[i,j] + b[i,j]u^{-1} + c[i,j]u^{-2} \tag{4.17}$$

where $a[i,j]$, $b[i,j]$ and $c[i,j]$ are independent parameters of the model. The fitting curve for the data in Figure 4.3(b) is shown as the black dash line. This model fitting is then applied to every two DDM samples at delays [-1,10] and Dopplers [-3,3]. This provides symmetric matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ in dimension of 84×84, where the diagonal values of $\boldsymbol{A}$ are all ones and the diagonal values of $\boldsymbol{B}$ and $\boldsymbol{C}$ are all zeros. Figure 4.4 shows the fitting curves for correlation coefficients between the specular bin and DDM samples at different delays and Dopplers.

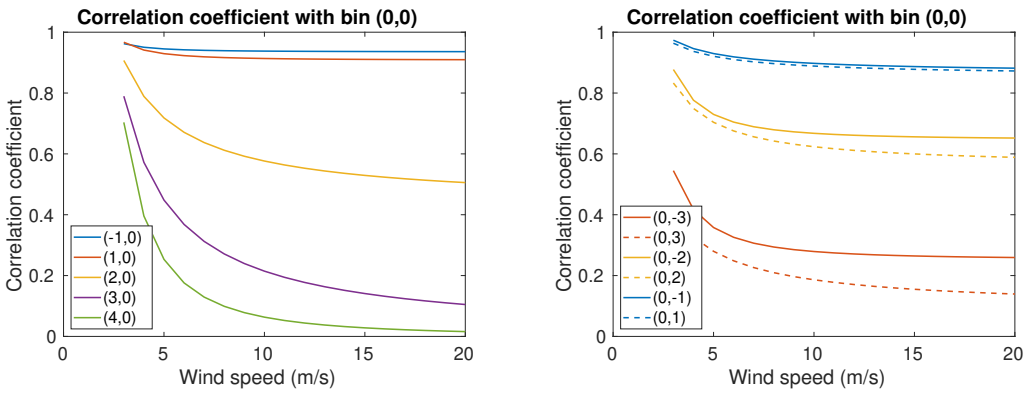


Figure 4.4.: Model for the correlation coefficient of the specular bin and DDM samples at different delays and Dopplers: (a) shifts in delay dimension and (b) shifts in Doppler dimension.

4.3.5 Implementation

Given the variance of the background noise term, σ_N^2, matrices $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ calulated in section 4.3.3 and 4.3.4, the modeled covriance matrix can be constructed from the real data as follow steps:

1) Get the location of the specular bin and informative samples of the DDM. For CYGNSS, the indices of the specular bin are in the Level 1 data. The informative samples are chosen as described in section 3.1.

2) Compute the wind speed at the specular point. In DA, this is the wind speed interpolated from the background.

3) Compute the variance of each informative samples (diagonal elements of the covriance matrix) as

$$\sigma_i^2 = \sigma_N^2 + \left(p[i]Y_i^{q[i]}\right)^2. \tag{4.18}$$

Most of the informative samples should fall into the region covered by the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$. If an informative sample is outside the region, the corresponding p and q values are computed by the nearest interpolation.

4) Compute the correlation coefficient between every two DDM samples by equation (4.17). If an informative sample is outside the region, the corresponding a, b and c values are computed by the nearest interpolation. Then compute the covariance between every two DDM samples (off-diagonal elements of the covriance matrix) as

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j. \tag{4.19}$$

where σ_i and σ_j are computed in the previous step.

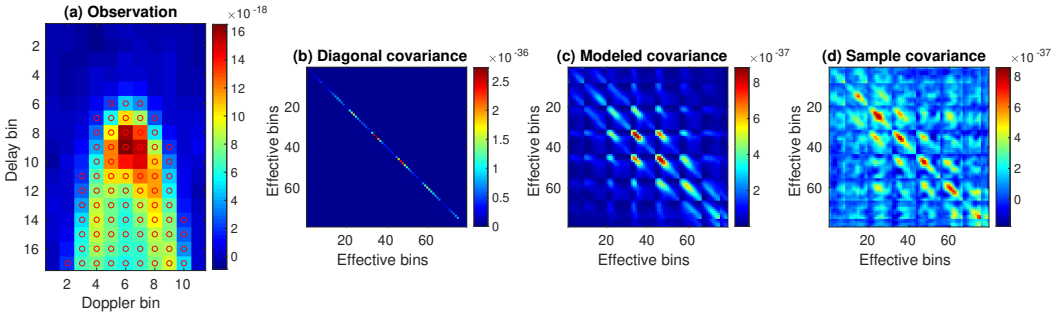


Figure 4.5.: Comparison between the modeled DDM error covariance matrices and DDM sample covariance matrix for the CYGNSS mission. (a) DDM observation with informative samples as red circles. (b) Diagonal covariance matrix computed by the scale method in section 4.2. (c) Non-diagonal covariance matrix computed by the model in section 4.3. (d) Sample covariance matrix computed from sequential 25 DDM samples.

Figure 4.5 presents an example for the comparsion of the different covaraince matrix models. Figure 4.5(a) is the DDM observation collected by CYGNSS SV2 with GPS PRN 20 at UTC Second of Day (SOD) 86339 on 1 June 2017. Figure 4.5(b) is the corresponding diagonal covariance matrix computed by the scale method in section 4.2. Figure 4.5(c) is the corresponding non-diagonal covariance matrix computed by the empirical model developed in section 4.3. Figure 4.5(d) is the sample covariance matrix computed from DDM observations between SOD 86327 and 86351 with CYGNSS SV2 and GPS PRN 20. Note that the sample covariance matrix can be noisy because it is computed using only 25 samples. It can be observed that the

empirical non-diagonal covariance matrix model captures much of the structures of the sample covariance matrix.

The inverse of the covariance matrix, $\boldsymbol{R}^{-1}$, is required in the VAM cost function (3.6). It is found that the covariance matrix computed by the empirical model is often ill-conditioned, making it difficult to compute an accurate inverse. Ridge regression [106], a reconditioning method, is applied to reduce the condition number of the matrix to ~100. This method increases the diagonal values of the matrix by a fixed number and thus will also increase the modeled variances of the observations.

5. DATA ASSIMILATION IN GLOBAL OCEAN SURFACE WIND ANALYSES

5.1 Introduction

The constellation and low-inclination orbit of CYGNSS provide wind observations across the global tropics with a 7 hour mean revisit time, filling the temporal and spatial gaps from conventional microwave instruments, which are mostly in polar orbits [39]. A comprehensive summary of results of assimilating CYGNSS DDMs into global NWP analyses is presented in chapter. One month of CYGNSS DDMs were assimilated using a 20 minute cycle. The background is from the 10-meter surface winds of the European Centre for Medium-Range Weather Forecasts (ECMWF) model. DA performance is assessed by comparison with collocated scatterometer winds. We will show two benefits of DDM assimilation:

1) A positive impact on the NWP analyses over a swath at least 150-km wide;

2) Lower error than wind retrievals (e.g., CYGNSS Level 2 products) when interpolated to CYGNSS specular points.

5.2 Data Description and Experimental Design

5.2.1 CYGNSS DDM Observations

CYGNSS version 2.1 Level 1 DDM data from 1 June 2017 to 30 June 2017 were used as observations. Level 1 data also include the transmitter EIRP and satellite geometries, estimated by the CYGNSS SOC. Receiver antenna patterns were separately provided by the SOC as well.

5.2.2 ECMWF Background

ECMWF is an independent intergovernmental organization aiming to provide accurate medium-range global weather forecasts supported by most European countries [107]. Zonal and meridional (u, v) components of the 10-meter ocean surface winds provided by the ECMWF operational forecast from 1 June 2017 to 30 June 2017 were used for the background wind field. The ocean surface winds in ECMWF are hourly forecasts initiated from analysis times at 00UTC and 12UTC on a grid spacing of 18 km.

5.2.3 Scatterometer Winds

A scatterometer is an instrument to measure the roughness of a surface using radar backscatter. Spaceborne scatterometers have provided accurate wind field information for meteorology and climate over the past decades. Scatterometer (SCAT) 10-meter ocean surface winds from ASCAT aboard the Metop satellites (Metop-A and Metop-B) and OSCAT aboard the ScatSat-1 satellite [108, 109] were used for validation in this study. The Metop satellites were developed by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and the ScatSat-1 satellite was developed by the Indian Space Research Organisation (ISRO). ASCAT has two sets of three antennas measuring ocean surface winds by two 550-km-wide swaths on both sides of the satellite ground track. It provides 10-meter wind products with 25-km and 12.5-km cell spacing. OSCAT uses a dish rotating antenna measuring ocean surface winds by an 1800-km-wide swath, providing 10-meter wind products in 50-km and 25-km cell spacing. The 25-km products from both instruments were used in this study to evaluate the result of DDM assimilation.

The 25-km zonal and meridional wind components measured by both instruments have been validated to have error standard deviation less than 1 m/s by a triple collocation method compared to buoy wind measurements and NWP models [110, 111].

5.2.4 CYGNSS Wind Products

The CYGNSS Level 2 product, CYGNSS Climate Data Record (CDR) product and NOAA CYGNSS wind product are three different CYGNSS wind speed products retrieved from the CYGNSS Level 1 product using different algorithms. They will be used to evaluate results of the DDM assimilation at the specular points.

- *CYGNSS Level 2 product v2.1:* Two observables, NBRCS and LES are first computed from a 3×5 window of the Level 1 DDM BRCS around the specular point. Two GMFs are developed to retrieve the 25-km surface wind speed at the specular point from these two observables. The two resulting wind speeds are then optimally combined to derive the minimum variance (MV) wind speed [37].

- *CYGNSS Level 2 CDR product v1.0:* This is a new wind product released by CYGNSS SOC in 2020 [112]. It is similar to the CYGNSS Level 2 product except that the observables NBRCS and LES are track-wise corrected using NASA's MERRA-2 wind product to calibrate unknown fluctuations in GPS transmitter EIRP. Additional QCs are also applied to the observables.

- *NOAA CYGNSS wind product:* Prepared by the National Oceanic and Atmospheric Administration (NOAA), this product is a 25-km surface wind speed at the CYGNSS specular points [87]. A new GMF was derived that expresses the CYGNSS NBRCS observable as a function of wind speed, incidence angle and significant wave height. The NBRCS observables are also track-wise corrected using the ECMWF model. 25-km gridding is implemented along each track to avoid overlapping observations. Additional rigorous QCs are applied to the data.

5.2.5 Experimental Design

The CYGNSS specular points were collocated with the SCAT wind vector cells (WVC) for all data from 1 June 2017 to 30 June 2017. Maximum differences of

40 minutes in time and 25 km in distance were used as criteria for collocation. If a CYGNSS specular point is collocated with several WVCs from different satellites (Metop-A, Metrop-B or ScatSat-1) then the average value of the wind speeds in all collocated WVCs was used. The DA experiment was done using a 20 minute cycle (0–20, 20–40, 40–60 minutes in each hour). In each 20-minute period, the analysis time is at the center of each cycle and the wind field is assumed to be constant. Hourly ECMWF surface winds were quadratically interpolated to the center time of each cycle from 0000 UTC on 1 June 2017 to 2400 UTC on 30 June 2017 and used as the background. The original ECMWF surface winds were also bilinearly interpolated to 0.125° grid spacing to match the working resolution of the DDM forward operator. In each cycle, all CYGNSS DDMs that were measured within the time period, passed the QC described in section 3.4, and were collocated with the SCAT WVCs were assimilated with the background using the VAM to produce the analysis on a 0.125° grid.

Two comparisons were made between the analysis winds and the reference SCAT winds.

- *Comparison at the specular points:* Wind vectors from the background and analysis wind field are linearly interpolated to the CYGNSS specular points and then compared to the collocated SCAT winds.

- *Comparison over a swath along the specular point track:* In order to evaluate the extent of the impact of assimilating DDMs, the wind vectors are compared over a much larger area than the one grid cell located at the specular point. CYGNSS data are first separated into different tracks corresponding to a specific pair of GPS transmitter and CYGNSS receiver. Along each track, background and analysis wind vectors on the 0.125° grid within a swath of a certain width are compared with collocated SCAT observations. Wind speeds at SCAT WVCs are linearly interpolated to the 0.125° grid of VAM wind field for the comparison.

Figure 5.1 shows an example of the collocation for CYGNSS specular points, an 80-km-wide swath of the VAM gridded wind field, and 25-km SCAT WVCs.

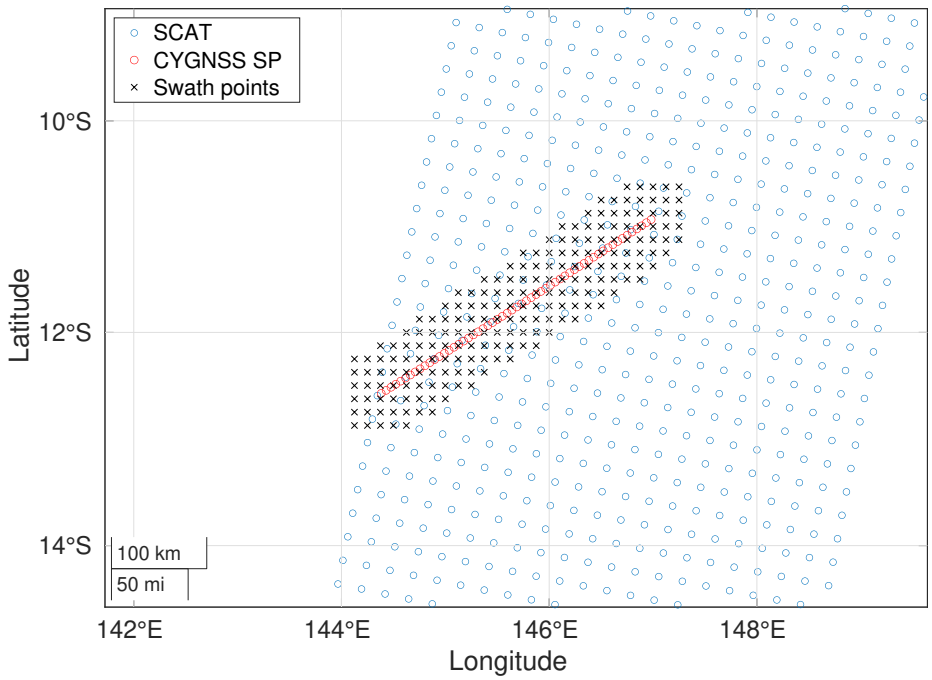


Figure 5.1.: An example of the collocation for CYGNSS specular points, 0.125° grid points of the CYGNSS 80-km-wide swath, and 25-km WVCs of the SCAT swath in the period 00:00–00:20 UTC on 10 June 2017. The CYGNSS observations are measured by GPS PRN 14 and CYGNSS SV 5. The SCAT measurements are from ASCAT-A.

The results of using three different DDM error covariance matrices are also compared:

a) A diagonal matrix using the scale method presented in section 4.2 (R-scale);

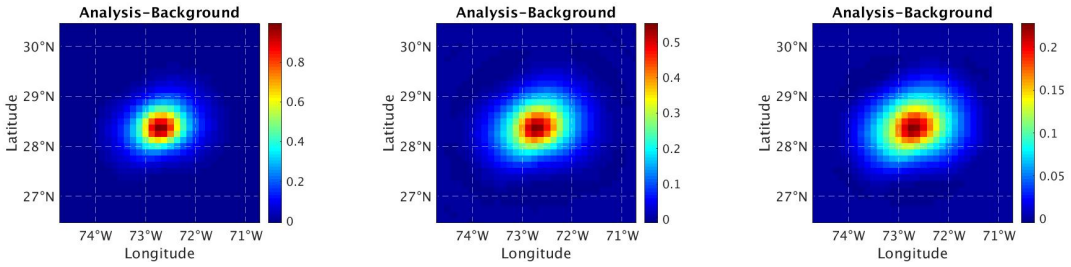
b) A diagonal matrix whose diagonal values are computed using the model presented in section 4.3 (R-model-diagonal);

c) A non-diagonal matrix computed using the model presented in section 4.3 (R-model).

5.3 Tuning the Weights

As introduced in section 3.2, there are a number of coefficients that can be used to weight the relative importance of the background winds vs. the new information. The constraint term and its weights describe background error correlations. In the study, the weight and standard deviation of the background wind components in equation (3.3) were fixed to be $\sigma_b = 1$ m/s and $\lambda_\beta = 4$. Only the ratio between these weights is important. The observation weight, λ_{ddm}, and constraint weights, λ_{lap}, λ_{div}, λ_{vor} were then determined by a series of sensitivity tests. In general, increasing the observation weight increases the intensity of the DA response, making the analysis closer to the observation, but does not change the shape of the response. Increasing the constraint weights increases the spatial scale of the response and decreases the intensity. λ_{lap} controls the smoothness of the response. λ_{div} and λ_{vor} control the shape of the response. Increasing the observation and constraint weights will also increase the number of iterations and computation cost in the minimization.

The constraint weights were first determined by a sensitivity test. Since they describe the background error correlations, the spatial scale of the response should be similar to the scale of the background effective resolution. It is important to note that the NWP grid spacing size and the model's effective resolution are different. In previous studies, the effective NWP model resolution was found to be 4-8 times larger than the grid spacing size [113, 114]. In our case, the effective model resolution of the ECMWF background is expected to be around 150 km [90]. Figure 5.2 shows the responses of assimilating a single DDM observation using three different sets of constraint weights. This example clearly show that increasing the constraint weight increases the area over which observations would have an effect. The DDM covariance is computed by the scale method and the observation weight λ_{ddm} is 1/4 in all three cases.



(a) (b) (c)

Figure 5.2.: Wind speed increments (analysis−background) of assimilating a single DDM using different constraint weights. $(\lambda_{div}, \lambda_{vor}, \lambda_{lap})$ = (a) (50, 100, 25); (b) (200, 400, 100); (c) (800, 1600, 400). Higher weights increase the extent of the impact of new observations and reduce the increment's intensity. The DDM is observed by CYGNSS SV 4 and GPS PRN 2 at UTC SOD 4723 on 1 June 2017. The background and observation weights are 4 and 1/4 in all three cases.

Considering that the footprint of a DDM observation is around 100 km and the model's effective resolution is around 150 km, the scale of the response should be about 250 km. By the sensitivity test, the constraint weights were chosen to be

$$(\lambda_{div}, \lambda_{vor}, \lambda_{lap}) = (200, 400, 100). \tag{5.1}$$

After determining the constraint weights, the observation weight λ_{ddm} is determined by another sensitivity test. As the CYGNSS specular point moves at about 6 km/s on the earth surface and the impact area of a DDM is about 250 km, the analysis wind speed at a point on the ocean surface can be impacted by 35-40 DDMs. Since the area impacted by a DDM through DA (~250 km) is larger than the area of its glistening zone (~100 km), the analysis wind speed at one point on the ocean surface can be affected by DDMs that, by themselves, are not sensitive to winds at that point. Due to this feature of overlapping measurements, in general λ_{ddm} should be much smaller than λ_b as a "deweighting" or equivalent "thining" of the observations. A total of ~25,000 DDMs from one day (10 June 2017) are processed by the VAM using a set of different observation weights, $\lambda_{ddm} = (1/64, 1/16, 1/4, 1, 4, 16)$, for each

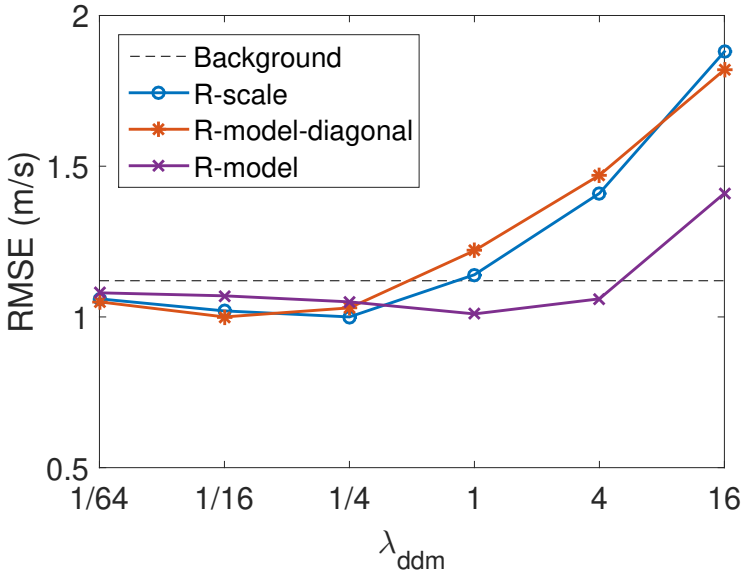


Figure 5.3.: Wind speed RMSE at CYGNSS specular points versus observation weight in the VAM for different DDM error covariance matrices (R-scale, R-model-diagonal, R-model). The background wind speed RMSE at specular points is shown as the black dash line. Results are computed using data of one day on 10 June 2017.

of the three DDM error covariance matrices. In each case, the Root Mean Square Error (RMSE) between the VAM and SCAT wind speeds, evaluated at the specular point, was computed. Figure 5.3 shows the RMSE for all cases in the sensitivity test. The optimal λ_{ddm} for each DDM covariance matrix can be found by choosing the one with the minimal RMSE.

