UNIVERSITY OF CALIFORNIA
SANTA CRUZ

DETERMINING FIRST-ORDER BOUNDARIES IN HIGH-FREQUENCY RADAR SENSING OF OCEAN SURFACE CURRENTS: A PROPOSED METHOD USING BAYESIAN ESTIMATION

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

by

Irene Rodriguez Alegre

March 2022

The Dissertation of Irene Rodriguez Alegre
is approved:

Professor Donald M. Wiberg, Chair

Professor John F. Vesecky

Dale Trockel, Ph.D.

Peter F. Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

List of Algorithms

Abstract

Determining First-Order Boundaries
in High-Frequency Radar Sensing
of Ocean Surface Currents:
A Proposed Method using Bayesian Estimation

by

Irene Rodriguez Alegre

High Frequency (3-30 MHz) radars measure the speed and direction of ocean surface currents in near real time by emitting a vertically polarized electromagnetic signal along the electrically conductive ocean surface. This signal is coherently back-scattered by Bragg resonant ocean waves – travelling radially away or towards the radar with ocean wavelength equal to half of the radar wavelength.

More than 400 oceanographic HF radars worldwide use this strong coherent return of energy (Bragg scattering) and its Doppler shift to infer the radial velocity of the surface water and to map the radial currents hourly with 1 to 5 km range resolution.

The primary objective of this study is to recognize and isolate the first-order Bragg back-scattered echoes, which are necessary for mapping currents, while excluding second-order echoes that lead to erroneous ocean current estimates.

Setting the boundaries of the first order Bragg peaks in the echo Doppler spectra is a very important task. We have observed how present algorithms fail under certain circumstances, e.g. when ocean conditions are influenced by tides or strong off-shore currents and in regions around islands.

In order to avoid some of the known shortcomings of conventional methods we have developed a new algorithm to find initial estimates of the boundaries between first and second order echoes by using more advanced null detection methods and normalized radar spectra across multiple ranges.

In addition, our approach keeps track of the spatial and temporal history of data and applies Bayesian estimation techniques to the observed radar back-scatter spectra to detect and filter out erroneous observations. Several Bayesian sequential estimation techniques were tested to improve the estimation of the first-order Bragg echo boundaries in the aforementioned challenging scenarios. We begin with simple filters like G-H and Kalman and include more advanced ones like Nearest Neighbor and Probabilistic Data Association filters.

Applying these filters to the initial estimates provided by the proposed null-finding method provides better first-order boundary results than filtering previous conventional estimates. By accepting more valid ocean current radial vectors, yet rejecting second-order spectrum and noise or interference, we improve radial current vector maps.

This dissertation is dedicated,
with genuine gratitude and love,
to my family and friends.
Gracias por todo ♡

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. John Vesecky. We first met back in 2011 when I visited California as part of my undergraduate student exchange program and through the years he has offered continuous support and guidance. Thank you for the opportunity of being your PhD student, for motivating me during the PhD classes, exams and research study to continue moving forward. For your patience, particularly with this dissertation. His guidance helped me in all the years of research and was highly appreciated when writing this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

My gratitude extends to my other PhD mentor, Dr. Dale Trockel, for his PhD advice, insightful comments and feedback, but also for the questions that instigated me to widen my research from various perspectives, particularly the statistical and mathematical point of view.

My sincere thanks to President Donald Barrick and Chad Whelan at CODAR Ocean Sensors. Without their financial support it would have not be possible to conduct this research. Their knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. Thanks for the amazing opportunity to join their team as an intern, and for giving me access to research facilities and an immense amount of resources. My gratitude as well to Bill Rector and Belinda Lipa who were extremely helpful in sharing their tests, memos and knowledge for me to understand and learn from their intensive work. Thanks to Hardik Parikh as well, who offered his help every time I needed it.

Additionally, I am grateful to the University of California Santa Cruz and the Electrical Engineering and Computer Science Department for the International Doctorate Fellowship which was such a helpful funding opportunity for three years and to the faculty that inspired me during their classes. Moreover, I would like to thank the other two members of my thesis committee: Prof. Donald Wiberg and Prof. Sara Abrahamsson, for their encouragement and interest in my research, and for going over the Advancement to Candidacy and this dissertation with us.

There are some other experts in the field who have been very helpful throughout this research and my academic life that I would like to acknowledge as well: Dr. Hugh Roarty at Rutgers University, Dr. Jeff Paduan at the Naval Postgraduate School, Prof. Jerome Fiechter at UC Santa Cruz Institute of Marine Sciences, Prof. John Largier at UC Davis Bodega Marine Laboratory and finally Dr. Rich Pawlowicz at University of British Columbia as well as Dr. Richard Dewey, Dr. Steve Mihály, Manman Wang and their team at Ocean Networks Canada for hosting me at their school and being so helpful sharing data from their radar sites in the Straits of Georgia.

Finally, all my love and gratitude to my supportive parents Eduardo and Carmen, brother Miguel, grandparents Maria and Antonio, aunt Maria del Mar and cousins, my whole family and dear friends in Spain. To Adam, Elena, Daria, Michelle, Denise, Jose and Carolina, my extended family in California as well as in Washington. Thank you all for your patience and support while I went through some of my toughest times the past few years. You are all amazing and I wouldn't be here without you.

Chapter 1

Introduction

In this chapter we first present an introduction to High-Frequency Radar technology and its main applications. Secondly, we describe the objectives of this research while introducing the concept of first-order boundaries and explain why are they relevant to the present study. Then, we describe the order of the chapters and the overall organization of this dissertation that will be mentioned in the fourth section: the contributions to knowledge. In addition, we also discuss the state of the art.

1.1 Oceanographic High-Frequency Radar

High-Frequency radars, which operate in the frequencies between 3 and 30 MHz, measure the speed and direction of ocean surface currents in near real time by emitting a vertically polarized electromagnetic signal along the electrically conductive ocean surface. This signal is coherently backscattered off the so-called Bragg resonant ocean waves – travelling radially away or towards the radar with ocean wavelength equal to half of the radar wavelength[2][3][11][19], as we will describe thoroughly in Chapter 3.

Oceanographic HF radars worldwide use this strong coherent return of energy, the Bragg scattering, and its Doppler shift to infer the radial velocity of the surface water and to map the radial currents hourly. Although limited by surface wave propagation loss, such radars can measure ocean surface currents over an area 100 km or more in radius offshore, approximately every 10 minutes with a surface current accuracy of about 10 cm/s and a spatial resolution of 1 to 6 km at a cost of around $150,000 per radar station. One can find more than 400 operating oceanographic HF radar stations in America, Europe, Asia and Oceania as shown in Figure 1.1.

The aforementioned ocean surface current maps are used in many oceanographic applications (see Figure 1.2) and some of them are listed below:

- physical oceanography,
- coastal engineering,
- search and rescue,
- maritime navigation and awareness,
- response to oil and other hazardous materials spills,
- water quality control,
- ocean biology,
- storm and hurricane research, prediction, and monitoring.