This result shows that the best observation weights λ_{ddm} for the three DDM error covariance matrices (R-scale, R-model-diagonal, R-model) are 1/4, 1/16 and 1, respectively. The optimal weight for the non-diagonal matrix (R-model) is larger than that for a diagonal matrix (R-model-diagonal) because adding error correlations and reconditioning the covariance matrix will reduce the weight of the observation [106]. When λ_{ddm} decreases, the analysis wind field approaches that of the background, so it is expected that the RMSE in each case would likewise approach the background RMSE. When λ_{ddm} increases beyond its optimal value, the RMSE increases dramat-

ically due to overfitting. Therefore, if the optimal λ_{ddm} cannot be precisely decided in an experiment, it is generally preferable to use a smaller one.

5.4 Use of Observation Error Covariance Matrix

Results from our study using one day of data (Figure 5.3) show that, if the optimal λ_{ddm} is selected, there is little difference in the RMSE from using either of the three DDM error covariance matrices. To additionally validate the performance of using the three matrices, a total of ~170,000 DDMs from 5 days data (10 June 2017 to 14 June 2017) were processed using the three matrices combined with the corresponding optimal weights. The comparison was made both at the specular points and over swaths with two different widths. The results are listed in Table 5.1.

Table 5.1.: Wind speed RMSE compared to SCAT at CYGNSS specular points, over 80-km swath, and over 120-km swath. Comparison of results using different error covariance matrices. 5 days (10 June 2017 to 14 June 2017) of data. All units in m/s.

DDM error covariance matrix	Specular	80-km swath	120-km swath
R-scale	1.03	1.05	1.07
R-model-diagonal	1.04	1.07	1.10
R-model	1.06	1.08	1.10

The conclusion of this study is that there is no significant difference in the accuracy of DA results, from comparisons at either the specular points or over a swath, using either of the three observation error covariance matrices. The slight differences in the results of using the three matrices are possibly because that only a set of discrete values of λ_{ddm} are tested. Similar performance for all three covariance matrics could be explained by the following reasons:

1) The VAM is heuristic. The observation error covariance matrix and the λ_{ddm} weight together determine the relative contribution of the observation in the analysis. Error in modeling the observation covariance matrix is compensated

by choosing the optimal weight in the sensitivity study. This explains why the optimal λ_{ddm} for the three different covariance matrices are different whereas their final RMSE results are almost the same.

2) Each DDM sample observes an area defined by its delay and Doppler coordinate. This area on the ocean surface is usually 10-50 km across, which is much smaller than the ECMWF effective model resolution (150 km). Although the error correlations between each DDM sample may provide extra information, this small-scale information is smoothed out by the constraint terms in the VAM which are controlled by the effective model resolution of the background.

3) The reconditioning method used to decrease the large condition number of the non-diagonal error covariance matrix could add extra noise to the DA process, counteracting the benefit of additional information contained in the off-diagonal elements.

It is valuable to note in Figure 5.3, that the RMSE for R-model increases more slowly than the RMSE for R-scale when λ_{ddm} increases beyond its optimal value. This means that results from using R-model would be less sensitive to the choice of λ_{ddm}. One possible reason for this effect could be that the performance of DDM assimilation is mainly dependent on the error variances of DDM bins near the specular point and the weight λ_{ddm}. So if λ_{ddm} is selected to accurately correct the observation error covariance, the result is not sensitive to the method computing the covariance matrix. Whereas, if λ_{ddm} is not optimal, more accurately estimated covariances of DDM bins away from the specular point (from R-model) could mitigate the effect of sub-optimal weighting.

Our conclusion is that the three DDM error covariance matrices should give similar results when the optimal λ_{ddm} is selected. For the remainder of this study, the covariance matrix R-scale with its optimal weight will be applied, due to its simplicity. The non-diagonal covariance matrix R-model accounting for error correlations in

the DDM could be valuable if DDMs are assimilated into DA systems at mesoscale or smaller spatial scales, e.g., a regional weather forecast model.

5.5 Assimilation Results

One month of CYGNSS Level 1 data from 1 June 2017 to 30 June 2017 (~663,000 DDMs, after applying the QC in section 3.4) was assimilated with the ECMWF background into the VAM to produce the analysis wind field (ECMWF-CY-DDM). The R-scale covariance matrix was used with weights determined in section 5.3 ($\lambda_{ddm} = 1/4$). Figure 5.4 shows an example of the wind field background, analysis, and increment for the 20-minute period from 6:40-7:00 UTC on 1 June 2017. This figure demonstrates that the impact of assimilating a track of DDMs extends over a 200-250 km wide swath. Figure 5.5 shows the wind vectors on the contour maps of the background, analysis and increment for a closer look at a region in the same time period. Since an isotropic slope PDF is assumed, with MSS a monotonic function of wind speed, the DDM observations will contain essentially no information of the wind direction. Analysis wind directions from the VAM are almost the same as those in the background, except for some negligibly slight changes due to the flow-dependent constraint terms.

A pair of density scatterplots showing a comparison of background and analysis wind speeds at CYGNSS specular points to SCAT winds is shown in Figure 5.6. The symmetric distribution of the samples with respect to the 1:1 line in both subfigures shows that both background and analysis are almost unbiased. The total wind speed RMSE at the specular points decreases from 1.17 to 1.07 m/s and the mean difference (bias) decreases from -0.14 to -0.08 m/s as a result of assimilating the DDMs. Wind speeds from both the background and analysis are smaller than SCAT wind speeds in general. The reduction of this bias, therefore, implies that the assimilation of CYGNSS DDMs increase the wind speeds from the ECMWF background on average. The wind speed RMSE and bias at the specular points for the background and analysis at different ranges of SCAT wind speed are shown in Table 5.2. Both the RMSE and

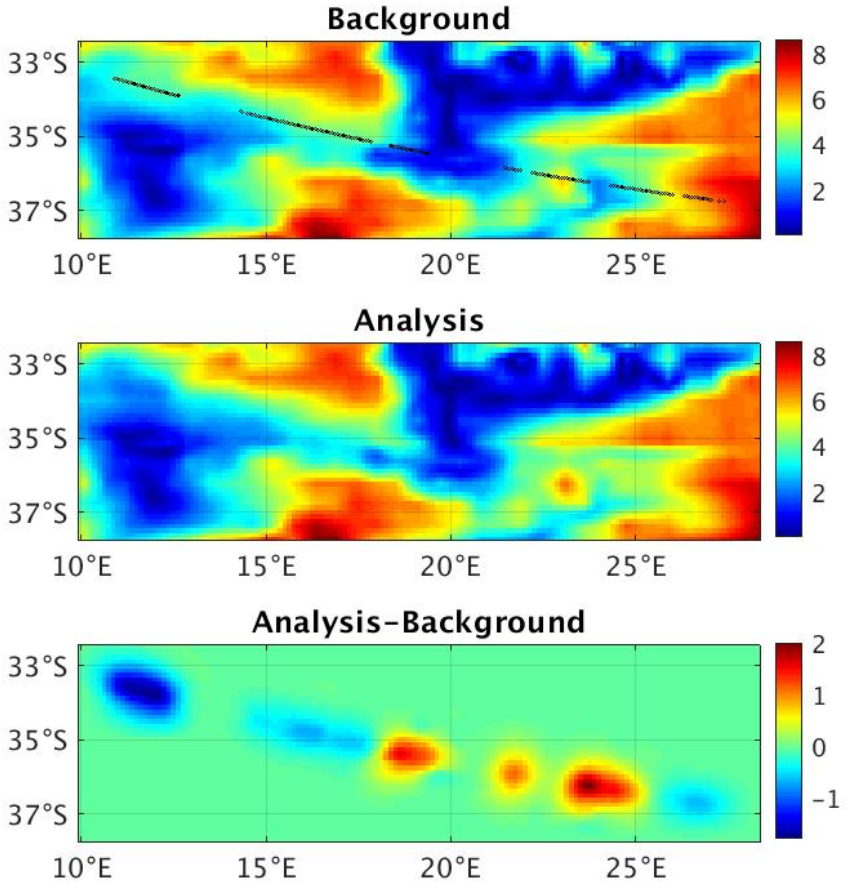


Figure 5.4.: Wind field maps of the ECMWF background, VAM analysis and increment (analysis−background) at 6:50 UTC on 1 June 2017. The CYGNSS specular point track is shown as the black circles on the background map. Units in m/s.

bias of the background are significantly decreased by the assimilation of CYGNSS DDMs for wind speed less than 15 m/s, while the statistics almost remain the same for wind speed larger than 15 m/s. The decrease of the performance on high wind speed cases is mainly related to the decrease in sensitivity of the DDM measurements (surface slope PDF) to wind speed at high wind speeds, which is an intrinsic limitation of the physics in GNSS-R [40]. Also, the impact of wave age and fetch length at high

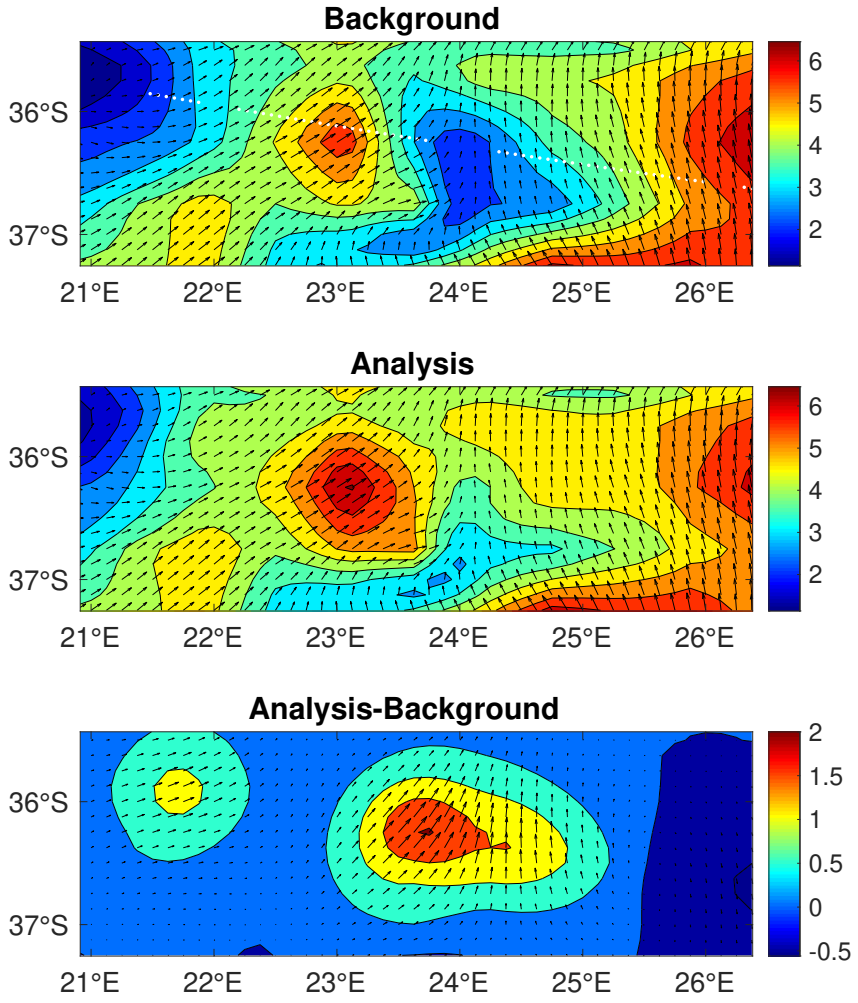


Figure 5.5.: Wind contour maps and wind vector fields of the ECMWF background, VAM analysis and increment (analysis−background) at 6:50 UTC on 1 June 2017. Only a small region of the whole map is presented here. The CYGNSS specular point track is shown as the white circles in the background map. Wind vectors on the increment map are shown at a scale 5 times larger than that used on the Background and Analysis maps. Units in m/s.

wind speeds, which are not considered in the forward operator, could be another source of error. Nevertheless, the bias correction scheme prevents the assimilation of

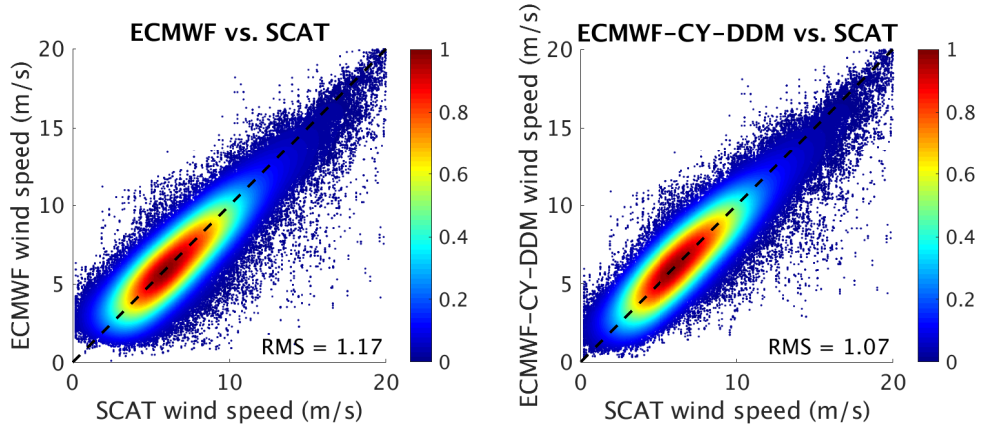


Figure 5.6.: Density scatterplots for the comparison of ECMWF background wind speeds (ECMWF) and VAM analysis wind speeds (ECMWF-CY-DDM) versus SCAT wind speeds at the CYGNSS specular points for one month of data (June 2017) . The color stands for the density (normalized number) of the samples.

DDMs from introducing additional errors into the background at high wind speeds. In the comparison of wind directions, data with collocated SCAT wind speeds less than 4 m/s are excluded because SCAT wind directions are less accurate at low wind speeds [115]. The wind direction RMSEs of the background and analysis at specular points for the one month of data are 20.73° and 20.70°, the biases are 0.011° and 0.003°, respectively, compared to SCAT wind directions. The analysis retains the wind direction information from the background while the wind speeds are changed by the DDM assimilation.

Wind speed error statistics are also computed over swaths of various widths (80, 120, and 150 km) along the CYGNSS specular point tracks. These results are listed in Table 5.3. Assimilation of CYGNSS DDMs is shown to improve the wind field accuracy, both at the specular point and over all swath widths. This improvement decreases as the swath width increases, which we interpret to be a consequence of the reduced sensitivity of the DDM away from the specular points. These results demonstrate the capability of CYGNSS DDM assimilation to improve the analyses of

Table 5.2.: Wind speed RMSE and mean difference (bias) of ECMWF background and VAM analysis (ECMWF-CY-DDM) compared to SCAT wind speeds over different ranges of SCAT wind speeds. The number of observations (Nobs) in each wind speed range is listed as well. The RMSE and bias are in unit of m/s.

Wind speed range	< 5 m/s	5–10 m/s	10–15 m/s	> 15 m/s	Total
Nobs	178,498	393,826	80,918	9,425	663,909
ECMWF RMSE	1.14	1.08	1.39	2.45	1.17
ECMWF-CY-DDM RMSE	0.98	0.99	1.34	2.45	1.07
ECMWF Bias	0.33	-0.21	-0.66	-1.48	-0.14
ECMWF-CY-DDM Bias	0.22	-0.07	-0.62	-1.50	-0.08

global NWP models. The reduction of RMSE and bias of the ECMWF background is comparable to results from assimilating conventional scatterometer winds at global NWP centers [115, 116].

Table 5.3.: Wind speed RMSE of the ECMWF background and VAM analysis (ECMWF-CY-DDM) at the CYGNSS specular points and over a swath with different widths (80-km, 120-km and 150-km) compared to SCAT wind speeds. All units in m/s.

	Specular	80-km swath	120-km swath	150-km swath
ECMWF	1.17	1.18	1.19	1.20
ECMWF-CY-DDM	1.07	1.10	1.11	1.13

Another benefit of DDM assimilation is that the interpolated wind vectors from the VAM analyses can subsequently be used as observations in other DA systems. To evaluate the performance of those wind speed retrievals, the interpolated wind speeds at the specular points from ECMWF-CY-DDM are compared to several other CYGNSS wind products: CYGNSS Level 2, CYGNSS Level 2 CDR, and NOAA-CYGNSS, which are described in section 5.2.4. All three products are 25-km wind speeds at the CYGNSS specular points retrieved from the CYGNSS Level 1 product. Both the CYGNSS Level 2 CDR product and NOAA CYGNSS wind product apply

a track-wise correction on the retrieved wind speeds using referenced NWP models. Wind speeds in the three products retrieved from the same CYGNSS Level 1 product for the one month of data in this study are compared to collocated SCAT winds. Note that all the three products apply some additional QCs and the NOAA CYGNSS wind product implements 25-km gridding along the track. Therefore, there are fewer collocated wind speeds from these three products (especially in the case of the NOAA product) than the number of CYGNSS Level 1 observations used in the DDM assimilation. RMSE and bias of all four products are compared in Table 5.4. The wind speeds from ECMWF-CY-DDM are shown to have smaller RMSE and bias than any of the other CYGNSS products. Another advantage of the DDM assimilation is that a wind direction is assigned to each specular point, which might be beneficial to DA systems.

Table 5.4.: Wind speed RMSE and bias at CYGNSS specular points compared to collocated SCAT wind speeds for CYGNSS Level 2 product (CYGNSS-L2), CYGNSS CDR product (CYGNSS-CDR), NOAA CYGNSS wind product (NOAA-CYGNSS), ECMWF background and VAM analysis (ECMWF-CY-DDM), for one month of data (June 2017).

	Nobs	RMSE (m/s)	Bias (m/s)
CYGNSS-L2	661,230	1.50	-0.45
CYGNSS-CDR	520,432	1.57	-0.44
NOAA-CYGNSS	135,931	1.20	-0.33
ECMWF	663,909	1.17	-0.14
ECMWF-CY-DDM	663,909	1.07	-0.08

5.6 Computational Efficiency

Although DDM assimilation has been shown to improve global NWP models and produce wind speed estimates at a higher accuracy than conventional Level 2 products, it does come with a significant computational cost. The DDM forward operator needs to be run at each iteration in the VAM's minimization. The cost function in the

VAM is minimized by a Quasi-Newton algorithm [84], using the convergence criteria listed in Table 5.5. About 30–50 function evaluations (running the forward operator) are generally required to reach the minimum.

Table 5.5.: The convergence criterion in VAM's minimization.

Maximum infinity norm for the gradient of the cost function	10^{-6}
Maximum infinity norm for the change of the state between two iterations	10^{-6}
Maximum number of iterations	30
Maximum number of function evaluations	50

The experiment tasks in this study were run in parallel on two servers using Intel Xeon processors (one with 10 cores at 3.10-GHz, another with 12 cores at 2.53-GHz). Running the forward operator one time to compute a simulated DDM and a Jacobian matrix takes 0.4–0.5 CPU seconds on either server. Assimilating one DDM in the VAM takes about 20–30 CPU seconds. In total, it takes about 20 days elapsed time to process one month of data with ~663,000 DDMs using both servers running in parallel by *GNU parallel* [117].

The wind field grid size in this study is small (0.125°), which makes the computational cost of the DDM assimilation relatively high. The computational cost can be reduced by using a larger grid size of the wind field. Computing the forward operator by GPUs should also significantly improve the computational efficiency [118].

5.7 Conclusion

Use of the VAM for assimilating CYGNSS Level 1 DDM power into global NWP analyses, including validation and assessment of the results, has been presented in this chapter. A track-wise bias correction scheme was found to be necessary. The best results were obtained using a simple diagonal observation covariance matrix combined with optimal selection of the cost function weights. However, we did find a lower sensitivity to the observation weight when a non-diagonal covariance matrix was

used. Our explanation for this effect is that the observation weight can counteract an inaccurate covariance matrix and the small-scale information in the error correlations is smoothed out by the constraint terms in the VAM. For some applications, such as regional forecast models, a full observation covariance matrix accounting for correlation between delay-Doppler coordinates may be beneficial.

We demonstrated our approach on one month (June 2017) of CYGNSS data collocated with SCAT observations, consisting of ~663,000 Level 1 DDMs. The VAM was used with ECMWF background winds in a cycle of 20 minutes to produce analysis winds on a 0.125° grid. Assimilation of a track of DDMs was shown to have an impact over a 200–250 km wide swath, corresponding approximately to total extent of the DDM footprint (~100 km) plus the ECMWF effective model resolution (~150 km). These results also showed a reduction of the RMSE from 1.17 to 1.07 m/s and bias from -0.14 to -0.08 m/s as compared to reference scatterometer wind speeds. Wind directions were not changed significantly in the analyses, with an RMSE of 20.7° and bias of 0.0° compared to scatterometer data. DDM assimilation was also shown to improve the background wind field over a swath up to 150 km wide, reducing the wind speed RMSE from 1.20 to 1.13 m/s. These results indicate that assimilation of GNSS-R DDMs can have a positive impact on NWP analyses. We found that improvement was mostly limited to wind speeds below 15 m/s, however, probably due to the decrease in the sensitivity of DDM observations to higher winds.

Wind vectors interpolated to the CYGNSS specular points can also be treated as a Level 2 product. These were compared to wind speeds from other CYGNSS wind products (Level 2, Level 2 CDR and NOAA). The RMSE and bias of ECMWF-CYDDM wind speeds were found to be lower than those of these other three products, as compared to scatterometer data.

Finally, the computational requirements of DDM assimilation were assessed. This could be a consideration when processing a larger dataset or including DDM observations into operational DA systems. Work needs to be done to streamline the implementation of this approach to improve computational efficiency.

6. OBSERVING SYSTEM EXPERIMENT FOR HURRICANE MODEL PREDICTION

6.1 Introduction

Accurate forecasts of tropical cyclones (TCs) including the track, structure and intensity are crucial to estimate the potential damage from wind and storm surge. Ocean surface wind observations have been shown to improve the forecasts of TCs as they can not only directly improve the NWP analyses and forecasts, but also indirectly improve the model parameterization and depiction of boundary layer [53]. Advances in NWP models have made improvements in TC forecasts over the past decades. However, accurately forecasting the intensity of TCs is still a challenge. The first reason is that most of the conventional spaceborne instruments (e.g., scatterometers and radiometers) using high frequency microwave are impacted by the intense rain attenuation in the TC's eyewall. The second reason is that conventional instruments are mostly in polar-orbit, which cannot provide frequent sampling in the TC's rapidly evolving stages [21]. Presently, wind observations in the inner core of TCs are collected by the NOAA P-3 aircraft using stepped frequency microwave radiometer (SFMR) and GPS dropwindsonde. Those instruments are expensive and can only provide limited sampling.

As the name implies, CYGNSS is specifically designed to provide frequent surface wind observations in the inner core of TCs. As described in section 1.3, The L-

Acknowledgement: Some work (assimilating CYGNSS baseline L2 winds, running the HWRF and calculating the forecast errors) in this chapter was done by the author's collaborators, Bachir Annane from the Cooperative Institute for Marine and Atmospheric Studies and Mark Leidner from Atmospheric and Environmental Research.

band signal used by CYGNSS and the constellation of eight (8) low-inclination small satellites address the issues above for conventional instruments. CYGNSS has two major mission science goals [102]:

1) *Measure ocean surface wind speed in most naturally occurring precipitating conditions, including those experienced in the TC eyewall.*

2) *Measure ocean surface wind speed in the TC inner core with sufficient frequency to resolve genesis and rapid intensification.*

Recent results show that CYGNSS wind observations can improve the forecast of TCs in track, intensity and structure [51–55]. As described in Chapter 5, one of the benefits of DDM assimilation is that it can provide a new type of wind observations at CYGNSS specular points. In this chapter, this new type of observations will be assimilated into the Hurricane Weather Research and Forecasting (HWRF) model for a specific TC. The results will be evaluated by an observing system experiment (OSE) and compared to the results of assimilating other types of CYGNSS observations. We will evaluate if the CYGNSS mission can meet the above science goals.

6.2 Case Selection

Hurricane Michael in 2018 was selected for this case study. Michael is a Category 5 hurricane. It formed on October 7, 2018 in the Atlantic Ocean and dissipated on October 16, 2018, after it made landfall in Florida, continental United States. Figure 6.1 and 6.2 show the track and intensity forecast of Michael starting from 12:00 UTC, October 7 by different hurricane forecast models vs. its accepted track from the National Hurricane Center (NHC) best track data. The best track data give "best" location (storm track) and intensity (central pressure or wind speed) estimates for TCs throughout the lifetime of each TC. The product is produced by post-event analysis using all available data related to the storm. Michael was still a tropical depression at its rapidly evolving stage on this day. It can be observed that all forecast models

including the operational HWRF had poor intensity forecasts for this hurricane. This is probably because the inner core of the hurricane was poorly observed by operational instruments and the depression was poorly organized at model initialization time. As CYGNSS can contribute information to the structure of the hurricane's nascent system by its high sampling frequency, it would be interesting to see if the assimilation of CYGNSS observation at the early stage of the hurricane can improve the hurricane forecasts.

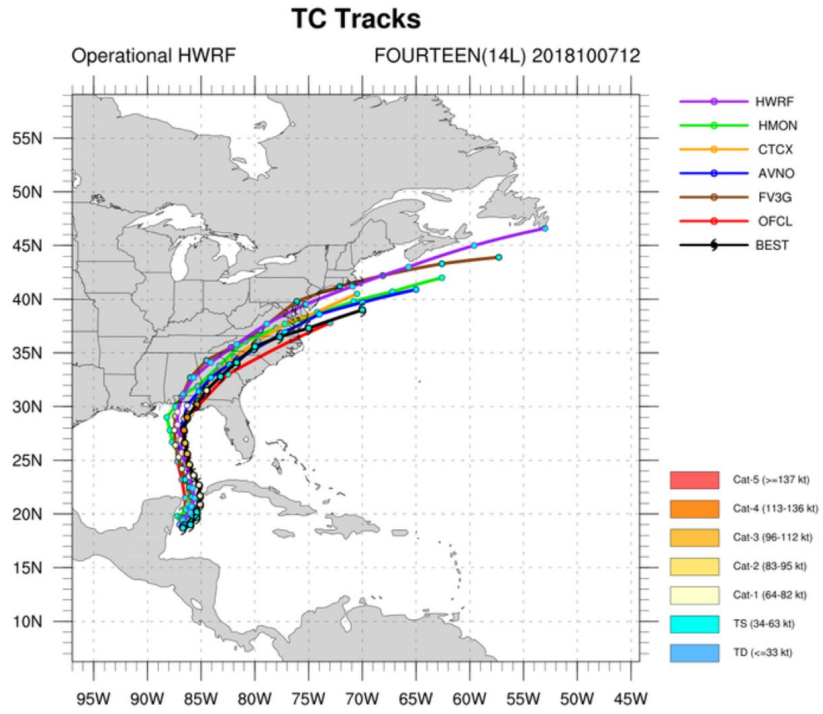
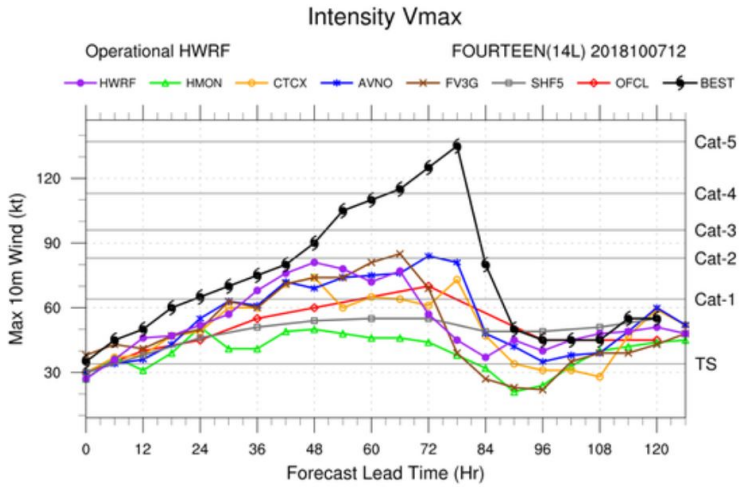
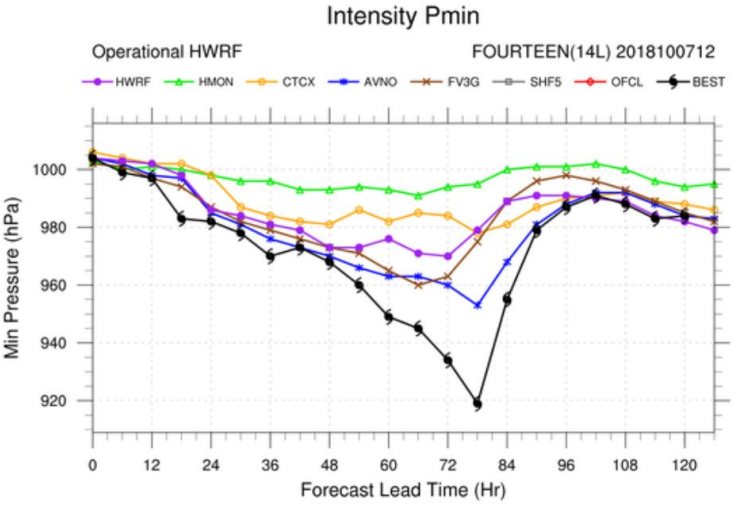


Figure 6.1.: Track forecasts of Hurricane Michael starting from 12:00 UTC, October 7 by different forecast models. The black line is the real track from the best track data.



(a)



(b)

Figure 6.2.: Intensity forecasts [maximum wind speed in (a) and minimum center pressure in (b)] of Hurricane Michael starting from 12:00 UTC, October 7 by different forecast models. Black lines are values from the best track data.

6.3 Experimental Design

6.3.1 Forecast Model

Since a global NWP model cannot resolve the small-scale information of the TC, a regional model with higher grid resolution and special parameterization is required to model the TC, especially for the rapid intensification process. The hurricane forecast model in this study is the 2019 version of the operational NCEP HWRF ("H219"), almost identical (with only bug fixes) to the version used to operationally forecast Michael in 2018. HWRF is a version of the Weather Research and Forecasting (WRF) model, specifically designed to model TCs. The model's initialization and lateral boundary conditions are from the operational Global Forecast System (GFS) analyses and forecasts (version FV3GFS). It uses a Hybrid 3D-Variational/Ensemble Kalman Filter data assimilation system in the Gridpoint Statistical Interpolation (GSI) framework. It is configured with a 14.5-km parent domain and two nested storm-following domain with 4.5-km and 1.5-km. The DA and model cycling perform every 6 hours, with each cycling producing a 5-day forecast [119].

6.3.2 Observing System Experiment

The OSE is an effective method to evaluate the impact of "new" observations on the NWP forecasts [120]. In an OSE, different observation sets are assimilated into a NWP model and the forecast results are compared. A flow chart of the OSE in this study is shown in Figure 6.3.

Four experiments using different sets of observations are run in parallel:

- **Control**: All operational observations (with no CYGNSS data) are assimilated into the HWRF.
- **CYG-L2**: CYGNSS Level 2 MV wind speed observations (a combination of NRBCS and LES FDS retrievals) are added to the Control experiment.

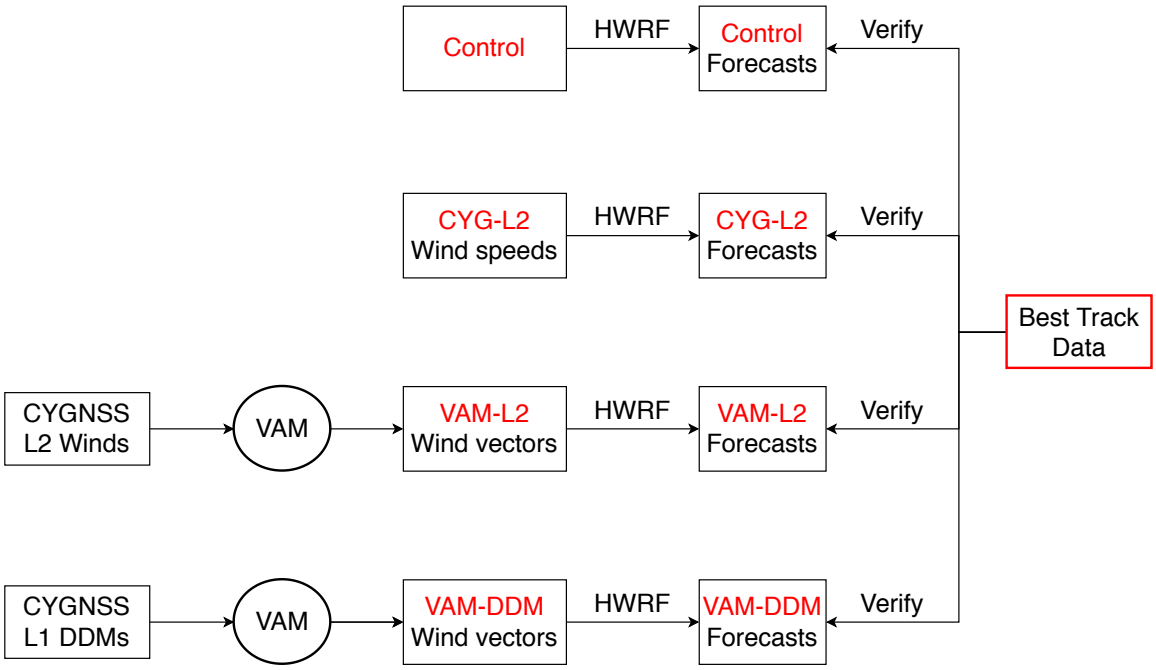


Figure 6.3.: Flowchart of the OSE. Four different experiments (Control, CYG-L2, VAM-L2, VAM-DDM) are run in parallel.

- **VAM-L2**: CYGNSS Level 2 MV winds are first assimilated by VAM with a similar implementation in [51]. Backgrounds of the VAM are from the ECMWF ERA5 reanalysis [105]. The VAM analyses are interpolated into the CYGNSS specular points to obtain the VAM-L2 wind vectors. The VAM-L2 wind vector observations are then added to the Control experiment.

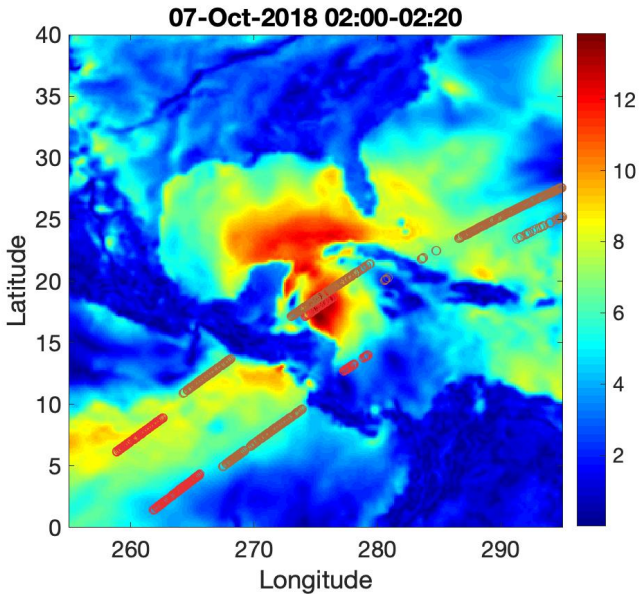
- **VAM-DDM**: CYGNSS Level 1 DDMs are first assimilated by VAM to obtain VAM-DDM wind vectors as described in Chapter 5. Backgrounds of the VAM are from the ECMWF ERA5. The implementation is almost the same as the one in Chapter 5. No bias correction is applied since the global model (background) usually underestimates wind speeds for a TC so that it cannot be used as the reference to correct observation bias. The VAM-DDM wind vectors are then added to the Control experiment.

As the CYGNSS specular point moves at about 6 km/s and the observation sampling frequency is 1 Hz, the spatial distance between each observation is about 6 km.

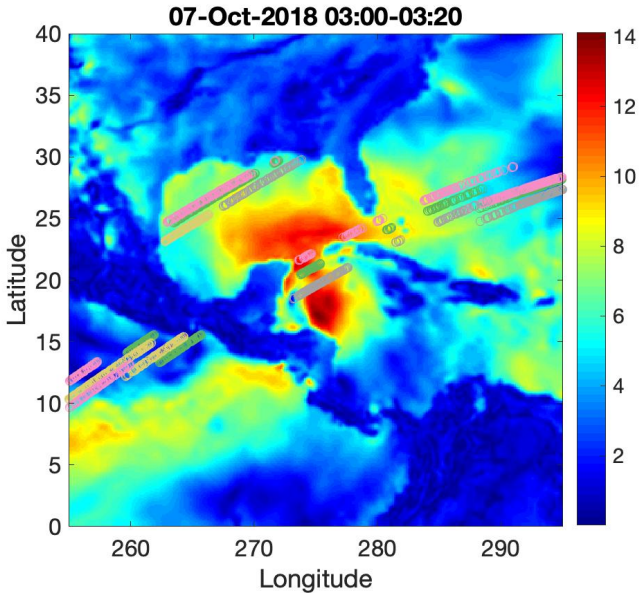
The spatial resolution of the CYGNSS L2 winds is 25 km and the footprint of the DDM is about 100 km. Therefore, there could be strong error correlations between observations in CYG-L2, VAM-L2 and VAM-DDM when they are assimilated. There are two approaches that can be used to avoid error correlations in DA. The first one is the thinning in which observations are thinned spatially. In this experiment, observations are thinned to one observation every 50 km before assimilated in HWRF. This discards more than 85% of the data. The second approach, no-thinning, uses all observations but inflates the observation error in DA. In this experiment, observation errors are inflated to be 3 times larger than they are in the thinning approach. A previous study has shown that the no-thinning approach has better results than the thinning approach for CYG-L2 and VAM-L2 observations [121]. In this study, the thinning and non-thinning approach for VAM-DDM observations will be compared.

Version 3 (sandbox163) of CYGNSS data (L1 DDM and L2 winds) from 21:00 UTC, August 6 to 9:00 UTC, August 8, within the region with latitude 0°–40° and longitude 255°–295° are used for the experiments. The time period was the early stage of the hurricane before it started the rapid intensification. The selected region is large enough to cover the whole structure and impact of the hurricane. Figure 6.4 shows the hurricane and CYGNSS observations (after the QC in section 3.4) in the region from two 20-minute time blocks. It shows CYGNSS can provide a great number of observations with sufficient frequency in the inner core of the storm.

In each experiment (Control, CYG-L2, VAM-L2 and VAM-DDM), observations are grouped by time at ±3-h time windows around the DA analysis times at 0000, 0600, 1200, 1800 UTC on August 7, 0000 and 0600 UTC on August 8. Note that there is no CYGNSS data in the ±3-h time window for the analysis time at 1800 UTC, August 7. Then all observations in each cycle are assimilated by the HWRF. With initialization times at those six (6) analysis times, the model will run to produce six (6) forecasts up to 96 hours in each experiment. Figure 6.5 shows a diagram of the forecast cycles. All forecasts from each experiment will be verified against the NHC best track data (HURDAT2).



(a)



(b)

Figure 6.4.: Wind field of Hurricane Michael and overpassed CYGNSS observations from two 20-minute time blocks. Colors stand for observations from different CYGNSS SVs.

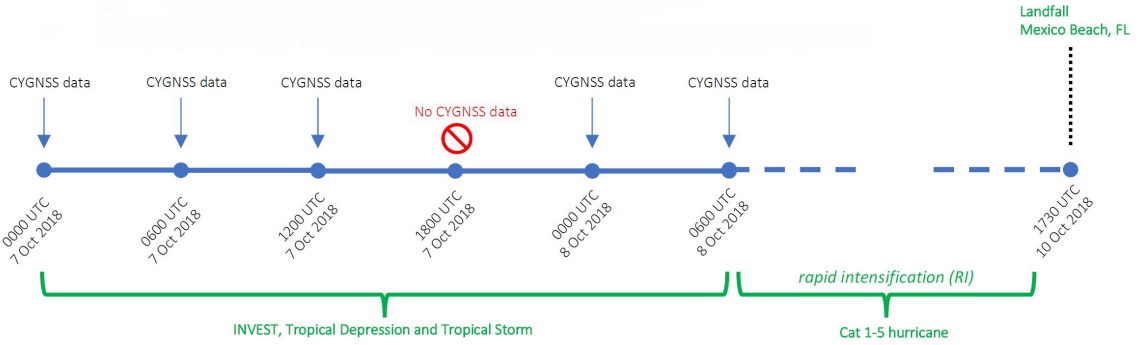


Figure 6.5.: Diagram of the forecast cycles.[1]

6.4 Results of VAM Analyses

This section presents the assessment for the results of VAM-DDM analyses. After the QC, around 30,000 DDMs are assimilated by VAM. Figure 6.6 presents a comparison of the wind speed distributions of VAM-DDM analyses and ECMWF backgrounds at CYGNSS specular points. Background wind speeds are significantly increased after assimilating the CYGNSS DDMs. It shows the ECMWF background could underestimate the wind speeds for the storm and CYGNSS observations could improve it.