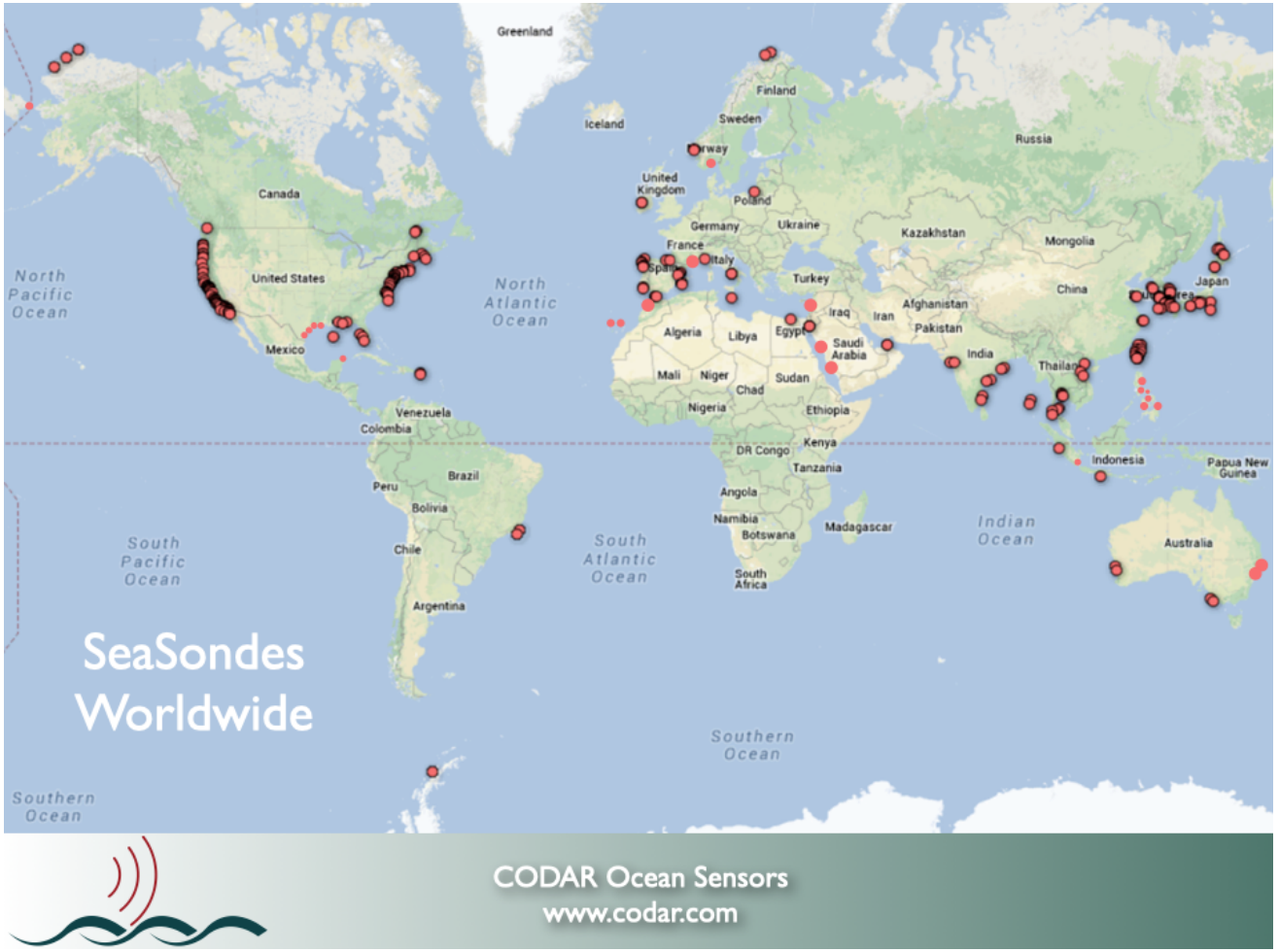

Figure 1.1: (Credit to Chad Whelan at CODAR Ocean Sensors Ltd.) Global HF Radar Network showing over 400 SeaSonde HF Oceanographic Surface Wave Radars in approximately 30 countries.

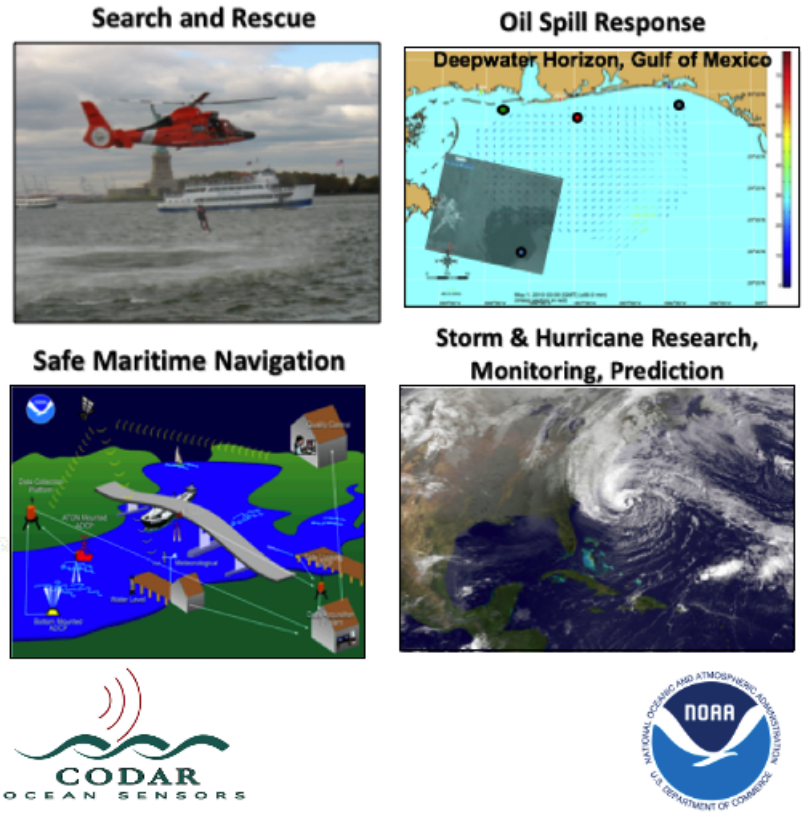

Figure 1.2: (Credit to Hardik Parikh and Chad Whelan at CODAR Ocean Sensors Ltd. in Mountain View, CA.) Figure presenting some of the most relevant applications of the SeaSonde High-Frequency Oceanographic Radar.

1.2 Motivation and Objectives of Research

This research focuses on the SeaSonde High-Frequency Oceanographic Radar data since more than 85% of all coastal oceanographic radars worldwide are SeaSonde by CODAR Ocean Sensors Ltd. As we discuss later in this dissertation, there are a number of HF radar sites in the SeaSonde network that we will particularly study because they present uncommon backscattered spectra – i.e. the Bragg regions show splits (first-order Bragg echoes have pronounced, multiple maxima) or an unusual structure.

Setting the boundaries of the first order Bragg peaks in the echo Doppler spectra is paramount for ocean current mapping and it has been observed how present algorithms fail under certain circumstances, e.g. when ocean conditions are influenced by tides or strong off-shore currents, in regions around islands, and during other air-sea interaction events. We define the radar sites that present those irregularities as "challenging". The main difficulty when working on a challenging site is the determination of the first-order boundaries of the Bragg region since they do not appear as at typical sites.

Hence, the motivation of this research is to improve the technique to properly determine the first-order boundaries, which in turn will improve the estimation and mapping of ocean surface currents. It is important to note that incorrectly estimating the boundaries of the first order Bragg regions in Doppler space will lead to two types of errors: we would include data that leads to erroneous radial currents being mapped and/or we would exclude valid data and thus lose valid ocean current measurements.

Therefore, the primary objective of this research is to recognize and isolate the first-order Bragg-backscattered echoes, which are necessary for mapping currents, while excluding second-order echoes and interference that lead to erroneous ocean current estimates.

We work on improving the accuracy of the surface current vector estimates generated by the *"conventional"* method that is currently being used. To do so we developed a new, proposed null finding algorithm that includes Bayesian Estimation techniques applied in both range and time and the use of a Probabilistic Data Association Filter to decide the optimal location for the first-order boundaries.

Through this proposed algorithm, we are able to discard erroneous ocean surface current vectors that the conventional method had placed in the map, and to add new vectors that agree with valid ocean currents present in the area nearby the corresponding radar site location.

It will be demonstrated that the new null-finding method used to determine the first-order boundaries is able to detect that a Bragg region has been split (first-order Bragg echoes present pronounced, multiple maxima), and it can track both sides of the split Bragg region, and therefore can add all the vectors that correspond to the area of the split that was missed with the conventional method.

Please refer to Section 1.3 for a description of every chapter and to find where to read more about each of the relevant concepts mentioned above.

1.3 Document Organization

This dissertation is divided into ten chapters. After this introduction in **Chapter 1**, we discuss the principles of oceanographic high-frequency radar and its applications to surface current mapping in **Chapter 2**.

Next, **Chapter 3** describes the importance of determining the boundaries between first-order and second-order Bragg regions backscatter properly. We also describe the present method being used to find the location of those first-order boundaries.

The proposed method that we have developed to determine the first-order Bragg boundaries consists of three major stages:

1. Proposed alternative multi-probabilistic null-finding algorithm developed, described in **Chapter 4**, which is based on the *conventional* method but provides some calculations and advantages that differ from it, plus it is able to output multiple possible locations for the first-order boundaries.

2. Object tracking and Bayesian estimation techniques, introduced in **Chapter 5**, that are very relevant for the Bayesian approach in our proposed algorithm to correct and smooth the first-order boundary locations.

3. The complete algorithm studied in this dissertation is summarized and listed in **Chapter 6**. When multiple possible first-order boundary locations are provided we apply the additional step of the Probabilistic Data Association filter, described in **Chapter 7**, for the spatial tracking in range as well as the Bayesian tracking in time.

Once the proposed method has been described, in **Chapter 8** we proceed with a series of statistical comparisons to demonstrate the validity of this alternative method versus the conventional method. We present comparisons of determined first-order boundaries on the radar backscatter spectra as well as comparisons of the corresponding results mapped into surface current radial maps.

Next, in **Chapter 9** we present a scenario from HF radar data collected in the Straits of Georgia where the statistical differences method did not seem relevant. In this study we compare the radial current vectors from both line boundary determination methods against a series of radial velocities obtained from drifters that were deployed by the University of British Columbia in the Straits of Georgia.

Finally, in **Chapter 10** we present a summary of the study and the results. In addition, in the conclusions we review the contributions to knowledge and sum up the success of using the proposed Bayesian method especially in challenging scenarios. Finally, we mention several topics that we think are worth pursuing to improve and validate the results of work as well as moving forward with improvements to the method.

Contributions to Knowledge and State of the Art

Contributions to knowledge:

1. Analysis algorithm to identify challenging observational situations where the first-order Bragg backscattered echoes (Chapter 2) in a given range bin either present unusual characteristics on the approaching (positive Doppler shift) and/or receding (negative Doppler shift) ocean waves or they form multiple peaks – see Figure 3.12 for an initial idea of those split Bragg peaks that will be discussed in Sections 3.3.2 and 4.9.

2. Surface current measurements using compact high-frequency radars are presently handicapped in instances where the present (conventional) method does not accurately determine the first-order boundaries. Instead, sometimes it includes second-order or interference as valid radar data that results in erroneous radial vectors, and misses some radar data that could produce useful radial current vectors.

3. By the application of a new null finding algorithm and Bayesian estimation techniques, plus probabilistic data association filtering (PDAF) to resolve multiple options we show that we can better define the first-order Bragg boundaries (see Chapter 3) and thus include more useful surface current vectors while at the same time excluding data that results in erroneous surface current vectors.

4. The method for finding the first-order boundaries developed here provides the most advantage in challenging situations where conventional algorithms are most susceptible to errors, namely:

    - Split Bragg regions when islands are present in the radar field of view,
    - Interference is present from ionospheric echoes, HF broadcast services or other sources,
    - Strong second order echoes are present i.e. in high sea states.

5. We validate our techniques by comparison with conventional techniques using both surrounding current measurements and drifter measurements of surface currents.

6. Our improvements mean that in the future the measurements performed by HF radars will be possible in some of the adverse conditions where useful measurements were previously impractical.

7. Further, these improvements mean that the oceanographic and practical applications of these radars are less susceptible to the present problems caused by erroneous radial vectors distorting results.

State of the Art:

The objective of this thesis is to develop and improve methods for selecting that portion of the radar echo data that allows valid radial current vectors to be retrieved from the radar observations of compact antenna, decameter wavelength, ground wave radars.

Although there are multiple publications with ideas for improving the retrievals of surface currents[10][11][20][23][28][33][36][39][42][50][69], a critical point in the radar signal processing is setting the first-order boundaries in the Doppler spectrum of the radar echoes, as discussed above and in Chapter 3.

Kirincich[37] developed an alternative methodology based on image processing techniques that uses a single and globally relevant smoothing length scale, careful pre-treatment of the spectra, and marker-controlled watershed segmentation, an image processing technique, to separate areas of spectral energy due to surface currents from areas of spectral energy due to more complex scattering by the wave field or background noise present. This novel idea shows some success in selecting the first-order Bragg line echoes better than the existing data processing software of the time of his work (around 2016). However, we are not aware of any adoption of the technique or further development of it.

The research to develop the algorithms presently in use rests on creative ideas plus trial and error with observational data. Most of the work on this critical task has been done in the development of the data processing software for CODAR SeaSonde radars and is not available beyond a few brief memos. Based on these memos and current practice we have constructed a *conventional* method to represent the state of the art. This *conventional* method is introduced in Chapter 4 and its algorithm is described in Section 6.3. However, we note that the data processing software for CODAR SeaSonde is continually evolving in efforts to improve the retrieval of ocean surface currents. Nevertheless, our *conventional* model is up to date and forms a good standard of comparison from which to validate the results of our thesis research.

In fact, in Chapter 8 we process radar data from the Straits of Georgia by both *conventional* and our new *multi-probabilistic* methods and statistically compare the results. The situations that the *conventional* method finds most challenging are when there are multiple peaks in the first-order Bragg region echoes due to islands in the radar coverage, flow in straits, tidal flows, and coastal upwelling. Another situation that presents problems for the conventional method is high sea states.

In this thesis we develop methods to better handle these challenging situations and test the validity of our new methods with both statistical tests and comparisons with drifting buoys. We choose that particular VCOL radar site in the Straits of Georgia for our comparisons because it is indeed a site where the Bragg regions present splits and therefore the first-order boundaries determination from our proposed *multi-probabilistic* method (Chapter 4) and the advantages of the Bayesian tracking we apply (Chapter 7) are more significant. In addition we also run the comparisons against drifting buoys on that same area of the Straits of Georgia for further verification of the favorable outcome of the proposed method we have developed.

Chapter 2

High-Frequency Radar and its Oceanographic Application

High Frequency (3-30 MHz) radars measure the speed and radial direction of ocean surface currents in near real time by emitting a vertically polarized electromagnetic signal along the electrically conductive ocean surface. This signal is coherently backscattered by Bragg resonant ocean waves – travelling radially away or towards the radar with ocean wavelength equal to half of the radar wavelength.

More than 400 oceanographic HF radars worldwide use this strong coherent return of energy (known as Bragg scattering) and its Doppler shift to infer the radial velocity of the surface water and to map the radial currents hourly with 1 to 5 km range resolution. Installing two radars allows the measurement of two-dimensional directions along the ocean surface in their field of view.

In this Chapter we present an introduction to some important concepts we will study in this dissertation, and the organization of the sections is presented next.

First, in Section 2.1, we provide an introduction to the radar, its history and principles of operation.

Later, in Section 2.2, we define a few of the most relevant concepts associated with a radar system. Amongst others, we will describe the radar wavelength, the concept of range to target, the properties of the antennas used on a radar system as well as the principal equations that describe the signals transmitted by a radar.

Finally, in Section 2.3 we describe the application of most interest for this dissertation: the oceanographic use of high-frequency radars to map ocean surface radial currents. After describing some history of the oceanographic radar, we introduce the concepts of Bragg scattering, Doppler-shift, and first-order Bragg peaks. In addition, we include a comparison of the two possible architectures for an oceanographic radar: phased-array beam-forming as opposed to direction-finding radars, in Section 2.3.1. At the end, Section 2.3.2 describes the MUSIC algorithm, a key technique in the CODAR SeaSonde compact direction-finding that chooses a radial surface current magnitude and determines the bearings of the patches of ocean surface where that current magnitude is present.

The system of measurement employed to express the values of quantities in this document is the International System of Units (SI). Only SI units and those units recognized for use with the SI system are utilized.

2.1 Introduction to Radar

A radar system uses radio waves to detect the presence of objects and to determine their range, angle, and velocity. The term *radar*, derived from radio detection and ranging, was first used by the U.S. Navy around 1940. The system was developed secretly, for military use, before and during World War II.

There are many **applications** for radar, on scale sizes that vary from a few centimeters – like measuring the thickness of a wall – to long range systems probing planets across the Solar System.

Some current radar applications include:

- Ground-based radar applications, i.e. air traffic control, sea traffic control, mapping ocean currents, speed limit radars on the roads and industrial radar measurements;
- Sea-borne applications for navigation or collision avoidance;
- Air-borne applications like altimeters, systems for navigation or to study weather formations;
- Space-borne systems to study the ocean and other Earth resources, mapping planets and other bodies in space;
- Military systems including the detection and tracking of targets or the guidance of weapons.

The **principle** of a radar system is that a transmitter sends electromagnetic waves, directed by an antenna, that propagate in the range of frequencies between VLF (starting at 3 kHz) to microwaves and higher (tens of GHz and above).
This signal will then reflect (also known as *scatter*) off anything it encounters – land, sea, ships, aircraft – and a small amount of its energy will scatter directly back to a receiver antenna which is usually, but not always, located near the transmitter. The term *backscatter* is given to the signal reflections that travel in the exact opposite direction to the incident rays. After amplification in the receiver, these signals are processed to sort out the required echoes from the 'clutter' of unwanted echoes by a combination of both analog electronic signal processing and digital computer data processing software.