Figure 6.7 shows the wind speed distribution of the CYGNSS L2 winds within the same area and time period. There are less DDM observations than the L2 winds because the QC on DDMs is more strict than on the L2 winds. The VAM-DDM wind speeds are also in general higher than the CYGNSS L2 wind speeds.

6.5 Results of Observing System Experiment

The results of OSE are presented in this section. Track and intensity forecast errors from all forecasts in each experiment (Control, CYG-L2, VAM-L2, VAM-DDM)

[1]Reference: Mark Leidner, Bachir Annane, Robert Atlas, Lidia Cucurull and Sharan Majumdar, "CYGNSS OSEs with HWRF: current results and plans", in *CYGNSS Science Team Telecon*, Oct. 2019.

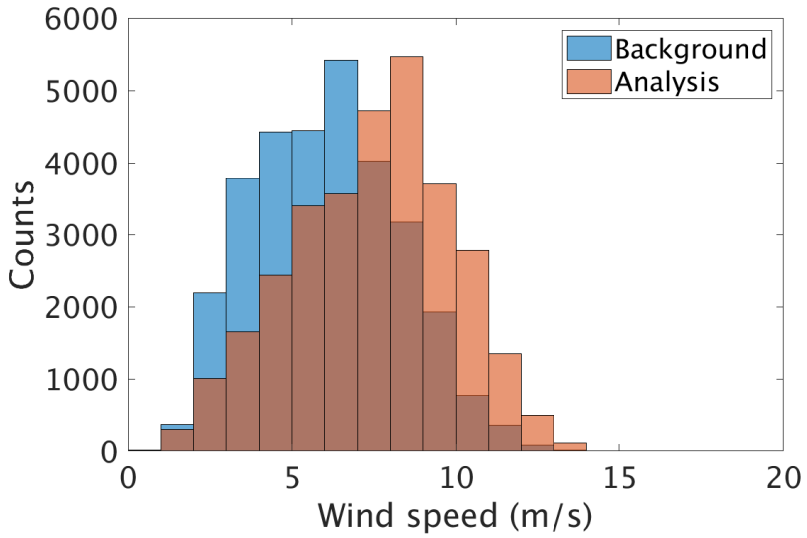


Figure 6.6.: Histograms for wind speeds of VAM-DDM analyses and ECMWF backgrounds within the area and time period described in section 6.3.2.

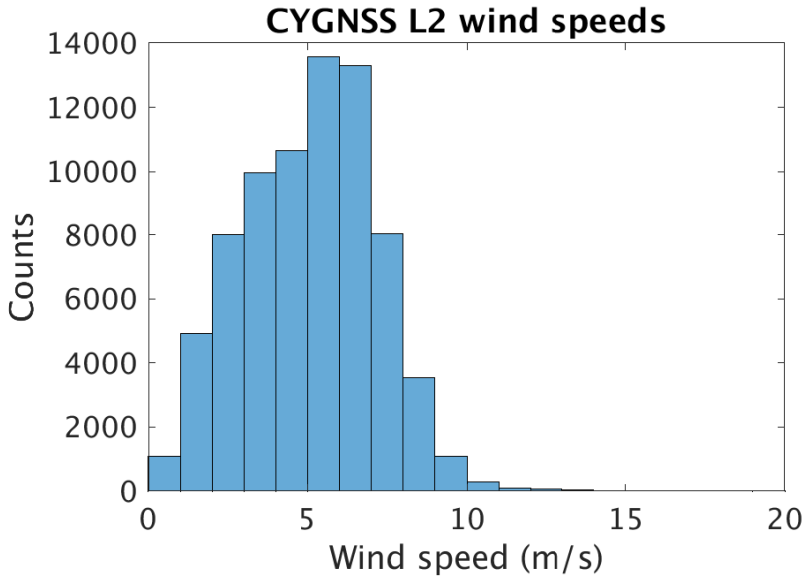


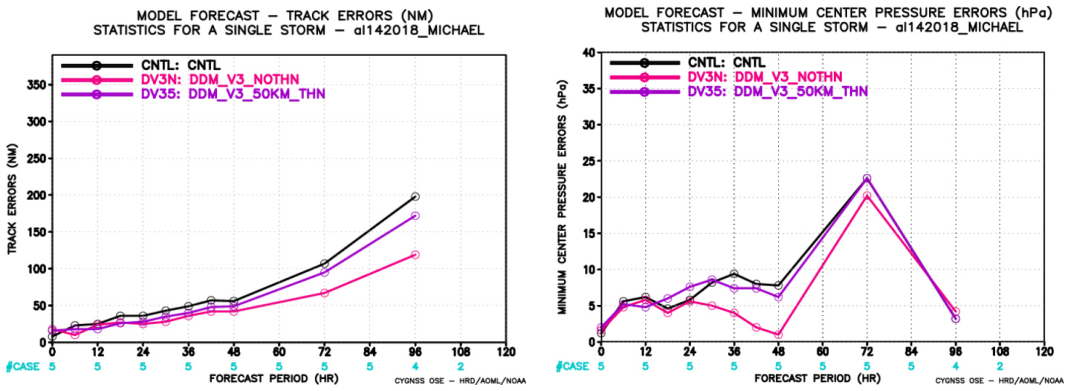
Figure 6.7.: Histograms for wind speeds of CYGNSS L2 winds within the area and time period described in section 6.3.2.

are computed. The metrics for track and intensity are storm center positions (nautical miles) and minimum center pressure (hPa), respectively. Then for each experiment,

the overall track and intensity errors are computed by averaging the errors of the six (6) forecasts. The overall errors from all experiments are then compared.

6.5.1 Thinning vs. No-thinning

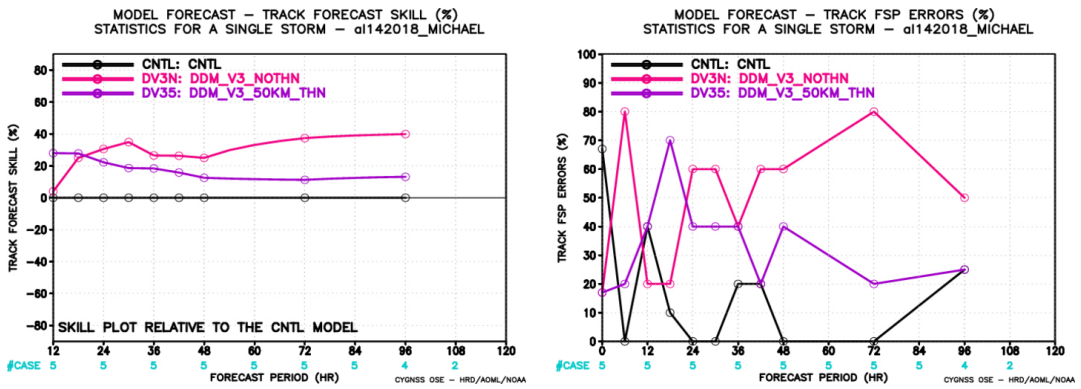
Forecast errors of Control, VAM-DDM with thinning and VAM-DDM with no-thinning are first compared. Figure 6.8 and 6.9 show the track error, minimum center pressure error, track forecast skill and track frequency of superior performance (FSP) of the three experiments. The #case in both figures stands for the number of independent forecast errors that have been averaged (5 by default). The #case at 96 hour is less than 5 because the storm decayed in one forecast.



(a) (b)

Figure 6.8.: Forecast errors of Hurricane Michael from Control (black), VAM-DDM with thinning (purple) and VAM-DDM with no-thinning (pink): (a) Storm center position (track) errors; (b) minimum center pressure (intensity) errors.

The forecast skill is assessed by comparing the performance of each experiment to that of the Control (in percent). The FSP is a measure (in percent) of how often one model produces a better forecast than another [122]. Results show that for VAM-DDM using all observation data with inflated errors (no-thinning) has a better performance than thinning the data in the forecasts of track and intensity. Therefore,



(a) (b)

Figure 6.9.: Track forecast skill (a) and track FSP (b) of Hurricane Michael from Control (black), VAM-DDM with thinning (purple) and VAM-DDM with no-thinning (pink).

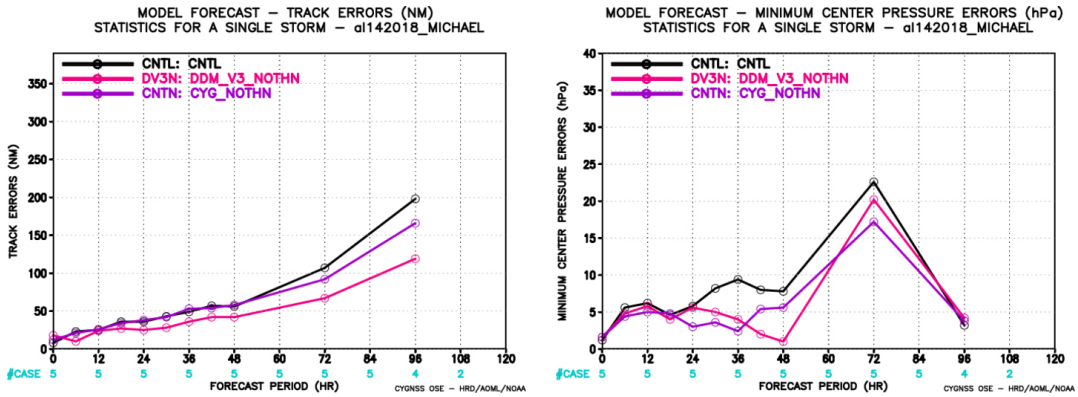
for the remainder of this study, the no-thinning configuration will be applied to the VAM-DDM case.

6.5.2 VAM-DDM vs. CYG-L2

Forecast errors of Control, VAM-DDM and CYG-L2 are compared. Figure 6.10 and 6.11 show the track error, minimum center pressure error, track forecast skill and track FSP of the three cases. Both VAM-DDM and CYG-L2 improve the forecasts compared to the Control. VAM-DDM has much better performance than CYG-L2 both in forecasts of both track and intensity.

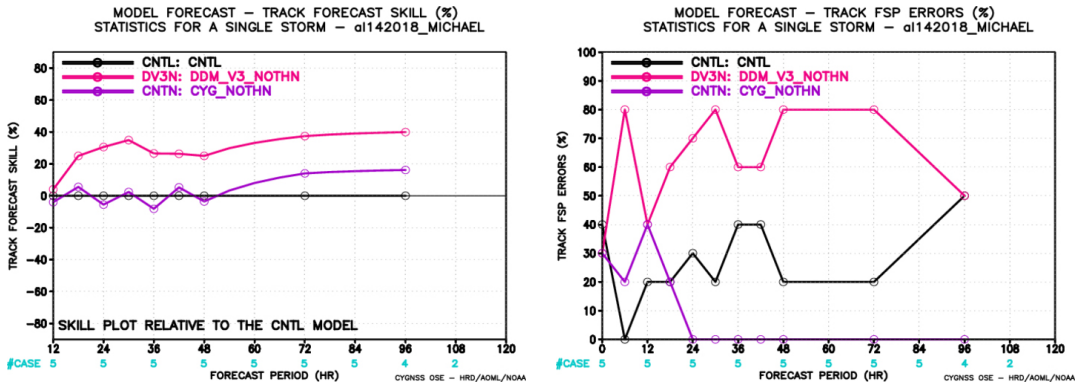
6.5.3 VAM-DDM vs. VAM-L2

Forecast errors of Control, VAM-DDM and VAM-L2 are compared. Figure 6.12 and 6.13 show the track error, minimum center pressure error, track forecast skill and track FSP of the three cases. Still, both VAM-DDM and CYG-L2 improve the forecasts compared to the Control. In the track and intensity errors (Figure 6.12), there is no significant difference between VAM-DDM and VAM-L2. In the track



(a) (b)

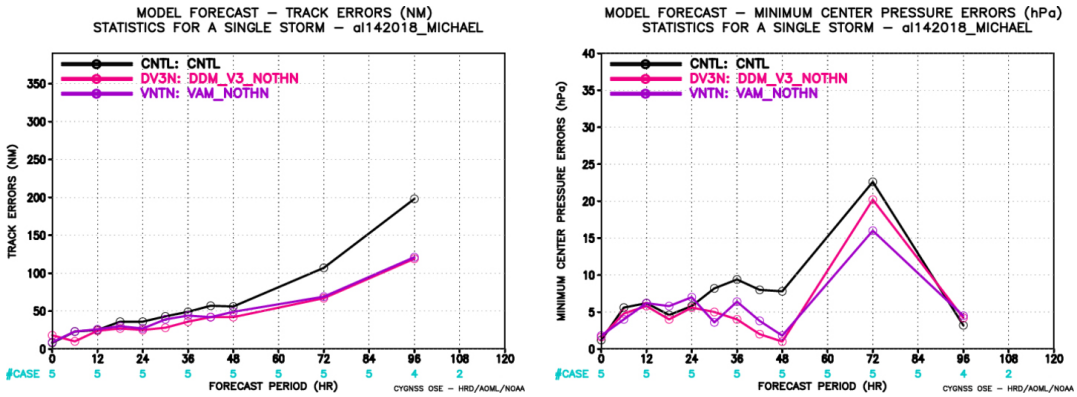
Figure 6.10.: Forecast errors of Hurricane Michael from Control (black), VAM-DDM (pink) and CYG-L2 with no-thinning (purple): (a) Storm center position (track) errors; (b) minimum center pressure (intensity) errors.



(a) (b)

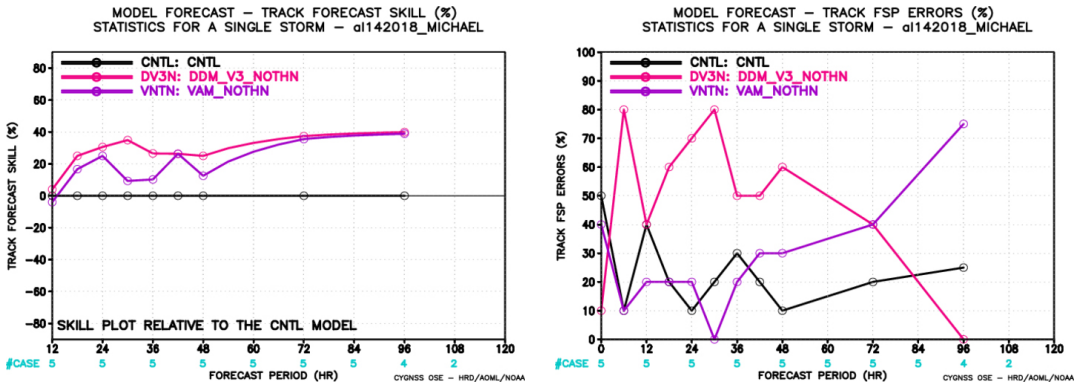
Figure 6.11.: Track forecast skill (a) and track FSP (b) of Hurricane Michael from Control (black), VAM-DDM (pink) and CYG-L2 (purple).

forecast skill and track FSP, results of VAM-DDM are slightly better than those of VAM-L2. The More investigations are required to compare VAM-DDM and VAM-L2 in future work.



(a) (b)

Figure 6.12.: Forecast errors of Hurricane Michael from Control (black), VAM-DDM (pink) and VAM-L2 with no-thinning (purple): (a) Storm center position (track) errors; (b) minimum center pressure (intensity) errors.



(a) (b)

Figure 6.13.: Track forecast skill (a) and track FSP (b) of Hurricane Michael from Control (black), VAM-DDM (pink) and VAM-L2 (purple).

6.6 Conclusion

An OSE was performed to evaluate the impact of assimilating wind vector observations derived from CYGNSS DDMs (VAM-DDM) on hurricane forecasts. The early stage (Oct. 6–8) of Hurricane Michael in 2018 was selected as the study case. HWRF was used as the DA system and forecast model. Results of assimilating VAM-DDM

observations were compared to the assimilation of other types of CYGNSS observations (CYG-L2 and VAM-L2) and also the Control (without CYGNSS data).

In general, the assimilation of CYGNSS DDMs increased the wind speeds of the VAM background from ECMWF ERA5. OSE results show that using all CYGNSS observations with inflated errors has better performance than thinning the data. The assimilation of CYGNSS observations, no matter the L1 DDMs or L2 winds, improves the HWRF forecasts both in track and intensity compared to control. Assimilating VAM derived wind vectors (VAM-L2, VAM-DDM) has better results than assimilating wind scalars (CYG-L2). More investigations are needed to compare the results of assimilating VAM derived wind vectors from L1 DDMs (VAM-DDM) and L2 winds (VAM-L2).

Only results from one hurricane are presented in this chapter. Experiments on more hurricanes are expected in the future to draw more generalized conclusions.

7. SUMMARY AND PERSPECTIVE

7.1 Summary

The baseline CYGNSS observables for wind retrieval utilize only the inner 15 samples of the fundamental measurement, DDM (with the dimension of 17 delays $\times$ 11 Dopplers), to form a 25 km resolution cell around the specular point. Prior studies have demonstrated DA using GNSS-R wind retrievals produced from DDMs. The complexity of the DDM dependence on winds, however, suggests that the alternative approach of directly ingesting DDM observables into DA systems, without performing a wind retrieval, may be beneficial. This dissertation demonstrated the approach for assimilating GNSS-R DDMs into weather models.

First, a forward model was presented to relate the DDM, in units of absolute power at the receiver, to the ocean surface wind field. This model and the related Jacobian were designed for use in the assimilation of DDMs. The forward model was assessed by comparing DDMs computed from HWRF model winds against measured DDMs from the CYGNSS mission. Limitations of the forward model were discussed, including the effects of swell, power bias, specular positioning error and physical limitations of the scattering theory.

Then, a two-dimensional variation analysis method (VAM) was developed to assimilate DDMs into an ocean surface wind vector field. A "track-wise" bias correction was proposed to correct the bias in the estimated transmitter EIRP and receiver antenna patterns. The observation quality control tests were presented. The impact of delay-Doppler ambiguity on the DDM assimilation was also discussed. Two methods to compute the DDM observation error covariance matrix were developed. One computes a simple diagonal matrix scaled with DDM magnitude. Another computes a

non-diagonal matrix considering the bin-bin correlations, empirically tuned to a large ensemble of CYGNSS observation data.

In the application of global data assimilation, DDM observations from the CYGNSS mission were assimilated into global ocean surface wind analyses. The background is from the ECMWF surface winds. Tuning of the VAM to the background's effective resolution was described. The comparison between using different DDM error covariance matrices found that the diagonal matrix performs as well as the fully populated matrix. Data assimilation results were assessed by Collocated scatterometer (ASCAT, OSCAT) winds. Results using one month (June 2017) of data show a reduction in the RMSE from 1.17 to 1.07 m/s and bias from -0.14 to -0.08 m/s for the wind speed at the specular point. Within a 150-km-wide swath along the specular point track, the RMSE was reduced from 1.20 to 1.13 m/s. Interpolated Winds at specular points from VAM analyses show smaller RMSE and bias than other CYGNSS wind products.

In the application of regional data assimilation, several types of CYGNSS observations were assimilated into the HWRF model to assess the impact of CYGNSS observations on hurricane forecast, in the case of Hurricane Michael (2018). The observations are CYGNSS L2 winds, VAM-derived wind vectors from CYGNSS L2 winds and VAM-derived wind vectors from CYGNSS L1 DDMs. Results show that the assimilation of CYGNSS observations improves the hurricane forecasts both in track and intensity. The assimilation of VAM-derived wind vectors has better performance than the assimilation of L2 wind speeds.

7.2 Recommendations for Future Work

7.2.1 Improvements on the Forward Model

The current forward model is based on the KA-GO model, Gaussian distribution of the isotropic surface slope and Katzberg empirical relation between wind speed and MSS. Those assumptions have their limitations and some of them have been

discussed in section 2.3.1. Currently we use quality controls on the wind speed to filter out cases for which those assumptions are not applicable. The bias correction in the data assimilation can also remove errors from the forward model to some extent. A more accurate and robust forward model can be developed if the following issues are considered, particularly in the operational application.