Radar transmissions beyond the horizon:

The curvature of the Earth limits the line-of-sight distance from the top of a 100-meter tower to about 30 km. The experiments of Guglielmo Marconi in 1901 proved that radio waves could propagate beyond the horizon, since signals were successfully transmitted from Cornwall, United Kingdom, to Newfoundland in Canada. The solution to the Maxwell's equations for a wave at a plane interface between two media gives a space wave (free space propagation) and a surface wave (guided propagation along the interface). The discovery of the Earth's ionosphere in the 1920s, proved that there can be a refracted mode of propagation, the ionospheric *skywave*: which is the explanation of Marconi's transatlantic communications.

Figure 2.1 shows the principle of operation of a *skywave* over-the-horizon radar, where the emitted electromagnetic waves refract off the ionosphere.

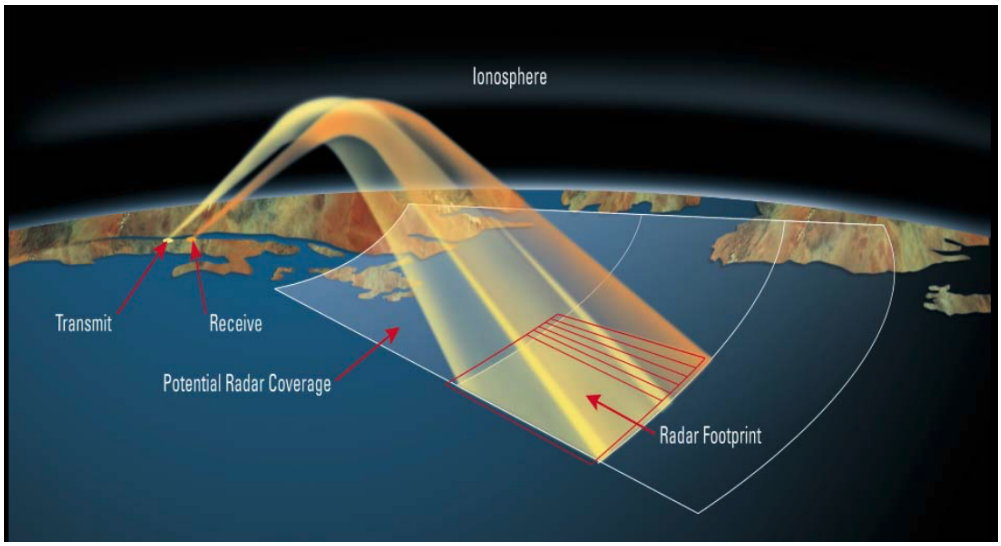

Figure 2.1: Graphical description of the principle of operation of a skywave over-the-horizon radar that employs two antenna systems installed in the coast to detect targets located hundreds of km away in the ocean. The transmitting antenna emits the electromagnetic waves that refract their signals off the ionosphere for long-range detection. Then the signal is reflected and travels back to the receiving antenna that will then process the backscatter spectra.

The SeaSonde *groundwave* radars that we use during this research operate in the decameter (3-30 MHz) band – also known as HF (high-frequency) band – and fall in the category of over-the-horizon radars. However, their transmitted electromagnetic wave closely follow Earth's surface, particularly over water, as a result of the wave's interaction with the terrestrial surface as shown in Figure 2.2. In particular the SeaSonde radars that we employ and require transmitting and receiving antennas as displayed in Figure 2.1.

More information about the modes of propagation of a radar signal will be described in Section 2.2.8.

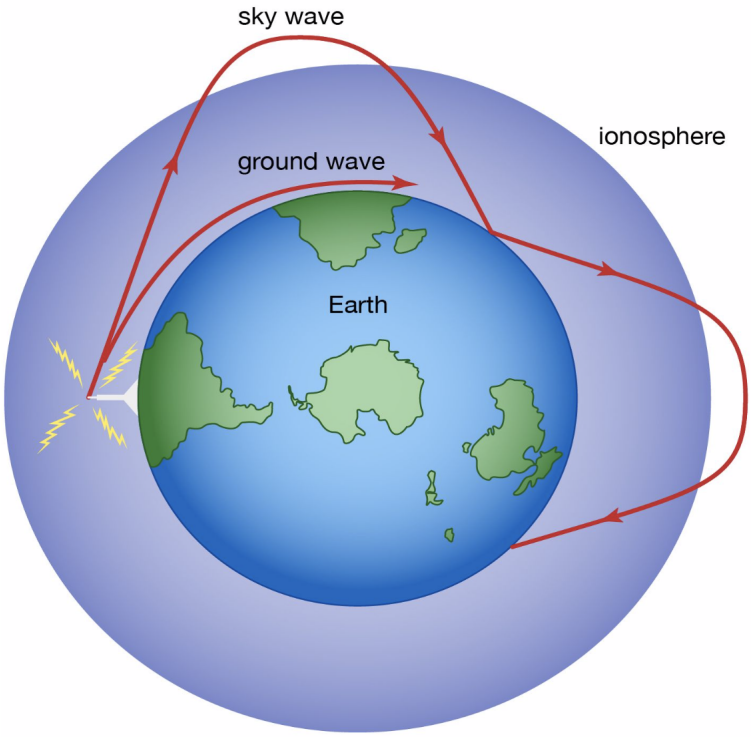

Figure 2.2: Radio wave transmission reaching beyond line of sight for over the horizon radars[15]. While the transmitted electromagnetic waves of a *skywave* radar refract off the ionosphere; the signals transmitted on a *groundwave* radar follow the Earth's surface as a result of the wave's interaction with the terrestrial surface. The exact range that the reflection of the sky wave by the ionosphere and the ground wave could reach depends on the system operating frequency. More details about a radar signal modes of propagation presented in Section 2.2.8.

2.2 Radar Fundamental Concepts

2.2.1 Radar Wavelength

The frequency f and wavelength λ of an electromagnetic wave are related to the velocity of light c by

$$c = \lambda f \qquad (\mathrm{m\,s^{-1}}) \tag{2.1}$$

At HF frequencies (3 to 30 MHz), electromagnetic waves can propagate very long distances over the horizon from a land based radar. The problem is that since wavelengths become very large (in the decameter size), antennas often have to be large as well in order to obtain high accuracy in the determination of the direction to the target or place antenna nulls toward unwanted echoes. For example, at 3 MHz, a quarter-wavelength monopole would be 25 meters high which makes it difficult to build a high-gain antenna that could potentially have a vertical null to avoid ionospheric echoes. Another issue found at HF frequencies is that layers and irregularities in the ionosphere can cause strong scattering, which creates unwanted echoes that could be confused with moving ocean waves.

2.2.2 Range to the Target

The radar transmitter antenna transmits a short pulse, focused in a direction given by the directivity (Section 2.2.6) of the antenna and this pulse propagates in this given direction at the propagation speed of EM waves for the situation of interest – typically very close to the speed of light. Once the pulse is transmitted, the radar clock starts counting and when it encounters obstacles or targets, part of the energy of this pulse is reflected back to the radar. The receiving antenna receives this energy, the clock registers the time and then the signal information is evaluated.

Since the radio pulse travels away from the radar at its known, propagation velocity, the distance (*range*) between the radar and the target that has reflected the signal can be calculated from the time that took to travel the distance to the target and back[35]. The range R to the target is

$$R = \frac{c \cdot \tau}{2} \qquad \text{(m)} \tag{2.2}$$

where c is the speed of light and τ is the time delay of the echo from the moment it is transmitted to the moment it is received i.e. having travelled the range distance twice.

2.2.3 Signal Bandwidth

A radar receiver must have the ability to process the signal *bandwidth* of the backscattered pulse that it receives. The bandwidth BW is the difference between the upper and the lower cut-off frequencies of the radar receiver and it is measured in Hertz.

For a single short pulse radar, the minimum bandwidth BW to receive a pulse of duration τ_{pulse} seconds is

$$BW = \frac{1}{\tau_{pulse}} \qquad \text{(Hz)} \tag{2.3}$$

2.2.4 Bearing and Range Resolution

The target resolution of a radar is its ability to distinguish between targets that are very close in either range or bearing.

Range resolution is the ability of radar system to distinguish between two or more targets at different ranges. In a basic short pulse radar (see Figure 2.3) the degree of range resolution depends primarily on the width of the transmitted pulse, τ_{pulse} seconds. The size and types of targets can also affect the resolution, as well as the efficiency of the receiver.

A well-designed and efficient radar system should be able to distinguish targets separated by one-half the pulse width time τ_{pulse}. Thus, the theoretical range resolution S_R of a radar system can be calculated as shown in Equation (2.4), which is based on Equation (2.2) for the range of a radar, expressed in terms of the pulse width time τ_{pulse}. As we see, this resolution can also be expressed in terms of bandwidth (Equation 2.3).

$$S_R \geq \frac{c \cdot \tau_{pulse}}{2} \quad \Rightarrow \quad S_R \geq \frac{c}{2 \cdot BW} \qquad \text{(m)} \tag{2.4}$$

Bearing resolution (also known as angular resolution) S_A is the minimum angular separation at which a radar is able to distinguish two targets located at the same range but at a different bearing. The angular resolution is typically determined by the antenna characteristics, specifically by the antenna beamwidth represented by the angle Θ, which is defined by the half-power point i.e. point where the power is –3 dB from the maximum. The angular resolution as a distance between two targets can be calculated as

$$S_A \geq 2R \cdot sin(\frac{\Theta}{2}) \qquad \text{(m)} \tag{2.5}$$

where Θ is the antenna beamwidth angle and R is the range in meters at which both targets are located.

2.2.5 Signal Modulation and FMCW

Initially, the principal radio frequency of a radar system is set by the frequency synthesizer. The first types of radar we describe are the **continuous-wave** radars (CW), where the transmitted signal is sent out continuously in time.

On the other hand, in a **pulsed wave radar**, this continuous signal is pulsed on and off by the modulator, creating shorts bursts of radio energy that are then amplified by the transmitter and sent to the antenna via a transmit – receive switch or duplexer. The two main functions of the switch are to connect the antenna to either the transmitter or the receiver at the appropriate times, and to protect the receiver part of the system from the higher energy of the transmitted pulse. The waves are indeed interrupted for a period of time so that the wave can reach a reflecting target or surface and a portion of the energy can return to the same antenna before the next burst of waves are transmitted.

Pulse compression is a method for improving the range resolution (and often signal to noise ratio) of a pulse radar by spreading out the pulse transmission in time and modulating the frequency of the transmission, typically a ramp in frequency over time, as shown in Figure 2.3. The range resolution of a pulse compression radar system is given by the length of the pulse after the pulse compressing stage[70], namely τ_c. Pulse compression combines the advantages of transmitting very short pulses with the energy advantages of very long pulses. The factor determining the ability to compress the pulse is not the pulse width (as in simple pulse-modulated radars), but the pulse bandwidth instead: the receiver needs at least the same bandwidth to process the full spectrum of the echo signals.

In this technique the transmission pulse is modulated internally, and it allows one to send long pulses and therefore high power but still obtain high range resolution. The Pulse Compression Ratio (PCR) is the ratio of the time length of the uncompressed transmitted pulse to the length of the compressed pulse, and it is approximately equal to the pulse compression gain. Finally, the range resolution S_R –compare with Eq.(2.4)– is therefore the range resolution of the FMCW compressed pulse length multiplied by the PCR.

$$\mathrm{PCR} = \frac{c \cdot \tau_{pulse}/2}{c\ /(2 \cdot BW)} = BW \cdot \tau_{pulse} \tag{2.6}$$

$$S_R = c \cdot \frac{\tau_{pulse}}{2} = \mathrm{PCR}\frac{c}{2 \cdot BW}$$

In our study, we are interested in a hybrid version of these two types of radar (continuous wave and pulsed wave radar), which is the **frequency-modulated continuous-wave** radar, described next and compared against pulsed wave radars in Figure 2.3.

A Frequency-Modulated Continuous Wave radar (FMCW) is a special kind of radar sensor which transmits continuously, like a simple continuous wave radar (CW), but that can modify its operating frequency during the measurement, i.e. the transmission signal is modulated either in frequency or in phase. On the other hand, pulse radars can only deliver their peak power in short bursts. With FMCW, a signal is transmitted and then the frequency (or phase) of this signal is increased or decreased periodically. When the echo signal is received, that change of frequency generates a delay Δt (known as 'runtime' shift), similar to the technique used on pulse radars. In a pulse radar this runtime must be

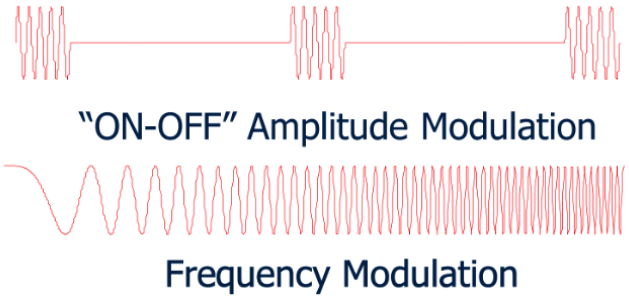

Figure 2.3: (Credit: Michael Davis, Georgia Tech Research Institute) Comparison of transmitting a pulsed wave (top) versus a frequency-modulated continuous-wave (lower). Pulsed Wave Radar uses electromagnetic waves that are emitted from the antenna in short bursts. On the other hand, Frequency-Modulated Continuous-Wave (FMCW) achieves a much better SNR by emitting a continuous-wave signal with a much larger energy as compared to the emitted signal of a pulse radar system. The frequency of this continuous-wave signal is linearly modulated over time and its sweep duration can be chosen independently from the bandwidth.

measured directly, but in FMCW radars we measure the differences in phase or frequency between the actual signal that was transmitted and the signal that is being received. Some of the advantages of choosing FMCW modulation are:

- Ability to measure very small ranges to the target, since the minimal measured range – Equation (2.4) – is comparable to the transmitted wavelength λ.
- High target range accuracy and ability to measure it simultaneously along with its relative velocity.
- The pulse radiation does not have to be at a high peak power. Rather, the pulse is spread out over a coherent integration time T, allowing more pulse energy for a given peak transmit power.

The range R to an object is defined in Eq. (2.7). The frequency difference in Hz is $\Delta f = \Delta t \cdot \frac{df}{dt}$, where Δt is the time delay in seconds, $(\frac{df}{dt})$ is the frequency shift per unit of time, also known as the *chirp rate*.

$$R = \frac{c|\Delta t|}{2} = \frac{c|\Delta f|}{2(\frac{df}{dt})} \qquad \text{(m)} \tag{2.7}$$

If the change in frequency is linear over a long time interval, then the radar range can be determined by a simple frequency comparison: the frequency difference Δf is proportional to the distance R to the target.

On the other hand, if the reflecting object has a *radial velocity* with respect to the receiving antenna, then the echo signal contains a Doppler frequency f_D caused by that velocity. Thus, the radar will measure the frequency difference $f_{\Delta t} = \frac{1}{\Delta t}$ caused by the runtime and, in addition, the Doppler frequency f_D caused by the radial velocity. These frequencies will be either added or subtracted depending if the movement direction and the modulation direction are equal or opposite to each other, resulting in $|\Delta f| = |f_{\Delta t} \pm f_D|$.

The coherent integration time T determines the frequency resolution

$$f_{res} = \frac{1}{T} \qquad \text{(Hz)} \tag{2.8}$$

and, therefore, the radial velocity resolution is calculated as

$$v_{res} = \lambda \frac{f_{res}}{2} \qquad (\mathrm{m\,s^{-1}}). \tag{2.9}$$

The signal-to-noise ratio (SNR) of an HF radar system can be improved by using waveforms that maximize the percentage of time that the transmitter is on, more about SNR can be found on Section 2.2.7.

As opposed to CW modulation, FMCW involves sweeping the frequency of the continuous wave to gain bandwidth and therefore range resolution too, as seen in Eq.(2.4) and described by Barrick[7].

2.2.6 Antenna Fundamentals

An antenna is a device for radiating and receiving electromagnetic energy. The three principal characteristics of an antenna are:

1. Directivity: Ability to focus the power beam in a given direction to increase the radar sensitivity in that direction.

2. Steering: Capacity to provide some beam steering such that an area of coverage is obtained.

3. Angle: To allow the measurement of angular information so that the direction of the target can be determined.

The terms above are generally applied to beam forming antennas. Direction finding antennas will be discussed in more detail in Section 2.3.1.

An isotropic antenna is a hypothetical radiator that radiates energy in all directions uniformly, being therefore omnidirectional.

Generally all practical antennas (which all show some directivity) are compared with the characteristics of the ideal isotropic radiator which would radiate power uniformly over the whole 4π steradians of a sphere.

This property of beaming power is known as the directive gain $D(\theta, \varphi)$, or the power gain $G(\theta, \varphi)$, and abbreviated as directivity D and gain G.

While the directive gain describes the radiation pattern of the antenna, it does not provide any details about how lossy the antenna is. The power gain includes the concept of losses, which occur through heating of the antenna itself, the ground plane and any matching devices, as well as power radiated into the sidelobes. Power gain might be defined as the ratio of the radiation intensity in the main lobe of the antenna to the radiation intensity from a 100 percent efficient isotropic antenna having the same power input.