- **Use a bi-directional slope distribution:** The current forward model simplifies the surface slope distribution as an isotropic Gaussian distribution by assuming the wind direction to be 0° and using an omnidirectional MSS (equation (2.5)). This implementation simplifies the computation of the Jacobian. However, a more generalized 2-D anisotropic Gaussian distribution for the slope is described as

$$P(\boldsymbol{s}, u, v) = \frac{1}{2\pi\sqrt{\det(\boldsymbol{M})}} \exp\left(-\frac{1}{2}\boldsymbol{s}^T\boldsymbol{M}^{-1}\boldsymbol{s}\right) \tag{7.1}$$

where $\boldsymbol{M}$ is the second order moment matrix of the slope:

$$\boldsymbol{M} = \begin{bmatrix} \cos\phi_w & -\sin\phi_w \\ \sin\phi_w & \cos\phi_w \end{bmatrix} \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_c^2 \end{bmatrix} \begin{bmatrix} \cos\phi_w & \sin\phi_w \\ -\sin\phi_w & \cos\phi_w \end{bmatrix} \tag{7.2}$$

and $\boldsymbol{s} = -\vec{q}_\perp/q_z = [s_x \ s_y]^T$ is the slope vector. ϕ_w is the wind direction, σ_u^2 and σ_c^2 are MSS in the up-wind and cross-wind direction. σ_u^2 and σ_c^2 can be computed from the zonal and meridional wind components, u and v using the Katzberg model [17] or a wind wave spectrum [123]. The wind direction angle ϕ_w is defined as 0° when wind is blowing to the North and follows a clockwise increment:

$$\phi_w = \begin{cases} \text{atan2}(u, v) & \text{if } u \geq 0 \\ \text{atan2}(u, v) + 360 & \text{if } u < 0 \end{cases} \tag{7.3}$$

in which ϕ_w is in the range of [0,360). Note that in the isotropic model, σ_u^2 and σ_c^2 are assumed to be same as the omnidirectional MSS: $m = \sigma_u^2 = \sigma_u^2$.

Although it requires a more complicated Jacobian, using a bi-directional slope distribution is possible to obtain wind direction information from the GNSS-R observations, especially at areas away from the specular point [124, 125]. It is also more suitable for the data assimilation since the state of NWP models are wind vector components. Other non-Gaussian slope distributions such as the Gram-Charlier distribution [124] can also be considered.

- **Sea state impact:** What GNSS-R signals actually measure is the ocean surface roughness characterized by MSS. Different from the scatterometer operating in C and Ku bands which is sensitive to centimeter-scale capillary waves and responds to local wind instantaneously through the Bragg scattering, GNSS-R operating in L band with a bistatic geometry, is sensitive to both capillary and part of gravity waves with longer wavelength ($\sim$20 cm). The wind-wave relation in GNSS-R is more complicated. The current forward model uses the Katzberg empirical model for the wind-MSS relation. Other than the limitations discussed in Appendix B, the most important thing is that it does not consider any sea state information, including the non-local swell [126] and the degree of wave development [127]. The swell effect mainly occurs at low wind speeds (< 5 m/s) while it is negligible at medium-high wind speeds. The degree of wave development is characterized by the wave age and fetch length. In normal cases, the sea can be regarded as fully-developed with infinite fetch. However, in or near the hurricane there could be young waves and limited fetch.

  The swell effect can be modeled by a swell-driven wave spectrum [128] or an excess MSS approach using the IFREMER WaveWatch III (WW3) data [129]. The wave development can be modeled by the Elfouhaily wind-driven spectrum [123] which uses the inverse wave age as an input. The limitations of those methods are that they all need ancillary data and the integration of the wave spectrum could be computationally expensive.

- **Improved scattering model:** The KA-GO model for the bistatic radar equation has its limiting assumptions. It only models the large-scale surface roughness, assuming strong diffuse (noncoherent) scattering. In general, the KA-GO model is only valid for wind speed at 3–35 m/s and incidence angle less than 70° [39]. This is because at very low wind speeds the coherent scattering dominates and at very high wind speeds the ocean waves can break down. A two-scale model (TSM) that calculates large-scale roughness by KA-GO and small-scale roughness by small perturbation method (SPM) can be applied to incorporate small-scale information [130]. A more generalized model, SSA1, considers both the coherent and incoherent scattering [131, 132]. Figure 7.1 and 7.2 show the flow charts for two possible configurations of the improved forward model.

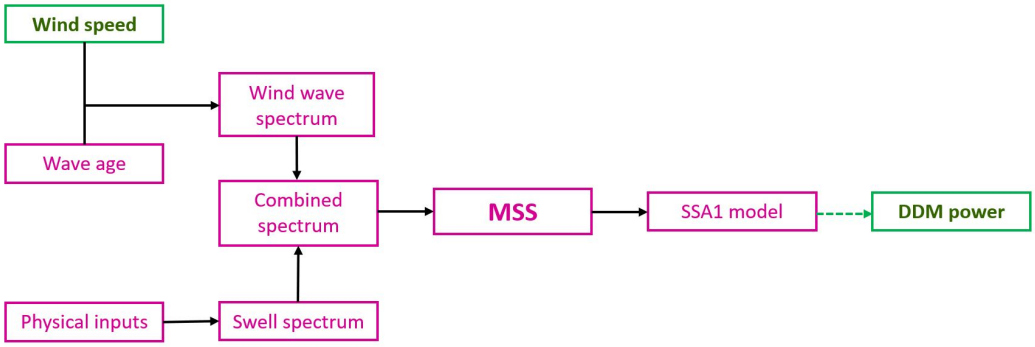


Figure 7.1.: Illustration of the forward model using the swell spectrum and SSA1 model.

7.2.2 Improvements on the EKF Wind Retrieval Algorithm

The advantages of the EKF wind retrieval method are that it utilizes the multi-look feature of GNSS-R observations and does not use GMFs that rely on some assumptions and ancillary data. Some updates on the EKF method have been implemented when applying it to the real CYGNSS data [133] including a new state on the latitude-longitude coordinates, updated receiver antenna patterns and using

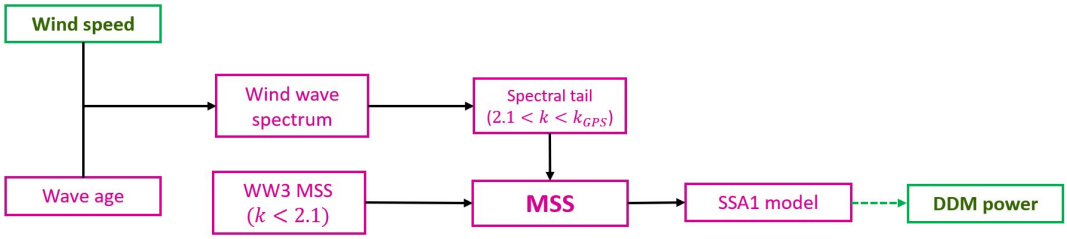


Figure 7.2.: Illustration of the forward model using the excess MSS method and SSA1 model.

parameters in the CYGNSS L1 data. First, all improvements on the forward model in the previous section 7.2.1 and the new wind-MSS model described in Appendix B can be considered to improve the EKF method. This section gives some additional recommendations for future improvements to the EKF method.

- **Number of iterations:** A point on the surface can be observed by more than 15 CYGNSS observed DDMs. Now the number of DDMs in each sequence is 7 which is determined by a simulation study [1]. A more optimal number of iterations can be obtained by a similar experiment using the real CYGNSS data.

- **Use delay-Doppler ambiguity-free observations:** As presented in [1], the EKF method is affected by the delay-Doppler ambiguity when there is a large wind variation in the observed area. Using DDM observations on the ambiguity-free line can potentially resolve this issue. A method for selecting DDM pixels on the ambiguity-free line has been presented in section 3.5.

- **Optimized error covariance matrices:** Now the EKF method uses diagonal matrices for the initial state and DDM observations. A non-diagonal DDM error covariance matrix is developed in Chapter 4. Although this covariance matrix does not improve results in the DDM assimilation mainly because the small-scale information of the DDM pixel-to-pixel correlations is smoothed out by the VAM, it has the potential to improve results of the EKF wind retrieval method

since the EKF method has no spatial smoothing and retrieves wind speeds at a very high spatial resolution. The covariance matrix for the initial state should also be optimized by a sensitivity experiment using the real CYGNSS data.

- **Kalman smoothing:** In the present implementation, the wind field is assumed to be constant in each sequence (7 seconds). If we consider the change of wind field during the time, it is promising to use Kalman smoothing rather than the Kalman filtering.

7.2.3 Improvements on the DDM Assimilation

The research in this dissertation mainly focuses on the feasibility and scientific application of the DDM assimilation. We use a simple 2D-VAR method to demonstrate the assimilation of CYGNSS DDMs into ocean surface wind analyses and show a positive impact. This 2D-VAR method does not include any time interpolation (e.g., First Guess at Appropriate Time (FGAT)) or running the NWP model to make forecasts. It is used as a pre-process to convert DDM observations into wind vectors before adding them to an operational DA system (shown in Chapter 6). This section gives some additional recommendations for future improvements to the DDM assimilation especially for the application of operational use. Still, all improvements on the forward model in section 7.2.1 can also be applied to improve the DDM assimilation.

- **Coupling with operational DA systems:** If any modern operational DA systems (e.g., 4D-VAR of ECMWF, GSI 3D-VAR of HWRF) want to add the GNSS-R DDM as one of their numerous observations, the DDM forward operator and its Jacobian have to be strongly coupled with the DA system. Main code of the forward model and VAM with some examples is open-source and has been published on Code Ocean [58]. The code can be educational use for implementing the DDM assimilation into an operational DA system. CYGNSS has a data latency of approximately 2 days, which could be an issue for operational use.

- **Generalized cost function:** Now the DDMs are sequentially assimilated due to the memory and computational cost. The wind field is updated every time when a DDM has been assimilated. The issue of this approach is that the observed area ($\approx$ 100 km) is less than the impact area ($\approx$ 150 km) of each DDM due to the constraint terms in the VAM. This can cause serious over-fitting issue and make it hard to characterize the error of DDM observations (now handled by the weight terms in the VAM). A more generalized method is to incorporate all DDM observations within the DA cycle into one cost function and assimilate them together, although the memory and computational cost might make this impossible. It is also possible to use a variational bias correction (VarBC) method, adding a bias uncertainty parameter into the state to better characterize the power uncertainty [134, 135].

7.2.4 Machine Learning Based Forward Model

The field of modern artificial intelligence (AI) including machine learning (ML), is expected to play a key role in future remote sensing and NWP. ML can improve dealing with larger observation data by increasing the accuracy and speed of computations [136, 137].

The main drawbacks of DDM assimilation are its complicated implementation and high computational cost because of the complex physically-based forward model. The high computational cost could make it difficult to implement DDM assimilation in an operational DA system. This limitation can be resolved by developing a machine learning (ML) emulation of the forward model. First, the most accurate physically based forward model can be used to produce a large training dataset where the inputs are geometry parameters, power parameters, wind speeds and the outputs are DDMs. Then an ML network can be trained from this dataset to emulate the forward problem. The Jacobian, which is the sensitivity of the outputs to some inputs (wind

speeds), is then easy to compute from the network.[1] Training an ML network is just a one-time cost and calculating the model is very efficient. With a large enough dataset, the ML model will have the ability to represent the physics in rare or unusual cases. Furthermore, the ML model is possible to resolve mechanisms that are poorly understood such as sea state impact and inaccurate transmitter power.

[1]Such as *tf.gradients* function in Tensorflow or *torch.autograd.grad* function in PyTorch.

REFERENCES

REFERENCES

[1] F. Huang, J. L. Garrison, N. Rodriguez-Alvarez, A. J. O'Brien, K. M. Schoenfeldt, S. C. Ho, and H. Zhang, "Sequential processing of GNSS-R delay-doppler maps to estimate the ocean surface wind field," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10 202–10 217, 2019.

[2] F. Huang, J. L. Garrison, S. M. Leidner, B. Annane, R. N. Hoffman, G. Grieco, and A. Stoffelen, "A forward model for data assimilation of GNSS ocean reflectometry delay-doppler maps," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[3] F. J. Wentz *et al.*, "Measurement of oceanic wind vector using satellite microwave radiometers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 5, pp. 960–972, 1992.

[4] S. H. Yueh, R. Kwok, F. Li, S. Nghiem, W. Wilson, and J. A. Kong, "Polarimetric passive remote sensing of ocean wind vectors," *Radio Science*, vol. 29, no. 04, pp. 799–814, 1994.

[5] P. W. Gaiser, K. M. St Germain, E. M. Twarog, G. A. Poe, W. Purdy, D. Richardson, W. Grossman, W. L. Jones, D. Spencer, G. Golba *et al.*, "The windsat spaceborne polarimetric microwave radiometer: Sensor description and early orbit performance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11, pp. 2347–2361, 2004.

[6] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave remote sensing: Active and passive. Volume 2-Radar remote sensing and surface scattering and emission theory*. Artech House, 1986.

[7] H. Hersbach, A. Stoffelen, and S. de Haan, "An improved C-band scatterometer ocean geophysical model function: CMOD5," *Journal of Geophysical Research: Oceans*, vol. 112, no. C3, 2007.

[8] F. M. Naderi, M. H. Freilich, and D. Long, "Spaceborne radar measurement of wind velocity over the ocean-an overview of the NSCAT scatterometer system," *Proceedings of the IEEE*, vol. 79, no. 6, pp. 850–866, 1991.

[9] D. L. Witter and D. B. Chelton, "A Geosat altimeter wind speed algorithm and a method for altimeter wind speed algorithm development," *Journal of Geophysical Research: Oceans*, vol. 96, no. C5, pp. 8853–8860, 1991.

[10] J. Horstmann, H. Schiller, J. Schulz-Stellenfleth, and S. Lehner, "Global wind speed retrieval from SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 10, pp. 2277–2286, 2003.

[11] V. U. Zavorotny, S. Gleason, E. Cardellach, and A. Camps, "Tutorial on remote sensing using GNSS bistatic radar of opportunity," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 4, pp. 8–45, 2014.

[12] M. Martin-Neira *et al.*, "A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry," *ESA journal*, vol. 17, no. 4, pp. 331–355, 1993.

[13] F. Soulat, M. Caparrini, O. Germain, P. Lopez-Dekker, M. Taani, and G. Ruffini, "Sea state monitoring using coastal GNSS-R," *Geophysical Research Letters*, vol. 31, no. 21, p. L21303, nov 2004.

[14] J. L. Garrison and S. J. Katzberg, "The application of reflected GPS signals to ocean remote sensing," *Remote Sensing of Environment*, vol. 73, no. 2, pp. 175–187, 2000.

[15] J. L. Garrison, A. Komjathy, V. U. Zavorotny, and S. J. Katzberg, "Wind speed measurement using forward scattered GPS signals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 1, pp. 50–65, 2002.

[16] E. Cardellach, G. Ruffini, D. Pino, A. Rius, A. Komjathy, and J. L. Garrison, "Mediterranean balloon experiment: Ocean wind speed sensing from the stratosphere, using GPS reflections," *Remote Sensing of Environment*, vol. 88, no. 3, pp. 351–362, 2003.

[17] S. J. Katzberg, O. Torres, and G. Ganoe, "Calibration of reflected GPS for tropical storm wind speed retrievals," *Geophysical Research Letters*, vol. 33, no. 18, 2006.

[18] N. Rodriguez-Alvarez, 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 Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 626–641, 2012.

[19] M. Clarizia, C. Gommenginger, S. Gleason, M. Srokosz, C. Galdi, and M. Di Bisceglie, "Analysis of GNSS-R delay-doppler maps from the UK-DMC satellite over the ocean," *Geophysical Research Letters*, vol. 36, no. 2, 2009.

[20] G. Foti, C. Gommenginger, P. Jales, M. Unwin, A. Shaw, C. Robertson, and J. Rosello, "Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission," *Geophysical Research Letters*, vol. 42, no. 13, pp. 5435–5441, 2015.

[21] C. Ruf, M. Unwin, J. Dickinson, R. Rose, D. Rose, M. Vincent, and A. Lyons, "CYGNSS: Enabling the future of hurricane prediction [remote sensing satellites]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 52–67, 2013.

[22] Weathernews and Axelspace, "Weathernews' meteorological and oceanographic satellite WNISAT-1R the first Japanese satellite to succeed in a GNSS-R mission," https://global.weathernews.com/news/11708/, publication date: 2018-02-14.

[23] C. Jing, X. Niu, C. Duan, F. Lu, G. Di, and X. Yang, "Sea surface wind speed retrieval from the first Chinese GNSS-R mission: Technique and preliminary results," *Remote Sensing*, vol. 11, no. 24, p. 3013, 2019.

[24] S. Gleason, C. S. Ruf, M. P. Clarizia, and A. J. O'Brien, "Calibration and unwrapping of the normalized scattering cross section for the cyclone global navigation satellite system," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2495–2509, 2016.

[25] S. Gleason, C. S. Ruf, A. J. O'Brien, and D. S. McKague, "The CYGNSS Level 1 calibration algorithm and error analysis based on on-orbit measurements," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 37–49, 2018.

[26] A. Komjathy, M. Armatys, D. Masters, P. Axelrad, V. Zavorotny, and S. Katzberg, "Retrieval of ocean surface wind speed and wind direction using reflected GPS signals," *Journal of Atmospheric and Oceanic Technology*, vol. 21, no. 3, pp. 515–526, 2004.

[27] J. F. Marchán-Hernández, N. Rodríguez-Álvarez, A. Camps, X. Bosch-Lluis, I. Ramos-Pérez, and E. Valencia, "Correction of the sea state impact in the L-band brightness temperature by means of delay-doppler maps of global navigation satellite signals reflected over the sea surface," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2914–2923, 2008.

[28] N. Rodriguez-Alvarez and J. L. Garrison, "Generalized linear observables for ocean wind retrieval from calibrated GNSS-R delay-Doppler maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 1142–1155, 2016.

[29] M. P. Clarizia, C. S. Ruf, P. Jales, and C. Gommenginger, "Spaceborne GNSS-R minimum variance wind speed estimator," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6829–6843, 2014.

[30] M. P. Clarizia and C. S. Ruf, "Bayesian wind speed estimation conditioned on significant wave height for GNSS-R ocean observations," *Journal of Atmospheric and Oceanic Technology*, vol. 34, no. 6, pp. 1193–1202, 2017.

[31] M. P. Clarizia and C. Ruf, "Statistical derivation of wind speeds from CYGNSS data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3955–3964, 2020.

[32] Y. Liu, I. Collett, and Y. J. Morton, "Application of neural network to GNSS-R wind speed retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9756–9766, 2019.

[33] M. Asgarimehr, I. Zhelavskaya, G. Foti, S. Reich, and J. Wickert, "A GNSS-R geophysical model function: Machine learning for wind speed retrievals," *IEEE Geoscience and Remote Sensing Letters*, 2019.

[34] J. Reynolds, M. P. Clarizia, and E. Santi, "Wind speed estimation from CYGNSS using artificial neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 708–716, 2020.

[35] C. S. Ruf, R. Atlas, P. S. Chang, M. P. Clarizia, J. L. Garrison, S. Gleason, S. J. Katzberg, Z. Jelenak, J. T. Johnson, S. J. Majumdar *et al.*, "New ocean winds satellite mission to probe hurricanes and tropical convection," *Bulletin of the American Meteorological Society*, vol. 97, no. 3, pp. 385–395, 2016.

[36] M. Unwin, P. Jales, P. Blunt, S. Duncan, M. Brummitt, and C. Ruf, "The SGR-ReSI and its application for GNSS reflectometry on the NASA EV-2 CYGNSS mission," in *Aerospace Conference, 2013 IEEE*. IEEE, 2013, pp. 1–6.

[37] M. P. Clarizia and C. S. Ruf, "Wind speed retrieval algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) mission," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4419–4432, 2016.

[38] C. S. Ruf and R. Balasubramaniam, "Development of the CYGNSS geophysical model function for wind speed," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018.

[39] C. Ruf, P. Chang, M. Clarizia, S. Gleason, Z. Jelenak, J. Murray, M. Morris, S. Musko, D. Posselt, D. Provost *et al.*, "CYGNSS handbook," *Ann Arbor, MI, USA: Michigan Publishing*, 2016.

[40] C. S. Ruf, S. Gleason, and D. S. McKague, "Assessment of CYGNSS wind speed retrieval uncertainty," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018.

[41] F. Bouttier and P. Courtier, "Data assimilation concepts and methods," *Meteorological training course lecture series. ECMWF*, vol. 718, p. 59, 2002. [Online]. Available: http://msi.ttu.ee/~elken/Assim_concepts.pdf

[42] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media, 2009.

[43] D. F. Parrish and J. C. Derber, "The National Meteorological Center's spectral statistical-interpolation analysis system," *Monthly Weather Review*, vol. 120, no. 8, pp. 1747–1763, 1992.

[44] D. M. Barker, W. Huang, Y.-R. Guo, A. Bourgeois, and Q. Xiao, "A three-dimensional variational data assimilation system for MM5: Implementation and initial results," *Monthly Weather Review*, vol. 132, no. 4, pp. 897–914, 2004.

[45] P. Courtier, J.-N. Thépaut, and A. Hollingsworth, "A strategy for operational implementation of 4D-Var, using an incremental approach," *Quarterly Journal of the Royal Meteorological Society*, vol. 120, no. 519, pp. 1367–1387, 1994.

[46] A. McNally, P. Watts, J. A. Smith, R. Engelen, G. Kelly, J. Thépaut, and M. Matricardi, "The assimilation of airs radiance data at ECMWF," *Quarterly Journal of the Royal Meteorological Society*, vol. 132, no. 616, pp. 935–957, 2006.

[47] M. Matricardi, F. Chevallier, G. Kelly, and J.-N. Thépaut, "An improved general fast radiative transfer model for the assimilation of radiance observations," *Quarterly Journal of the Royal Meteorological Society*, vol. 130, no. 596, pp. 153–173, 2004.

[48] L. Cucurull, J. Derber, and R. Purser, "A bending angle forward operator for global positioning system radio occultation measurements," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 1, pp. 14–28, 2013.

[49] H. Lievens, B. Martens, N. Verhoest, S. Hahn, R. Reichle, and D. G. Miralles, "Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates," *Remote Sensing of Environment*, vol. 189, pp. 194–210, 2017.

[50] T. Janjić, N. Bormann, M. Bocquet, J. Carton, S. Cohn, S. Dance, S. Losa, N. Nichols, R. Potthast, J. Waller *et al.*, "On the representation error in data assimilation," *Quarterly Journal of the Royal Meteorological Society*, vol. 144, no. 713, pp. 1257–1278, 2018.

[51] S. M. Leidner, B. Annane, B. McNoldy, R. Hoffman, and R. Atlas, "Variational analysis of simulated ocean surface winds from the Cyclone Global Navigation Satellite System (CYGNSS) and evaluation using a regional OSSE," *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 8, pp. 1571–1584, 2018.

[52] S. Zhang, Z. Pu, D. J. Posselt, and R. Atlas, "Impact of CYGNSS ocean surface wind speeds on numerical simulations of a hurricane in observing system simulation experiments," *Journal of Atmospheric and Oceanic Technology*, vol. 34, no. 2, pp. 375–383, 2017.

[53] B. Annane, B. McNoldy, S. M. Leidner, R. Hoffman, R. Atlas, and S. J. Majumdar, "A study of the HWRF analysis and forecast impact of realistically simulated CYGNSS observations assimilated as scalar wind speeds and as VAM wind vectors," *Monthly Weather Review*, vol. 146, no. 7, pp. 2221–2236, 2018.

[54] Z. Cui, Z. Pu, V. Tallapragada, R. Atlas, and C. S. Ruf, "A preliminary impact study of CYGNSS ocean surface wind speeds on numerical simulations of hurricanes," *Geophysical Research Letters*, vol. 46, no. 5, pp. 2984–2992, 2019.

[55] X. Li, J. R. Mecikalski, and T. J. Lang, "A study on assimilation of CYGNSS wind speed data for tropical convection during 2018 January MJO," *Remote Sensing*, vol. 12, no. 8, p. 1243, 2020.

[56] Y. Ying and F. Zhang, "Potentials in improving predictability of multiscale tropical weather systems evaluated through ensemble assimilation of simulated satellite-based observations," *Journal of the Atmospheric Sciences*, vol. 75, no. 5, pp. 1675–1698, 2018.

[57] S. M. Leidner, S. Majumdar, J. Hegarty, and B. D. McNoldy, "CYGNSS data impact on global analyses of ocean surface winds," in *100th American Meteorological Society Annual Meeting*. AMS, 2020.

[58] F. Huang, A. O'Brien, N. Rodriguez-Alvarez, and J. Garrison, "GNSS-R DDM forward model and variational analysis method for data assimilation," https://www.codeocean.com/capsule/1268465/tree/v2, June 2020.

[59] A. O'Brien, "End-to-End Simulator Technical Memo," *CYGNSS official documentation, University of Michigan*, no. 148-0123, 2014.

[60] V. U. Zavorotny and A. G. Voronovich, "Scattering of GPS signals from the ocean with wind remote sensing application," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 951–964, 2000.

[61] T. Elfouhaily, D. R. Thompson, and L. Linstrom, "Delay-doppler analysis of bistatically reflected signals from the ocean surface: theory and application," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 560–573, 2002.

[62] J. F. Marchán-Hernández, A. Camps, N. Rodríguez-Álvarez, E. Valencia, X. Bosch-Lluis, and I. Ramos-Pérez, "An efficient algorithm to the simulation of delay–doppler maps of reflected global navigation satellite system signals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2733–2740, 2009.

[63] G. Grieco, A. Stoffelen, M. Portabella, M. B. Rivas, W. Lin, and F. Fabra, "Quality control of delay-doppler maps for stare processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2990–3000, 2018.

[64] V. Tallapragada, L. Bernardet, M. K. Biswas, S. Gopalakrishnan, Y. Kwon, Q. Liu, T. Marchok, D. Sheinin, M. Tong, S. Trahan *et al.*, "Hurricane Weather Research and Forecasting (HWRF) model: 2015 scientific documentation," *HWRF Development Testbed Center Tech. Rep*, vol. 99, 2014.

[65] R. M. Yablonsky, I. Ginis, B. Thomas, V. Tallapragada, D. Sheinin, and L. Bernardet, "Description and analysis of the ocean component of NOAA's operational Hurricane Weather Research and Forecasting Model (HWRF)," *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 1, pp. 144–163, 2015.

[66] J. P. Cangialosi, "National hurricane center forecast verification report, 2017 hurricane season," *NOAA National Hurricane Center Tech. Rep*, 2017.

[67] J. Miranda, M. Vall-Ilossera, A. Camps, and N. Duffo, "Sea surface emissivity at L-band: Swell effects," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*, vol. 5. IEEE, 2002, pp. 2623–2625.

[68] M. P. Clarizia, "Investigating the effect of ocean waves on GNSS-R microwave remote sensing measurements," Ph.D. dissertation, University of Southampton, 2012.

[69] H. L. Tolman *et al.*, "User manual and system documentation of WAVEWATCH III TM version 3.14," *Technical note, MMAB Contribution*, vol. 276, p. 220, 2009.

[70] T. W. Group, "The WAM model–a third generation ocean wave prediction model," *Journal of Physical Oceanography*, vol. 18, no. 12, pp. 1775–1810, 1988.

[71] A. G. Voronovich and V. U. Zavorotny, "The transition from weak to strong diffuse radar bistatic scattering from rough ocean surface," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 11, pp. 6029–6034, 2017.

[72] T. Wang, C. S. Ruf, B. Block, D. S. McKague, and S. Gleason, "Design and performance of a GPS constellation power monitor system for improved CYGNSS l1b calibration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 26–36, 2018.

[73] V. U. Zavorotny and A. G. Voronovich, "Recent progress on forward scattering modeling for GNSS reflectometry," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. IEEE, 2014, pp. 3814–3817.

[74] P. Jales, "Spaceborne receiver design for scatterometric GNSS reflectometry," Ph.D, University of Surrey, 2012.

[75] J. Tye, P. Jales, M. Unwin, and C. Underwood, "The first application of stare processing to retrieve mean square slope using the SGR-ReSI GNSS-R experiment on TDS-1," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, pp. 4669–4677, 2016.

[76] M. Caparrini, O. Germain, F. Soulat, L. Ruffini, and G. Ruffini, "A system for monitoring a feature of a surface with broad swath and high resolution," p. 5, 2004.

[77] C. Li, W. Huang, and S. Gleason, "Dual antenna space-based GNSS-R ocean surface mapping: Oil slick and tropical cyclone sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 1, pp. 425–435, 2014.

[78] R. G. Brown, P. Y. Hwang *et al.*, *Introduction to random signals and applied Kalman filtering*. Wiley New York, 1992, vol. 3.

[79] P. S. Maybeck, *Stochastic models, estimation, and control*. Academic press, 1982, vol. 3.

[80] R. Hoffman, "SASS wind ambiguity removal by direct minimization," *Monthly Weather Review*, vol. 110, no. 5, pp. 434–445, 1982.

[81] R. N. Hoffman, "SASS wind ambiguity removal by direct minimization. part II: Use of smoothness and dynamical constraints," *Monthly Weather Review*, vol. 112, no. 9, pp. 1829–1852, 1984.

[82] R. Hoffman, S. Leidner, J. Henderson, R. Atlas, J. Ardizzone, and S. Bloom, "A two-dimensional variational analysis method for NSCAT ambiguity removal: Methodology, sensitivity, and tuning," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 5, pp. 585–605, 2003.

[83] R. Atlas, R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, and D. Gombos, "A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications," *Bulletin of the American Meteorological Society*, vol. 92, no. 2, pp. 157–174, 2011.

[84] J.-F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006.

[85] T. Wang, C. S. Ruf, B. Block, D. S. McKague, and S. Gleason, "Design and performance of a GPS constellation power monitor system for improved CYGNSS L1B calibration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 26–36, jan 2019.

[86] F. Huang, G. Grieco, and A. Stoffelen, "GNSS-R processing and NWP assimilation," *EUMETSAT OSI SAF*, 2019. [Online]. Available: http://www.osi-saf.org/sites/default/files/dynamic/page_with_files/file/osisaf_vs19_01_Huang_Report_v1.0.pdf

[87] F. Said, Z. Jelenak, J. Park, S. Soisuvarn, and P. S. Chang, "A 'track-wise'wind retrieval algorithm for the CYGNSS mission," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 8711–8714.

[88] A. Stoffelen and J. Vogelzang, "Wind bias correction guide," *EUMETSAT, Darmstadt, Germany, Tech. Rep. NWPSAF-KN-UD-007, v1.3*, 2018. [Online]. Available: http://projects.knmi.nl/publications/fulltexts/wind_bias_correction_guide_v1.3_def.pdf

[89] M. P. Clarizia and C. S. Ruf, "On the spatial resolution of GNSS reflectometry," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 8, pp. 1064–1068, 2016.

[90] V. Stoffelen and G. Marseille, "High resolution data assimilation guide," *EUMETSAT, Darmstadt, Germany, Tech. Rep. NWPSAF-KNUD-008, v1.2*, 2018. [Online]. Available: http://projects.knmi.nl/scatterometer/publications/pdf/high_resolution_data_assimilation_guide_1.2_def.pdf

[91] S. Gleason, J. Johnson, C. Ruf, and C. Bussy-Virat, "Characterizing background signals and noise in spaceborne GNSS reflection ocean observations," *IEEE Geoscience and Remote Sensing Letters*, 2019.

[92] S. Gleason, C. Gommenginger, and D. Cromwell, "Fading statistics and sensing accuracy of ocean scattered GNSS and altimetry signals," *Advances in Space Research*, vol. 46, no. 2, pp. 208–220, 2010.

[93] M. Martín-Neira, S. D'Addio, C. Buck, N. Floury, and R. Prieto-Cerdeira, "The PARIS ocean altimeter in-orbit demonstrator," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2209–2237, 2011.

[94] O. Germain and G. Ruffini, "A revisit to the GNSS-R code range precision," *GNSS-R'06 Workshop*, 2006.

[95] F. Martín, S. D'Addio, A. Camps, and M. Martín-Neira, "Modeling and analysis of GNSS-R waveforms sample-to-sample correlation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 5, pp. 1545–1559, 2014.

[96] J. L. Garrison, "A statistical model and simulator for ocean-reflected GNSS signals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6007–6019, 2016.

[97] W. Li, A. Rius, F. Fabra, E. Cardellach, S. Ribo, and M. Martín-Neira, "Revisiting the GNSS-R waveform statistics and its impact on altimetric retrievals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2854–2871, 2018.

[98] G. Desroziers, L. Berre, B. Chapnik, and P. Poli, "Diagnosis of observation, background and analysis-error statistics in observation space," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 613, pp. 3385–3396, 2005.

[99] J. A. Waller, S. P. Ballard, S. L. Dance, G. Kelly, N. K. Nichols, and D. Simonin, "Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics," *Remote Sensing*, vol. 8, no. 7, p. 581, 2016.

[100] M. Cordoba, S. L. Dance, G. Kelly, N. K. Nichols, and J. A. Waller, "Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 702, pp. 333–341, 2017.

[101] H. You, J. L. Garrison, G. Heckler, and V. U. Zavorotny, "Stochastic voltage model and experimental measurement of ocean-scattered GPS signal statistics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2160–2169, 2004.

[102] M. P. Clarizia, V. Zavorotny, and C. Ruf, "CYGNSS algorithm theoretical basis document Level 2 wind speed retrieval, document 148-0138," University of Michigan, Ann Arbor, MI, USA, Tech. Rep., Sep 2020. [Online]. Available: https://podaac-tools.jpl.nasa.gov/drive/files/allData/cygnss/L2/docs/148-0138-6_ATBD_L2_v3.0_Wind_Speed_Retrieval.pdf

[103] C. Zuffada, T. Elfouhaily, and S. Lowe, "Sensitivity analysis of wind vector measurements from ocean reflected GPS signals," *Remote Sensing of Environment*, vol. 88, no. 3, pp. 341–350, 2003.

[104] H. You, J. L. Garrison, G. Heckler, and D. Smajlovic, "The autocorrelation of waveforms generated from ocean-scattered GPS signals," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 78–82, 2006.

[105] ECMWF, "Copernicus climate change service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate," 2020. [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/home

[106] J. M. Tabeart, S. L. Dance, A. S. Lawless, N. K. Nichols, and J. A. Waller, "Improving the condition number of estimated covariance matrices," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 72, no. 1, pp. 1–19, 2020.

[107] R. Owens and T. Hewson, "ECMWF forecast user guide," *Reading: ECMWF*, 2018.

[108] OSI SAF/EARS Winds Team, "ASCAT wind product user manual, version 1.16," OSI SAF, 2019. [Online]. Available: https://projects.knmi.nl/scatterometer/publications/pdf/ASCAT_Product_Manual.pdf

[109] OSI SAF Winds Team, "ScatSat-1 wind product user manual, version 1.3," OSI SAF, 2018. [Online]. Available: http://projects.knmi.nl/scatterometer/publications/pdf/osisaf_cdop2_ss3_pum_scatsat1_winds.pdf

[110] A. Stoffelen, J. A. Verspeek, J. Vogelzang, and A. Verhoef, "The CMOD7 geophysical model function for ASCAT and ERS wind retrievals," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2123–2134, 2017.

[111] A. Verhoef, J. Vogelzang, and A. Stoffelen, *ScatSat-1 wind validation report*, OSI SAF, 2018. [Online]. Available: http://projects.knmi.nl/scatterometer/publications/pdf/osisaf_cdop3_ss3_valrep_scatsat1_winds.pdf

[112] C. Ruf and D. Twigg, "Algorithm theoretical basis document Level 1 & 2 trackwise corrected climate data record," University of Michigan, Ann Arbor, MI, USA, Tech. Rep., 2020. [Online]. Available: https://podaac-tools.jpl.nasa.gov/drive/files/allData/cygnss/L2/docs/148-0389-1_ATBD_Trackwise_Corrected_CDR.pdf

[113] W. C. Skamarock, "Evaluating mesoscale NWP models using kinetic energy spectra," *Monthly Weather Review*, vol. 132, no. 12, pp. 3019–3032, 2004.

[114] S. Abdalla, L. Isaksen, P. Janssen, and N. Wedi, "Effective spectral resolution of ECMWF atmospheric forecast models," *ECMWF Newsletter*, vol. 137, pp. 19–22, 2013. [Online]. Available: https://www.ecmwf.int/node/17358

[115] R. Singh, P. Kumar, and P. K. Pal, "Assimilation of Oceansat-2-scatterometer-derived surface winds in the weather research and forecasting model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1015–1021, 2011.

[116] P. Laloyaux, J.-N. Thépaut, and D. Dee, "Impact of scatterometer surface wind data in the ECMWF coupled assimilation system," *Monthly Weather Review*, vol. 144, no. 3, pp. 1203–1217, 2016.

[117] O. Tange, *GNU Parallel 2018*. Ole Tange, Mar. 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1146014

[118] O. C. i Nogues, D. Pascual, R. Onrubia, and A. Camps, "Advanced GNSS-R signals processing with GPUs," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1158–1163, 2020.

[119] M. Biswas, S. Abarca, L. Bernardet, I. Ginis, E. Grell, M. Iacono, E. Kalina, B. Liu, Q. Liu, T. Marchok, A. Mehra, K. Newman *et al.*, "Hurricane weather research and forecasting (HWRF) model: 2018 scientific documentation," *Developmental Testbed Center*, 2018. [Online]. Available: https://www.dtcenter.org/sites/default/files/community-code/hwrf/docs/scientific_documents/HWRFv4.0a_ScientificDoc.pdf

[120] R. Dumelow, "Overview of observing system experiments," in *ECMWF Seminar Proceedings. Seminar on Recent developments in data assimilation for atmosphere and ocean.* European Centre for Medium-Range Weather Forecasts, 2003.

[121] B. Annane, M. Leidner, B. McNoldy, R. Atlas, S. Majumdar, and R. Hoffman. Forecast impact experiments to optimize utilization of CYGNSS wind oservations. IGARSS 2019. [Online]. Available: https://bmcnoldy.rsmas.miami.edu/papers/ALMAMH_39IGARSS.pdf

[122] S. B. Goldenberg, S. G. Gopalakrishnan, V. Tallapragada, T. Quirino, F. Marks Jr, S. Trahan, X. Zhang, and R. Atlas, "The 2012 triply nested, high-resolution operational version of the hurricane weather research and forecasting model (HWRF): Track and intensity forecast verifications," *Weather and Forecasting*, vol. 30, no. 3, pp. 710–729, 2015.

[123] T. Elfouhaily, B. Chapron, K. Katsaros, and D. Vandemark, "A unified directional spectrum for long and short wind-driven waves," *Journal of Geophysical Research: Oceans*, vol. 102, no. C7, pp. 15 781–15 796, 1997.

[124] J. Park and J. T. Johnson, "A study of wind direction effects on sea surface specular scattering for GNSS-R applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 4677–4685, 2017.

[125] G. Zhang, D. Yang, Y. Yu, and F. Wang, "Wind direction retrieval using spaceborne GNSS-R in nonspecular geometry," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 649–658, 2020.

[126] D. Chen, "GNSS-R remote sensing of the ocean: Surface waves and related phenomena," Ph.D. dissertation, University of Michigan, 2016.

[127] R. Balasubramaniam and C. Ruf, "Azimuthal dependence of GNSS-R scattering cross-section in hurricanes," *Journal of Geophysical Research: Oceans*, vol. 125, no. 7, p. e2020JC016167, 2020.

[128] B. Li, L. Yang, B. Zhang, D. Yang, and D. Wu, "Modeling and simulation of GNSS-R observables with effects of swell," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020.

[129] T. Wang, V. U. Zavorotny, J. Johnson, C. Ruf, and Y. Yi, "Modeling of sea state conditions for improvement of CYGNSS L2 wind speed retrievals," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 8288–8291.

[130] J. C. Leader, "Incoherent backscatter from rough surfaces: the two-scale model reexamined," *Radio Science*, vol. 13, no. 3, pp. 441–457, 1978.

[131] A. G. Voronovich and V. U. Zavorotny, "Bistatic radar equation for signals of opportunity revisited," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 1959–1968, 2017.

[132] V. U. Zavorotny and A. G. Voronovich, "Validity of the kirchhoff-geometric optics approach for modeling of ocean bistatic radar scattering," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 668–671.

[133] H. Zhang, "New algorithms for ocean surface wind retrievals using multi-frequency signals of opportunity," Ph.D. dissertation, Purdue University Graduate School, 2019.

[134] D. Dee and S. Uppala, *Variational bias correction in ERA-Interim*. ECMWF Reading, UK, 2008. [Online]. Available: https://www.ecmwf.int/node/8936

[135] Y. Zhu, J. C. Derber, R. J. Purser, B. A. Ballish, and J. Whiting, "Variational correction of aircraft temperature bias in the NCEP's GSI analysis system," *Monthly Weather Review*, vol. 143, no. 9, pp. 3774–3803, 2015.

[136] S. Boukabara, V. Krasnopolsky, J. Stewart, S. Penny, R. Hoffman, and E. Maddy, "Artificial intelligence may be key to better weather forecasts," *Eos, 100*, 2019. [Online]. Available: https://doi.org/10.1029/2019EO129967

[137] S.-A. Boukabara, V. Krasnopolsky, J. Q. Stewart, E. S. Maddy, N. Shahroudi, and R. N. Hoffman, "Leveraging modern artificial intelligence for remote sensing and NWP: Benefits and challenges," *Bulletin of the American Meteorological Society*, vol. 100, no. 12, pp. ES473–ES491, 2019.

[138] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global positioning system: theory and practice*. Springer Science & Business Media, 2012.