The gain of a lossless antenna is given by

$$G(\theta, \varphi) \cong \frac{4\pi}{\Delta\theta\Delta\varphi} \tag{2.10}$$

where $\Delta\theta$ [radians] is the width of the beam in azimuth direction and $\Delta\varphi$[radians] is the width of the beam in the elevation direction.

The azimuth angle θ gives directional or bearing information, measured clockwise from the north. The elevation angle φ is measured from the horizon upwards.

The power gain is most commonly used to denote the maximum gain of an antenna, which can also be calculated from its size:

$$G = (4\pi A_e)/\lambda^2 \tag{2.11}$$

where A_e is the effective area of the antenna, usually less than the real area A by an efficiency factor ε.

If the practical antenna has ohmic losses (I^2R) represented by P_{loss}, the power radiated P_{rad} is less than input power P_{in}, and they are related by η_r, the radiation efficiency of the antenna.

$$\eta_r = \frac{P_{rad}}{P_{in}} = \frac{P_{rad}}{(P_{rad} + P_{loss})} \tag{2.12}$$

The ratio of the power radiated in a particular direction (θ,φ) to the actual power input of the antenna is called power gain of an antenna, denoted by G_p (θ,φ) and given by

$$G_P(\theta, \varphi) = \frac{(P_d(\theta, \varphi))}{P_{in}} \tag{2.13}$$

2.2.7 Range Equation and Signal-to-Noise Ratio

The ability of a radar to detect the presence of a target is expressed in terms of the radar equation.

First, at the transmitter we have a peak power output P_t in Watts. If this power was radiated isotropically by the antenna, then the power flux Φ_R at a range R would be given by:

$$\Phi_R = \frac{P_t}{4\pi R^2} \qquad (\mathrm{W\,m^{-2}}) \tag{2.14}$$

If the transmitting antenna is not isotropic, concentrating the power towards the target with a gain factor G_t, then the power flux at a range R will be:

$$\Phi_R = \frac{P_t G_t}{4\pi R^2} \qquad (\mathrm{W\,m^{-2}}) \tag{2.15}$$

The target intercepts a portion of this incident power and re-radiates it. The measure of the incident power intercepted by the target and radiated back towards the radar is σ, the radar cross section. On the return path the power gain will be spread out over the sphere of area $4\pi R^2$, and in addition the amount of returning power that is intercepted by the antenna is determined by its effective area A_e.

Finally the mean power reflected and received by the radar becomes

$$P_r = \frac{P_t G_t \sigma A_e}{(4\pi R^2)^2} \qquad \text{(W)} \tag{2.16}$$

and the receiving antenna gain is

$$G_r = \frac{4\pi A_e}{\lambda^2} \tag{2.17}$$

Adding the inevitable inefficiencies in the radar system, all grouped in the loss factor L_s, we obtain the power received by the radar from the target defined by:

$$P_r = \frac{P_t G_t \sigma \lambda^2 L_s}{(4\pi)^3 R^4} \qquad \text{(W)} \tag{2.18}$$

Finally, adding the concept of noise level N, we can now compare the power received from the target with the noise power, which is known as the signal-to-noise ratio:

$$\text{SNR} = \frac{P_t G_t \sigma \lambda^2 L_s}{(4\pi)^3 R^4 N} \tag{2.19}$$

The noise level N is an average of the background noise. In some cases we need consider additional sources of interference, e.g. other discrete transmitters, machinery or power lines, to obtain a signal to interference ratio SIR.

2.2.8 Signal Propagation

An accurate study of the performance of a radar requires studying the significant effects that the Earth and its atmosphere cause in the propagation of radio waves. Some of these effects are described next, and Figure 2.4 illustrates the layers of the Earth's atmosphere mentioned below.

- Ground reflections from the surface of the Earth.
- Refraction (bending) of the radar electromagnetic waves. The refraction effect will happen in the first layer of the atmosphere i.e. the troposphere, which extends up to an altitude of around 20 km. The refraction indices are a function of pressure, temperature, water vapor, and gaseous content. At the HF frequency band (3-30 MHz) refraction also happens in the ionosphere, discussed below.
- Attenuation or absorption of part of the radar signal energy by the gases in the atmosphere. Atmospheric attenuation increases significantly in the presence of rain, fog, dust, and clouds, as well as with the increase of the frequency of the transmitted electromagnetic wave. On the other hand, it decreases with air pressure and therefore with altitude, which is why the atmospheric attenuation is maximum in the troposphere. Typically, at HF frequencies attenuation is not significant in the troposphere.
- The ionosphere extends from about 50 km to about 600 km. Compared to the troposphere, the ionosphere has very low gas density and a significant amount of ionized free electrons, which will affect the electromagnetic waves being propagated through it. The primary cause for this ionization is the sun ultraviolet and X-rays. These effects are highly dependent on the frequency at which the waves are incident, and they include refraction,

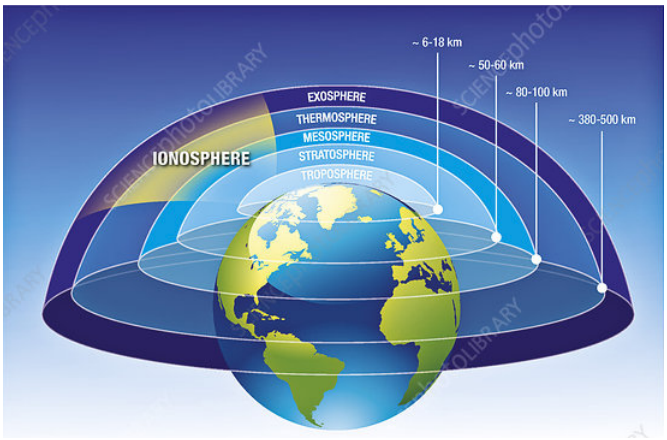

Figure 2.4: (Credit: NASA's Science Photo Library) Illustration of the Earth's ionosphere and atmospheric layers. The ionosphere (partially shown in yellow at upper left) is the ionised part of the Earth's atmosphere, reaching from around 60 to 1000 kilometres above the Earth's surface. The ionisation is caused by the Sun, and studying the ionosphere is a key part of understanding the effect space weather has on satellite communications and related technology. Other atmosphere layers shown in the figure are the troposphere and the stratosphere, the two lower layers where most surface weather occurs. The altitude boundaries of each layer is shown in kilometres (km).

absorption, noise emission, and polarization rotation. For incidence near vertical, frequencies lower than about 4 to 6 MHz are completely reflected in the lower region of the ionosphere and will not travel through it. Frequencies higher than 30 MHz may penetrate the ionosphere and will suffer some level of attenuation. For oblique incidence frequencies in the HF band (3-30 MHz) are often refracted back toward the Earth and enable "skywave" propagation between points on the Earth's surface at long distances.

- Diffraction of electromagnetic waves. The region below the horizon, close to the Earth's surface, is called the diffraction region. Diffraction is a term used to describe the bending of radar waves around physical objects and it generally increases with frequency. *Groundwave* radars and their operation were presented in Figure 2.2.

2.3 HF Radar for Oceanography

In his publication from 1955, Crombie[19] noted that part of the received signal from propagation over the ocean seemed to be related to surface waves. Further development of radar systems, driven primarily by military needs, quickly moved to shorter centimetre-scale wavelengths. However, numerous researchers proposed and demonstrated that high-frequency (HF) radar systems operating between 3 and 30 MHz frequency (10-100 meters wavelength) could be used to measure ocean surface currents and sea state. Among those researchers, publishing their research in the 1970s are Barrick[8], Hasselmann[27] and Stewart and Joy[65].

HF radar systems estimate surface currents by analyzing the spectra of the resonant scattering peak of the echo –returned signal–, which is also known as the first-order Bragg peak. (Refer to Chapter 3 for a complete study on first-order Bragg).

First-order resonant scattering occurs when outgoing electromagnetic waves reflect from radially-propagating surface waves with half the wavelength of the radar. Since ocean waves travel with a known speed set by the deep water gravity wave dispersion relation, the reflected signal is Doppler-shifted by a known amount. In addition, if the waves are being transported by an underlying, near surface, ocean current, then this Doppler shift is further modified. Thus, a measured Doppler shift is proportional to the sum of the known velocity of ocean waves and the unknown velocity of the surface current.

Radio waves in this frequency band propagate primarily as ground waves because they are effectively trapped on the electrically conductive ocean surface.