[139] P. Jales and J. Rawlinson, "MERRByS product manual: GNSS reflectometry on TDS-1 with the SGR-ReSI," Surrey Satellite Technol. Ld. Guildford, UK, Tech. Rep., 2019. [Online]. Available: http://merrbys.co.uk/resources/documentation

[140] D. McKague and D. Starkenburg, "CYGNSS science data processing coordinate systems definitions," University of Michigan, Tech. Rep. UM Doc. 148-0336, 2016.

APPENDICES

A. COORDINATE SYSTEMS AND THE SURFACE GRID

This chapter summarizes the definition of coordinate systems, coordinate transformations and surface grid in the forward model from Chapter 2. The forward model uses the same coordinate systems in the CYGNSS E2ES [59].

Basics of Coordinate Transformation

Given two Cartesian coordinate systems A and B with the same origin, defined by unit vector sets $\hat{a}_1, \hat{a}_2, \hat{a}_3$ and $\hat{b}_1, \hat{b}_2, \hat{b}_3$, respectively, the rotation (transformation) matrix from A to B in the form of row vector is calculated as Table A.1. Note that $\cos(\hat{a}_i, \hat{b}_j) = \hat{a}_i \cdot \hat{b}_j$ stands for the cosine of the angle between the two vectors.

Table A.1.: Rotation matrix from coordinate systems A to B in the form of row vector

${}^{A}\boldsymbol{C}^{B}$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
$\hat{a}_1$	$\cos(\hat{a}_1, \hat{b}_1)$	$\cos(\hat{a}_1, \hat{b}_2)$	$\cos(\hat{a}_1, \hat{b}_3)$
$\hat{a}_2$	$\cos(\hat{a}_2, \hat{b}_1)$	$\cos(\hat{a}_2, \hat{b}_2)$	$\cos(\hat{a}_2, \hat{b}_3)$
$\hat{a}_3$	$\cos(\hat{a}_3, \hat{b}_1)$	$\cos(\hat{a}_3, \hat{b}_2)$	$\cos(\hat{a}_3, \hat{b}_3)$

For any vector, its coordinates in frame B can be computed from the coordinates in frame A by

$$[X_B \quad Y_B \quad Z_B] = [X_A \quad Y_A \quad Z_A]{}^{A}\boldsymbol{C}^{B}. \tag{A.1}$$

The rotation matrix from A to B in the form of column vector is the transpose of the one in the form of row vector, $\boldsymbol{R}_A^B = \left({}^A\boldsymbol{C}^B\right)^T$, giving

$$\begin{bmatrix} X_B \\ Y_B \\ Z_B \end{bmatrix} = \boldsymbol{R}_A^B \begin{bmatrix} X_A \\ Y_A \\ Z_Z \end{bmatrix}. \tag{A.2}$$

Vectors in the rest of this chapter will all be presented as column vectors.

If frame B is produced by rotating frame A along +X axis by angle θ, the rotation matrix from A to B is

$$\boldsymbol{R}_A^B = \boldsymbol{R}_X(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{A.3}$$

The rotation matrices along +Y axis and +Z axis are respectively given by

$$\boldsymbol{R}_Y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{A.4}$$

$$\boldsymbol{R}_Z(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{A.5}$$

Specular Frame and Surface Grid

Computation of the discretized forward model, equations (2.14) requires a wind field $\boldsymbol{x}'$ defined on a $n_r \times n_v$ grid of equally-spaced points distributed around the specular point called "surface frame" (SURF). The input wind field $\boldsymbol{x}$, however, will be in geodetic latitude and longitude coordinates on the Earth's surface. In this

section, we will define the practical implementation of a coordinate transformation and interpolation of the wind field between these two reference frames.

First, the specular point position on WGS-84 ellipsoid, $\vec{r}_s$, is computed and used to define the specular frame (SPEC) in Table A.2.

Table A.2.: Definition of the specular frame (SPEC)

Component	**Definition**
Origin	Center of the Earth
+X axis	Orthogonal to the z-axis and in the plane of the transmitter and receiver
+Y axis	Completes the right hand system
+Z axis	Normal to the Earth's surface at the specular point

When coordinates of the transmitter, receiver and specular point are given in the Earth-centered Earth Fixed (ECEF) reference frame (WGS-84), the unit vectors of the specular frame $\hat{x}_s$, $\hat{y}_s$, $\hat{z}_s$ can be computed as:

$$\hat{z}_s = \frac{\vec{r}_s}{|\vec{r}_s|} \tag{A.6}$$

$$\hat{x}_s = \frac{\vec{r}_{tr} - (\vec{r}_{tr} \cdot \hat{z}_s)\hat{z}_s}{|\vec{r}_{tr} - (\vec{r}_{tr} \cdot \hat{z}_s)\hat{z}_s|} \tag{A.7}$$

$$\hat{y}_s = \hat{z}_s \times \hat{x}_s \tag{A.8}$$

where $\vec{r}_{tr}$ is the position vector from transmitter to receiver in ECEF frame and $\vec{r}_s$ is the position vector of the specular point in ECEF frame. The transformation matrix from the ECEF frame to SPEC frame is therefore $\boldsymbol{R}_{ECEF}^{SPEC}$ given by

$$\boldsymbol{R}_{ECEF}^{SPEC} = [\hat{x}_s \quad \hat{y}_s \quad \hat{z}_s]^T. \tag{A.9}$$

Second, by approximating the Earth's surface near the specular point as a spherical surface, a curvilinear system called the surface frame (SURF) is defined by the angles (θ,ϕ). (θ,ϕ) stands for the intersection point between z-axis of the specular

frame and the Earth's surface after rotating SPEC in θ along x-axis and then ϕ along y-axis. The transformation matrix from the specular frame after rotation (SPEC') to SPEC $\boldsymbol{R}_{SPEC'}^{SPEC}$ is:

$$\boldsymbol{R}_{SPEC'}^{SPEC}(\theta,\phi) = [\boldsymbol{R}_Y(\phi)\boldsymbol{R}_X(\theta)]^T. \tag{A.10}$$

Coordinates of the point at (θ,ϕ) in the ECEF frame are given by

$$\begin{bmatrix} x(\theta,\phi) \\ y(\theta,\phi) \\ z(\theta,\phi) \end{bmatrix} = \boldsymbol{R}_{SPEC}^{ECEF}\boldsymbol{R}_{SPEC'}^{SPEC}(\theta,\phi) \begin{bmatrix} 0 \\ 0 \\ r_c \end{bmatrix} \tag{A.11}$$

where $\boldsymbol{R}_{SPEC}^{ECEF} = \left(\boldsymbol{R}_{ECEF}^{SPEC}\right)^T$; r_c is the local radius of curvature; $[0\ 0\ r_c]^T$ is the coordinate of (θ,ϕ) in SPEC'. The SPEC frame and SURF frame are shown in Fig. A.1.

The surface is discretized using $n_r \times n_v$ points in the surface frame with surface resolution of d. Surface frame coordinates (θ, ϕ) corresponding to the grid point (r, v) are defined as

$$\theta = \left(\frac{n_r}{2} - r\right)\left(\frac{d}{r_c}\right) \tag{A.12}$$

$$\phi = \left(\frac{n_v}{2} - v\right)\left(\frac{d}{r_c}\right) \tag{A.13}$$

for $r = 1, 2, ..., n_r$, $v = 1, 2, ..., n_v$. In the forward model, $n_r = n_v = 120$ and $d = 1$ km.

The input wind field in geodetic latitude and longitude coordinates is interpolated into the $n_r \times n_v$ discretized wind field to compute the forward model in (2.14) through the following procedure:

1. An $n_r \times n_v$ grid is generated around the specular point using (A.12) (A.13).
2. ECEF coordinates of each point in the grid are computed using (A.11).
3. Geodetic coordinates of each point in the grid are computed from the ECEF positions using a standard transformation [138].

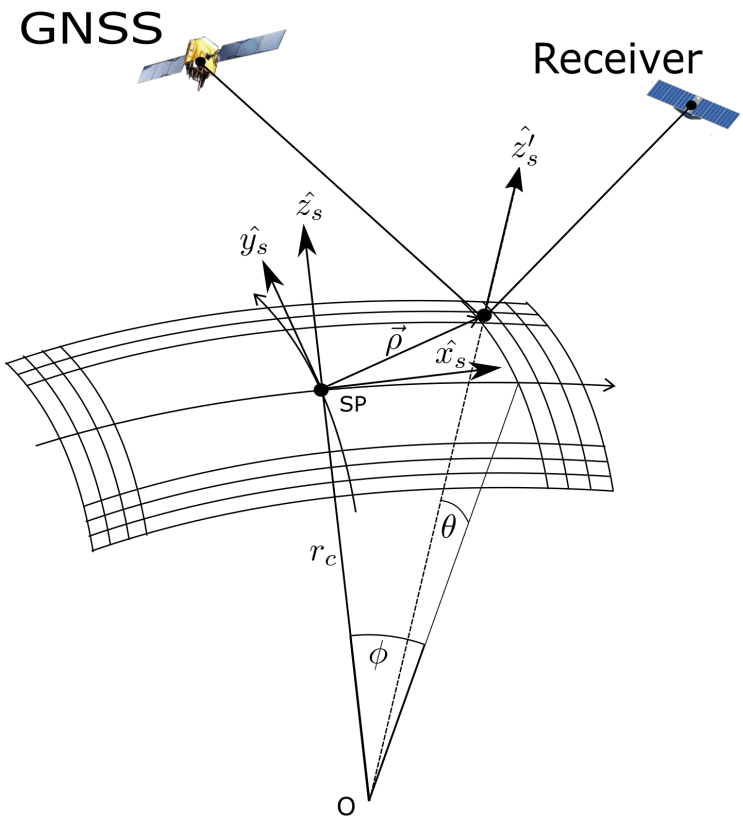


Figure A.1.: Illustration of the coordinate systems: O is the center of the Earth; r_c is the local radius of curvature at the specular point; $\hat{x}_s, \hat{y}_s, \hat{z}_s$ are unit vectors of the specular frame (SPEC); position of the scattering point in the surface frame (SURF) is defined by the rotation angles, θ and ϕ, with $\hat{z}_s{}'$ is $\hat{z}_s$ after rotating the SPEC. © 2020 IEEE.

4. The input wind field is then interpolated to the latitude and longitude of each of the grid points.

Orbit Frame Accounting for the Earth Rotation

The orbit frame (ORB) is defined by the position and velocity vector of the spacecraft at a given instant. The definition is listed in Table A.3.

Table A.3.: Definition of the orbit frame

Component	**Definition**
Origin	The spacecraft center position
+X axis	Along-track direction
+Y axis	Toward the negative orbit normal
+Z axis	Toward the center of the Earth

As the position and velocity of the spacecraft are provided in the non-inertial ECEF coordinates, a correction for the Earth rotation is needed [139]. The angular velocity of ECEF frame in the Earth-centered inertial frame is

$$\vec{w}_e = \begin{bmatrix} 0 \\ 0 \\ w_e \end{bmatrix} \tag{A.14}$$

where $w_e = 7.2921158553 \times 10^{-5}$ rad/s is the Earth rotation rate. The inertial velocity of the spacecraft is computed by the Basic Kinematic Equation (BKE)

$$\vec{v}_I = \vec{v} + \vec{w}_e \times \vec{p} \tag{A.15}$$

where $\vec{p}$ and $\vec{v}$ are position and velocity of the spacecraft in ECEF frame.

Unit vectors of the orbit frame $\hat{x}_o$, $\hat{y}_o$, $\hat{z}_o$ can be computed as:

$$\hat{y}_o = -\frac{\vec{p} \times \vec{v}_I}{|\vec{p} \times \vec{v}_I|} \tag{A.16}$$

$$\hat{z}_o = -\frac{\vec{p}}{|\vec{p}|} \tag{A.17}$$

$$\hat{x}_o = \hat{y}_o \times \hat{z}_o \tag{A.18}$$

The rotation matrix from ECEF frame to the orbit frame is therefore given by

$$\boldsymbol{R}_{ECEF}^{ORB} = [\hat{x}_o \quad \hat{y}_o \quad \hat{z}_o]^T. \tag{A.19}$$

The rotation matrix from the specular frame to orbit frame is then given by

$$\boldsymbol{R}_{SPEC}^{ORB} = \boldsymbol{R}_{ECEF}^{ORB}\boldsymbol{R}_{SPEC}^{ECEF}. \tag{A.20}$$

Body Frame

The body (vehicle) frame is produced by rotating the orbit frame determined by the spacecraft attitude (pitch, roll and yaw angles). Roll (ϕ) is rotation about the +X axis. Pitch (θ) is rotation about the +Y axis. Yaw (ψ) is rotation about the +Z axis. The definition of the body frame is listed in Table A.4.

Table A.4.: Definition of the body frame

Component	**Definition**
Origin	The spacecraft center position
+X axis	Toward the ram side of the spacecraft
+Y axis	Toward the starboard side of the spacecraft
+Z axis	Completes the right hand system

Figure A.2 shows the orbit frame and body frame. The transformation from the orbit frame to the body frame follows a 2-1-3 (pitch-roll-yaw) Euler angle rotation sequence [140]. The rotation matrix from the orbit frame to the body frame is given by

$$\boldsymbol{R}_{ORB}^{BODY} = \boldsymbol{R}_Z(\psi)\boldsymbol{R}_X(\phi)\boldsymbol{R}_Y(\theta). \tag{A.21}$$

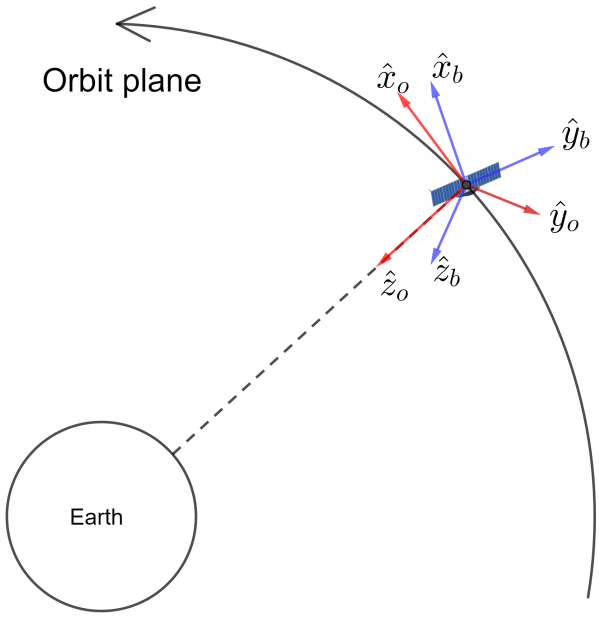


Figure A.2.: Illustration of the orbit frame and body frame. $\hat{x}_o, \hat{y}_o, \hat{z}_o$ are unit vectors of the orbit frame. $\hat{x}_b, \hat{y}_b, \hat{z}_b$ are unit vectors of the body frame.

Nadir Antenna Pattern

The orbit frame and body frame are used to calculate the receiver nadir antenna gain $G_r(\vec{\rho})$ for a point on the Earth surface. The receiver antenna patterns are expressed in the antenna spherical coordinates converted from the body frame as $G_R(\theta,\phi)$, where θ and ϕ are the elevation angle and azimuth angle of the target from the perspective of the spacecraft, respectively. The antenna spherical coordinates is shown in Figure A.3.

After creating the surface grid shown in section A, given the position vectors of the spacecraft and a grid point in the specular frame, the position vector of the point respect to the spacecraft in the body frame is calculated as

$$\vec{r} = \boldsymbol{R}_{ORB}^{BODY} \boldsymbol{R}_{SPEC}^{ORB} (\vec{r}_g - \vec{r}_R) \tag{A.22}$$

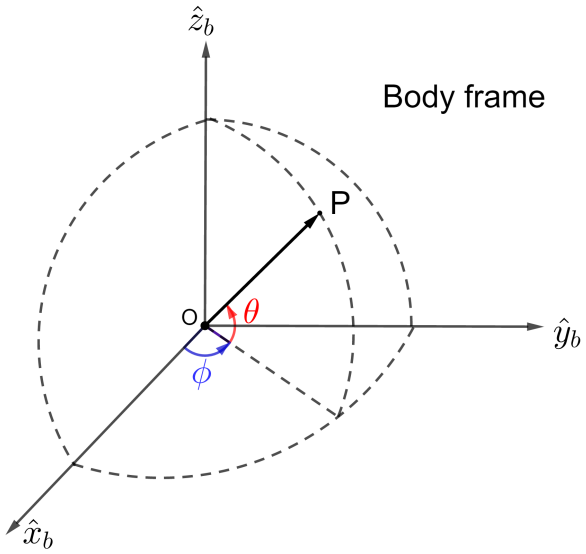


Figure A.3.: Illustration of the antenna spherical coordinates. P is the target. θ is the elevation angle and ϕ is the azimuth angle.

where $\vec{r}_R$ and $\vec{r}_g$ are the position vectors of the spacecraft and the grid point in the specular frame, $\boldsymbol{R}_{SPEC}^{ORB}$ is given in equation (A.20) and $\boldsymbol{R}_{ORB}^{BODY}$ is given in equation (A.21).

The angles of elevation θ and azimuth ϕ are then calculated by

$$\theta = \text{atan2}(r_z, \sqrt{r_x^2 + r_y^2}) \tag{A.23}$$

$$\phi = \text{atan2}(r_y, r_x) \tag{A.24}$$

where atan2(x, y) returns the angle between the ray to the point (x, y) and +X axis, confined to $(-\pi, \pi]$. The antenna gain of the point at $\vec{\rho}$ is then given by

$$G_r(\vec{\rho}) = G_R(\theta(\vec{\rho}), \phi(\vec{\rho})). \tag{A.25}$$

Figure A.4 shows an example of the CYGNSS antenna pattern.

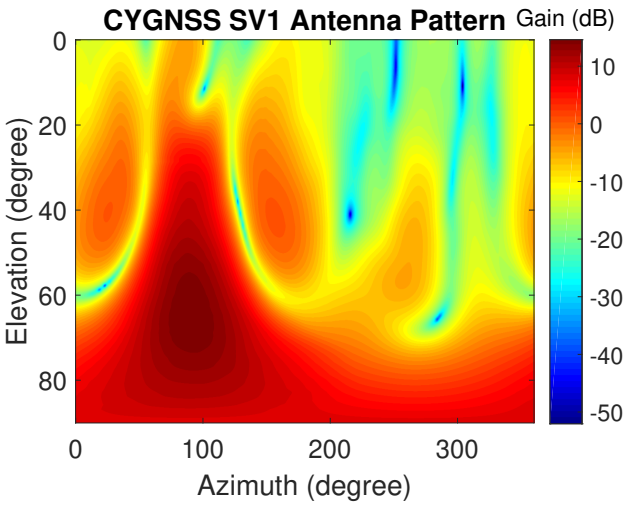


Figure A.4.: CYGNSS SV1 starboard antenna pattern (v6).

The CYGNSS receiver nadir antenna file format used in the forward model is shown in the table A.5. A MATLAB script for reading the CYGNSS antenna files is also provided as below.

Table A.5.: Antenna Data File Format

Data Type	**Definition**
double	Number of elevation samples N_{el}
double	Number of azimuth samples N_{az}
$N_{el} \times N_{az}$ repetitions of the following field	
double	Gain value (dB)
double	Elevation value (degree), -90–90
double	Azimuth value (degree), 0–360

```
fid = fopen('E2ES_antennas/V6/Rx1_starboard_E2ES_v6.bin');
antenna = fread(fid, 'double'); % 19456205=1801*3601*3+2
fclose(fid);
numEl = antenna(1); % 1801
numAz = antenna(2); % 3601

N = numEl * numAz; % 1801*3601
M = 3;
dataNum = N*M; % 1801*3601*3
data = antenna(3:end);

gain = data(1:N);
el_deg = data(N+1:2*N);
az_deg = data(2*N+1:3*N);

map = zeros(3601,1801);
for i = 1:3601
    map(i,:) = gain((i-1)*1801+1:i*1801);
end
elevation = -90:0.1:90;
azimuth = 0:0.1:360;

map1 = map(:,901:1801)';
figure;imagesc(azimuth,elevation(901:1801),map1);
h = colorbar;set(get(h,'title'),'string','Gain (dB)');
xlabel('Azimuth (degree)');ylabel('Elevation (degree)');
title('CYGNSS SV1 Antenna Pattern');
```

B. A MODIFIED WIND-MSS MODEL FROM CYGNSS GMF

Introduction

In the forward model, the normalized bistatic radar cross section (NBRCS) is expressed by assuming the sea surface slope PDF is in a Gaussian distribution. Combining equation (2.3) and (2.5), we can get

$$\sigma^0(m, \vec{q}(\vec{\rho})) = \frac{\pi \left|\Re(\vec{\rho})\right|^2 q^4}{q_z^4} \frac{1}{2\pi m} \exp\left(-\frac{q_\perp^2}{2mq_z^2}\right) \tag{B.1}$$

where m is the omnidirectional mean square slope (MSS) and $\vec{q}$ is the bisector vector determined by the location of the surface point $\vec{\rho}$. The geometry is shown in Figure B.1.