Therefore, in principle, HF radar is capable of long range measurements beyond the horizon – like presented in Figure 2.2 – because the waves do follow the curvature of the Earth. However, in practice, electromagnetic wave energy is lost by scattering and absorption, and also diluted geometrically, all of which reduce the signal to noise ratio (Eq.(2.19)) and therefore the range and quality of the data.

Summarizing, a surface-current radar system requires:

1. A conductive body of water.
2. The presence of ocean waves with a wavelength of half the radar wavelength, propagating in the radial direction, of amplitude large enough to generate a measurable reflection but not so large that most energy is scattered away.
3. A method of localization so that a measured Doppler shift can be associated with its specific patch of ocean at a particular range and azimuthal direction.

In practice, most seawater is salty enough to support propagation, and although surface wave spectra can be highly directional, there is generally enough wave energy in the radial direction to generate a suitable reflection. Range is easily found from time-of-flight estimates, as described in Section 2.2.2. However, as we will see next, obtaining directional information can be more challenging[25].

An HF surface wave radar does indeed receive echoes from multiple ranges, as shown in Figure 2.5. Using a time series of echoes, each range bin shows a Doppler spectrum for all the targets (regions) within an arc of constant range, as described next in Figure 2.6.

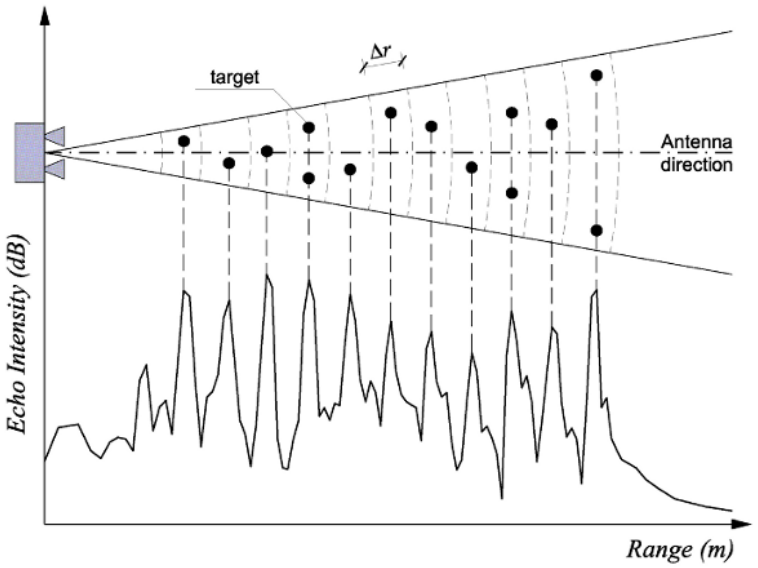

Figure 2.5: Schematic view of a high frequency surface wave radar with echoes being received from multiple range cell arcs. A time series of the echoes from each range cell arc are used to compute a Doppler spectrum for each range cell arc.

Figure 2.6 presents a system schematic showing the signal information flow, starting at the radar sensing and ending as a surface current map that includes the radial surface current vectors present in the radars field of view.

Following the arrows in Figure 2.6, the radar first collects radar echoes for several minutes in range-Doppler mode to form a Doppler spectrum (as seen in the upper right portion of the figure). Each Doppler bin corresponds to some location along the range arc and corresponds to a radial surface current speed. Each Doppler bin in placed in azimuth along the range arc. The Doppler shift is converted into a radial current speed and mapped to location in range and azimuth (as we see in the bottom left portion)[47]. Two radar stations allow the generation of a vector current map as the one presented in Figure 2.9.

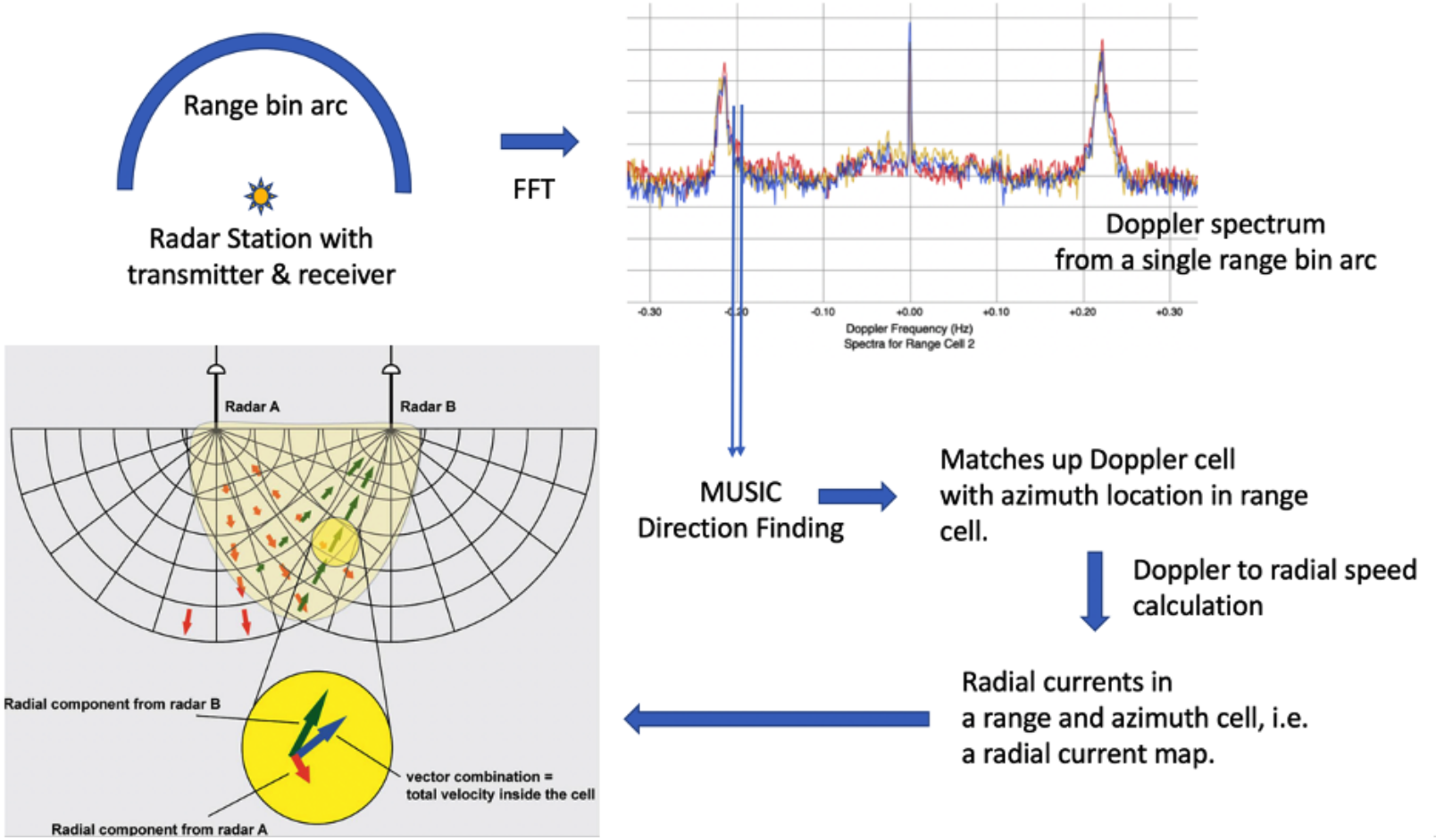

Figure 2.6: Compact radar system schematic showing signal information flow from radar sensor to surface current map. The radar (upper left) collects radar echoes for several minutes in range-Doppler mode to form a Doppler spectrum (upper right). Each Doppler bin is from some location along the range arc and corresponds to a radial surface current speed. Each Doppler bin in placed in azimuth along the range arc. The Doppler shift is converted into a radial current speed and mapped to location in range and azimuth (bottom left). Two radar stations, like shown in the bottom left portion (Credit to Mantovani et al.[47]), allow the generation of a vector current map as the one presented in Figure 2.9.