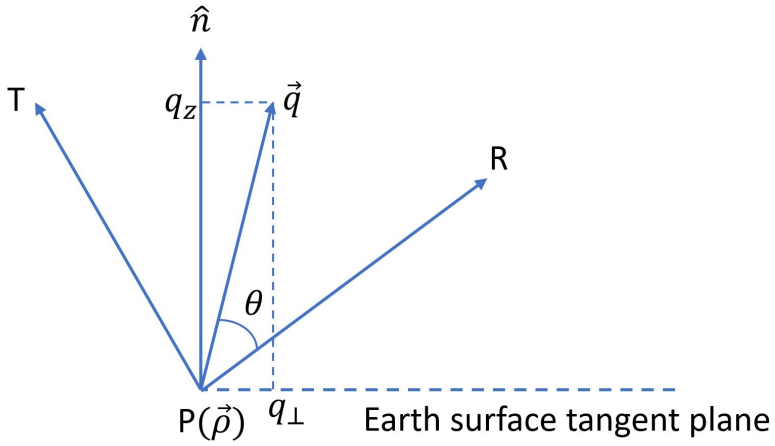


Figure B.1.: Illustration of the geometry. P is a point on the ocean surface; T is the transmitter; R is the receiver; $\vec{q}$ is the bisector; $\hat{n}$ is normal to the Earth surface; θ is the incidence angle at this point.

The MSS is computed from the wind speed monotonically by the Katzberg empirical model [17] expressed in equation (2.6). However, the Katzberg model was derived from GNSS-R data only at high elevation angles from an aircraft experiment over hurricanes with a very primitive receiver. The validity and applicability of the model are suspicious. In this section, a new wind-MSS empirical model will be derived from the CYGNSS Geophysical Model Functions (GMF). It has the potential to have better performance than the Katzberg model.

Generalized CYGNSS GMF

There are two observables for CYGNSS wind speed retrievals, the NBRCS (σ^0) and Leading Edge Slope (LES) which are computed from the average values of 3×5 box around the specular bin in the L1b DDM [37]. GMF is a function between an observable and the observed wind speed. GMFs were developed for the CYGNSS mission by matchups between the observables and reference wind speeds. The GMF for NBRCS is used here. In the GMF, the NBRCS is a function of the wind speed and incidence angle at the specular point as

$$g = g(u, \theta_{sp}) \tag{B.2}$$

where g is the GMF value for the NBRCS, u is the wind speed and θ_{sp} is the incidence angle at the specular point.

The GMF also has two versions, a fully developed sea (FDS) version and a young sea/limited fetch (YSLF) version. The FDS version was developed by matchups with global NWP model winds and the YSLF version was developed by matchups with aircraft measurements over hurricanes [38]. At each incidence angle, the FDS and YSLF models are piece-wise functions of the wind speed as

$$g_{fds} = \begin{cases} a_0 + a_1 u^{-1} + a_2 u^{-2} & \text{if } u \leq u_{fds} \\ b_0 + b_1 u + b_2 u^2 & \text{if } u > u_{fds} \end{cases} \tag{B.3}$$

$$g_{yslf} = \begin{cases} c_0 + c_1 u^{-1} + c_2 u^{-2} & \text{if } u \leq u_{yslf} \\ d_0 + d_1 u & \text{if } u > u_{yslf} \end{cases} \tag{B.4}$$

where $a_0, a_1, a_2, b_0, b_1, b_2, c_0, c_1, c_2, d_0, d_1$ are coefficients, u_{fds} and u_{yslf} are transition points between the two curves in each version.

Since the FDS version is mainly for low-to-medium wind speeds and YSLF version is mainly for high wind speeds, a generalized GMF model that combines the FDS and YSLF versions is derived by transferring between the two versions at 10-20 m/s in order to get a continuous function

$$g = \begin{cases} g_{fds} & u \leq 10 \\ \frac{20-u}{10} g_{fds} + \frac{u-10}{10} g_{yslf} & 10 < u \leq 20 \\ g_{yslf} & u > 20 \end{cases} \tag{B.5}$$

where $g = g(u, \theta_{sp})$ is the generalized CYGNSS GMF. Figure B.2 shows the generalized CYGNSS GMF.

Modified Wind-MSS Model

The CYGNSS GMF is derived for wind speed retrievals at the specular point and thus can only be applied to specular-point geometry. At the specular point, $q = q_z$ and $q_\perp = 0$. Then equation (B.1) becomes

$$\sigma^0 = \frac{|\Re(\theta_{sp})|^2}{2m}. \tag{B.6}$$

Substituting the σ^0 in equation (B.6) by the generalized CYGNSS GMF $g(u, \theta_{sp})$, a modified wind-MSS model from the CYGNSS GMF (CYGNSS model) can be derived as

$$m(u, \theta_{sp}) = \frac{|\Re(\theta_{sp})|^2}{2g(u, \theta_{sp})}. \tag{B.7}$$

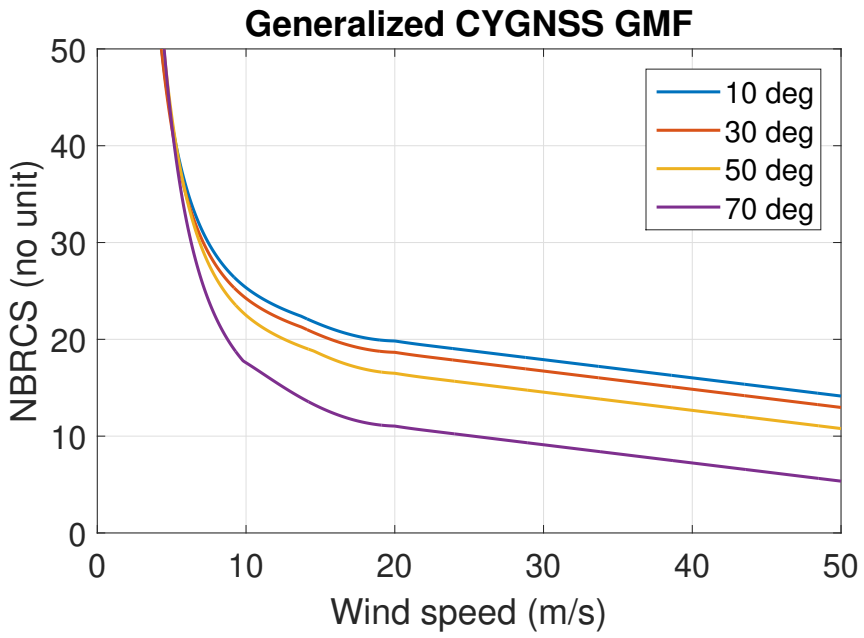


Figure B.2.: Generalized CYGNSS GMF (wind speed vs. NBRCS at different incidence angles).

For a surface point not at the specular point, a more generalized model will be

$$m(u,\theta,\theta_{sp}) = \frac{|\Re(\theta)|^2}{2g(u,\theta_{sp})} \tag{B.8}$$

where θ is the incidence angle (defined by the bisector) at the surface point.

Figure B.3 shows a comparison between different wind-MSS models: Katzberg model, CYGNSS model and integration of the Elfouhaily wind wave spectrum [123] at a cutoff wavenumber [15]. It can be observed that the CYGNSS model at low incidence angle (high elevation) matches more with the Katzberg model than at large incidence angle.

In the forward model, in order to compute the Jacobian, the derivative of MSS respect to wind speed is also required, similar to equation (2.22) for the Katzberg model. The derivative is given by

$$\frac{dm}{du} = -\frac{|\Re|^2}{2}\frac{1}{g^2}g' \tag{B.9}$$

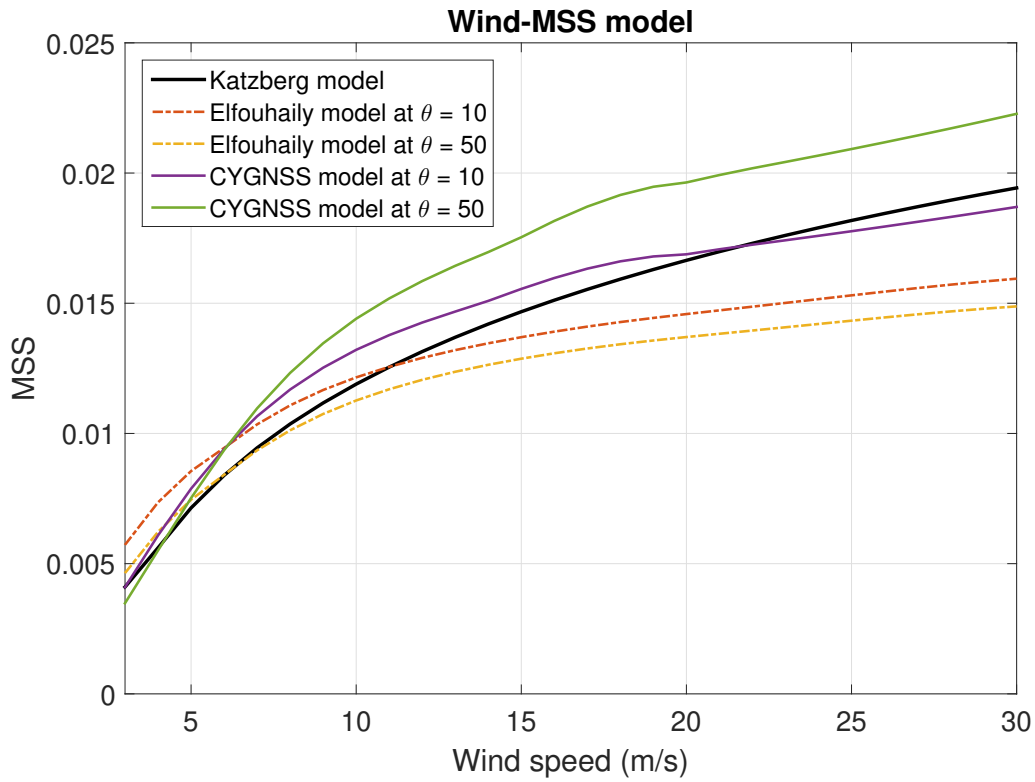


Figure B.3.: A comparison between different wind-MSS models: Katzberg model, CYGNSS model and integration of the Elfouhaily spectrum (Elfouhaily model). θ is the incidence angle.

where

$$g' = \frac{dg}{du} = \begin{cases} g'_{fds} & u \leq 10 \\ -0.1g_{fds} + \frac{20-u}{10}g'_{fds} + 0.1g_{yslf} + \frac{u-10}{10}g'_{yslf} & 10 < u \leq 20 \\ g'_{yslf} & u > 20 \end{cases} \tag{B.10}$$

in which

$$f'_{fds} = \begin{cases} -a_1 u^{-2} - 2a_2 u^{-3} & u \leq u_{fds} \\ b_1 + 2b_2 u & u > u_{fds} \end{cases} \tag{B.11}$$

and

$$f'_{yslf} = \begin{cases} -c_1 u^{-2} - 2c_2 u^{-3} & u \leq u_{yslf} \\ d_1 & u > u_{yslf} \end{cases} \tag{B.12}$$

Model Assessment

An experiment was conducted to compare the performance of the CYGNSS model and Katzberg model. From the collocation dataset between ECMWF, SCAT and CYGNSS described in Chapter 5, a dataset with ~100,000 DDMs was collected in which the root mean square difference between ECMWF and SCAT for each CYGNSS track is less than 0.8 m/s. For this dataset, the ECMWF wind field is regarded as the unbiased ground truth. Then the simulated DDMs computed by the forward model using the two wind-MSS models are compared to the CYGNSS observed DDMs for the dataset. The relative power difference ϵ described in equation (2.27) is used as the metric to compare the two DDMs. The mean of ϵ and mean of the $|\epsilon|$ for the two models are computed for the entire dataset. The results are listed in Tabel B.1.

Table B.1.: Assessment of the Katzberg model and CYGNSS model

	mean ϵ	mean $\|\epsilon\|$
Katzberg model	-0.198	0.249
CYGNSS model	-0.047	0.166

From the values of the mean of ϵ, it is observed that the Katzberg model has a systematic bias, while the CYGNSS model has almost no bias. The mean of $|\epsilon|$ from the two models also show that the CYGNSS model is more accurate than the Katzberg model.

Results of Data Assimilation

This CYGNSS model was applied to the global data assimilation experiment described in chapter 5. Same as before, the best observation weight λ_{ddm} was found by a sensitivity test and the wind speed RMSE from 5 days data is computed. The results are listed in Table B.2.

Table B.2.: Data assimilation results using the Katzberg model and CYGNSS model

	Best λ_{ddm}	5-day RMSE (m/s)
Background		1.18
Katzberg model	1/4	1.03
CYGNSS model	1/16	1.06

Nevertheless, it is found that using the new wind-MSS model does not improve the data assimilation results. The possible reason is that the difference between the Katzberg model and CYGNSS model is compensated by the bias correction and finding the optimal observation weight in the VAM.

The GMF models used in this section are the ones for CYGNSS version 2.1 science data products. Improved GMF models from CYGNSS future versions are expected in the future. It is also recommended to validate this new wind-MSS model in the EKF wind speed retrieval algorithm in the future work.

C. PUBLICATIONS AND CONFERENCE PRESENTATIONS

Journal Publications

Huang, Feixiong, James L. Garrison, Mark Leidner, Giuseppe Grieco, Ad Stoffelen, Bachir Annane, and Ross Hoffman. "Assimilation of GNSS Reflectometry Delay-Doppler Maps into Global Ocean Surface Wind Analyses." *Quarterly Journal of the Royal Meteorological Society* (2020). In review.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, Ross Hoffman, Giuseppe Grieco, and Ad Stoffelen. "A Forward Model for Data Assimilation of GNSS Ocean Reflectometry Delay-Doppler Maps." *IEEE Transactions on Geoscience and Remote Sensing* (2020). URL: https://doi.org/10.1109/TGRS.2020.3002801

Huang, Feixiong, James L. Garrison, Nereida Rodriguez-Alvarez, Andrew J. O'Brien, Kaitie M. Schoenfeldt, Soon Chye Ho, and Han Zhang. "Sequential Processing of GNSS-R Delay-Doppler Maps to Estimate the Ocean Surface Wind Field." *IEEE Transactions on Geoscience and Remote Sensing* (2019). URL: https://doi.org/10.1109/TGRS.2019.2931847

Preprints

Huang, Feixiong, James L. Garrison, Mark Leidner, Giuseppe Grieco, Ad Stoffelen, Bachir Annane, and Ross Hoffman. "Assimilation of GNSS Reflectometry Delay-Doppler Maps into Global Ocean Surface Wind Analyses." *EathArXiv*. 2020. URL: https://doi.org/10.31223/X5DS36

Conference Proceeding Publications

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, Giuseppe Grieco, Ad Stoffelen and Ross Hoffman. "Assimilation of GNSS-R Delay-Doppler Maps into Weather Models." *2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2020. Accepted.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman. "A GNSS-R Forward Model for Delay-Doppler Map Assimilation." *2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2018. URL: https://doi.org/10.1109/IGARSS.2018.8518987

Technical Reports

Huang, Feixiong, Giuseppe Grieco, and Ad Stoffelen. "GNSS-R Processing and NWP Assimilation." EUMETSAT OSI SAF (2019). URL: http://www.osi-saf.org/?q=content/gnss-r-processing-and-nwp-assimilation

Code Publications

Feixiong Huang, Andrew O'Brien, Nereida Rodriguez-Alvarez, James Garrison (2020) GNSS-R DDM Forward Model and Variational Analysis Method for Data Assimilation [Source Code]. https://doi.org/10.24433/CO.5369859.v2

Conference and Workshop Presentations

Huang, Feixiong, Bachir Annane, James L. Garrison, Mark Leidner, Giuseppe Grieco, Ad Stoffelen and Ross Hoffman, "The Influence of Assimilating CYGNSS DDMs on Global Analyses and Tropical Cyclone Forecasts", in *2020 AGU Fall Meeting*, Online, Dec. 2020.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, Giuseppe Grieco, Ad Stoffelen and Ross Hoffman, "Assimilation of GNSS-R Delay-Doppler

Maps into Weather Models", in *Geoscience and Remote Sensing Symposium (IGARSS) 2020*, Online, Sep. 2020.

Huang, Feixiong, James L. Garrison , Mark Leidner, Bachir Annane, Giuseppe Grieco, and Ad Stoffelen, "Assimilation of CYGNSS DDMs into Global Surface Wind Analyses", in *CYGNSS Science Team Virtual Meeting*, Online, June 2020.

Annane, Bachir, Mark Leidner, Ross Hoffman, **Feixiong Huang**, and James Garrison, "Influence of Assimilating CYGNSS Ocean Surface Wind Data on Tropical Cyclone Analyses and Predictions", in *EGU General Assembly Conference 2020*, Online, May 2020.

Grieco, Giuseppe, **Feixiong Huang**, Ad Stoffelen, James Garrison, and Marcos Portabella, "Are multi-look approaches feasible for GNSS-R winds?", in *6th Workshop on Advanced RF Sensors and Remote Sensing Instruments, ARSI'19*, Noordwijk, The Netherlands, Nov. 2019.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "Assimilation of CYGNSS Delay-Doppler Maps by a Two-Dimensional Variational Analysis Method", in *IEEE GNSS+R 2019*, Benevento, Italy, May 2019.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "Direct assimilation of Delay-Doppler Maps into Hurricane models by a Variational Analysis Method", in *ESA Living Planet Symposium 2019*, Milan, Italy, May 2019.

Huang, Feixiong, James L. Garrison, S. Mark Leidner, Ross N. Hoffman, and Han Zhang, "Demonstration of Assimilation of CYGNSS Delay-Doppler Maps (DDMs) in a 2-Dimensional Variational Ocean Vector Wind Analysis", in *99th American Meteorological Society Annual Meeting*, Phoenix, USA, Jan. 2019.

Garrison, James, Han Zhang, **Feixiong Huang**, Mark Leidner, and Bachir Annane, "Progress in Extended-DDM Processing for Wind-Field Retrievals", in *CYGNSS Science Team Meeting*, Pasadena, USA, Jan. 2019.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "A GNSS-R forward model for delay-doppler map assimilation", in *Geoscience and Remote Sensing Symposium (IGARSS) 2018*, Valencia, Spain, July 2018.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "Assessment of the DDM Forward Model in CYGNSS data assimilation", in *CYGNSS Science Team Meeting*, Ann Arbor, USA, May 2018.

Garrison, James, **Feixiong Huang**, Mark Leidner, Bachir Annane, and Ross Hoffman, "Assessment of a Delay-Doppler Map (DDM) Observation Operator for use in CYGNSS Data Assimilation", in *33rd Conference on Hurricanes and Tropical Meteorology*, Ponte Vedra, USA, April, 2018.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "Development of a Forward Model for the Assimilation of Delay-Doppler Maps (DDMs)", in *2017 AGU Fall Meeting*, New Orleans, USA, Dec. 2017.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "GNSS-R DDM Forward Model Status Update", in *CYGNSS Application Workshop II*, Monterey, USA, Oct. 2017.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, Kaitie Schoenfeld and Ross Hoffman, "Evaluation of CYGNSS winds retrieved using an Extended Kalman Filter (EKF) and direct assimilation of Delay-Doppler Maps (DDMs)", in *GNSS+R 2017*, Ann Arbor, USA, May 2017.

Huang, Feixiong, James L. Garrison, Mark Leidner, Bachir Annane, and Ross Hoffman, "EKF and Forward Model: Status Update", in *CYGNSS Science Team Meeting*, Ann Arbor, USA, May 2017.

VITA

VITA

Feixiong Huang is from Wuhan, China. He received the B.S. degree in Geodesy and Geomatics Engineering from Wuhan University, Wuhan, China, in 2014, and the M.S. degree in Civil Engineering from Purdue University, West Lafayette, IN, USA, in 2015. From 2016, has has been pursuing the Ph.D. degree with the School of Aeronautics and Astronautics from Purdue University, under the supervision of Professor James Garrison. He also has concentrations in Ecological Sciences and Engineering and Computational Science and Engineering. In the summer of 2019, he worked as a visiting scientist at the Royal Netherlands Meteorological Institute (KNMI), The Netherlands, supervised by Dr. Ad stoffelen who is the Active Instruments group leader of Satellite Division.

Since 2016, he has been a member of the CYGNSS Science Team. He is also currently a member of the IEEE Geoscience and Remote Sensing Society (GRSS), American Geophysical Union (AGU) and The Institute of Navigation (ION). His research interests include spaceborne GNSS remote sensing and atmospheric data assimilation.