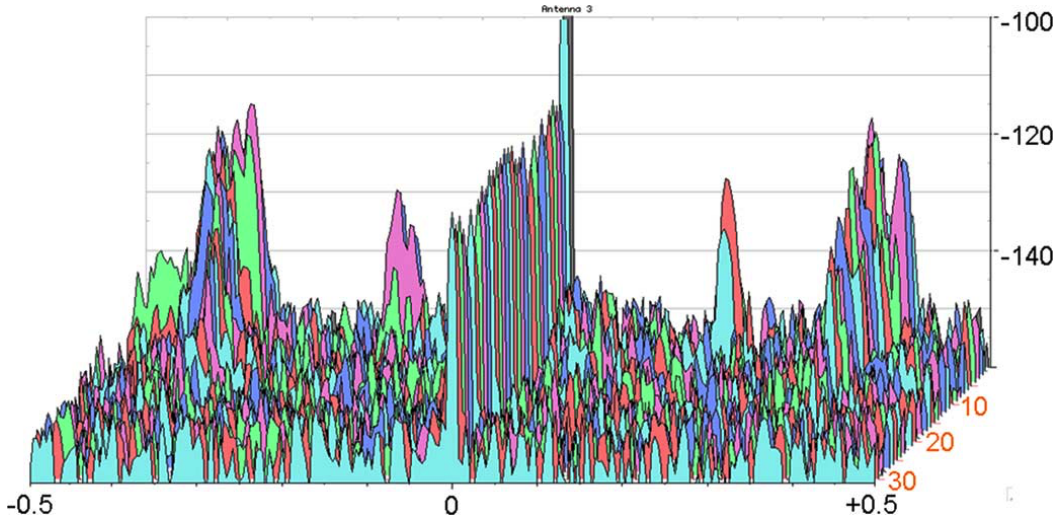

Figure 2.7: (Credit to Hugh Roarty – Rutgers University) Picture of power spectra for Antenna 3 of the SeaSonde at 00:15 GMT on February 26, 2009 [59]. The x-axis corresponds to the Doppler shift (Hz), while the y-axis is the signal strength (dB). Finally, the z-axis denotes the range bin from the radar (scalar). We observe how the signal strength decreases with increasing range, as expected.

All those range Doppler spectra are superimposed and then presented in Figure 2.7, as an example graph of the power spectra for the monopole antenna in a SeaSonde radar i.e. Antenna 3, which is the antenna we use for first-order boundaries determination[9][10][40][41]. In addition, there are two other receiving antennas in a SeaSonde radar system, namely Antenna 1 and Antenna 2, which are a dipole as seen in Figure 2.10 and Figures 3.4 through 3.7.

Each **Doppler bin** in the Doppler frequency shift axis represents the amount of energy that the signal has at that particular Doppler frequency and range.
The x-axis in Figure 2.7 corresponds to the Doppler shift in Hz, while the y-axis is the signal strength in dB. Finally, the z-axis denotes the range bin from the radar.

In this case we can see the energy spectrum from each different range bins superimposed and, as expected, we observe how the signal strength decreases as we move further away from the radar site. Also the higher energy regions (first-order regions that will be detailed in Section 3.1.2) are found constantly around ± 0.35 Hz Doppler frequency shift.

Moreover, Figure 2.8 presents the full backscatter spectra energy received at the monopole antenna (Antenna 3) for a single range at another SeaSonde site. Its horizontal axis represents the Doppler frequency, and the vertical axis is the range in kilometers from the radar site location, again we see the energy decreasing with increasing range. The two stronger first-order Bragg regions are clearly located in Figure 2.8 around the ± 0.22 Hz from the center.

From a complete ocean backscattered spectra plot – like the one presented in Figure 2.8 – we study each 'Range Spectra' portion separately. Every 'Range Spectra' corresponds to the backscattered energy from that specific range arc, usually simplified to 'Range Cell', where all azimuth bins within the arc are located at the same distance (range) from the radar site.

As we described in Figure 2.6, each Doppler bin in each range spectrum represents a radial surface current for an azimuth somewhere within the range bin. Looking at the radial range bins from the top of the ocean surface we can plot a full map of radial currents, like the radial surface currents map from the SeaSonde radar site near Bodega Bay, CA shown in Figure 2.9.

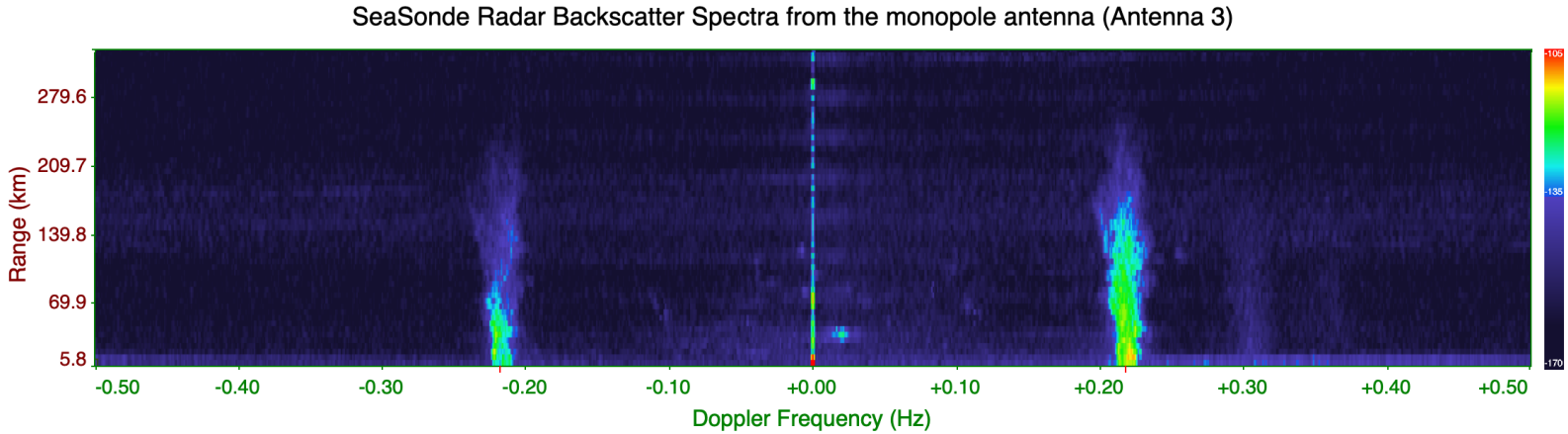

Figure 2.8: Full backscatter spectra energy received at the SeaSonde monopole antenna (Antenna 3). While the vertical axis corresponds to the range from the radar in km, the horizontal axis shows the Doppler frequency of each of the Doppler bins.

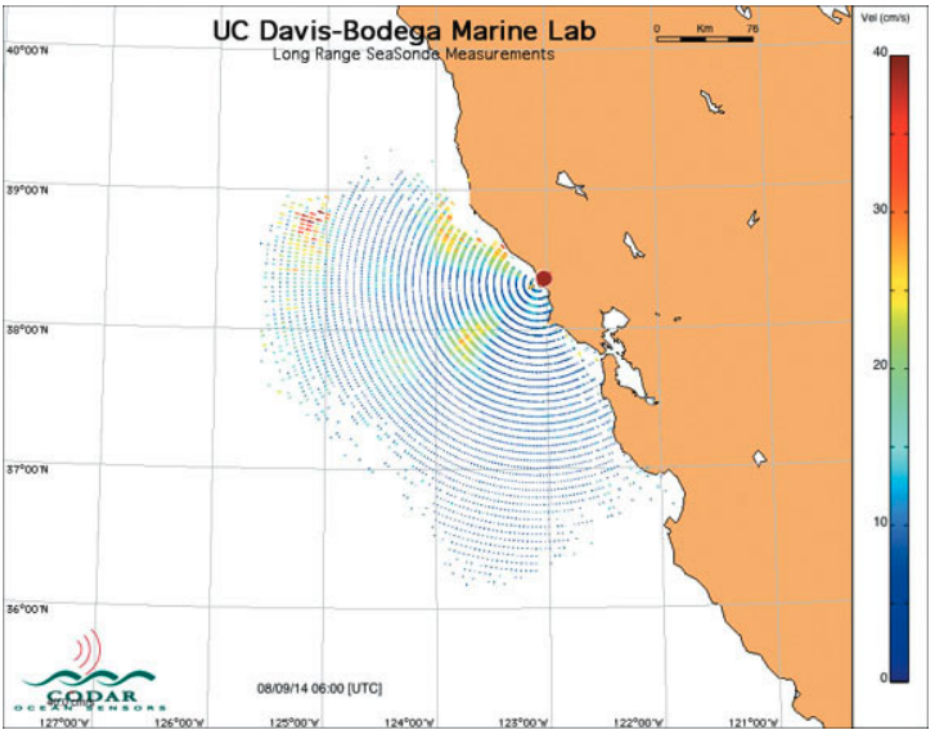

Figure 2.9: Map showing the radial surface currents and the constant range arcs (range cells) at the SeaSonde radar site near Bodega Bay, CA. Each range arc provides a separate 'Range Spectra' (like the ones in Figure 2.7 and Figure 2.8 from which we extract the first-order region boundaries.

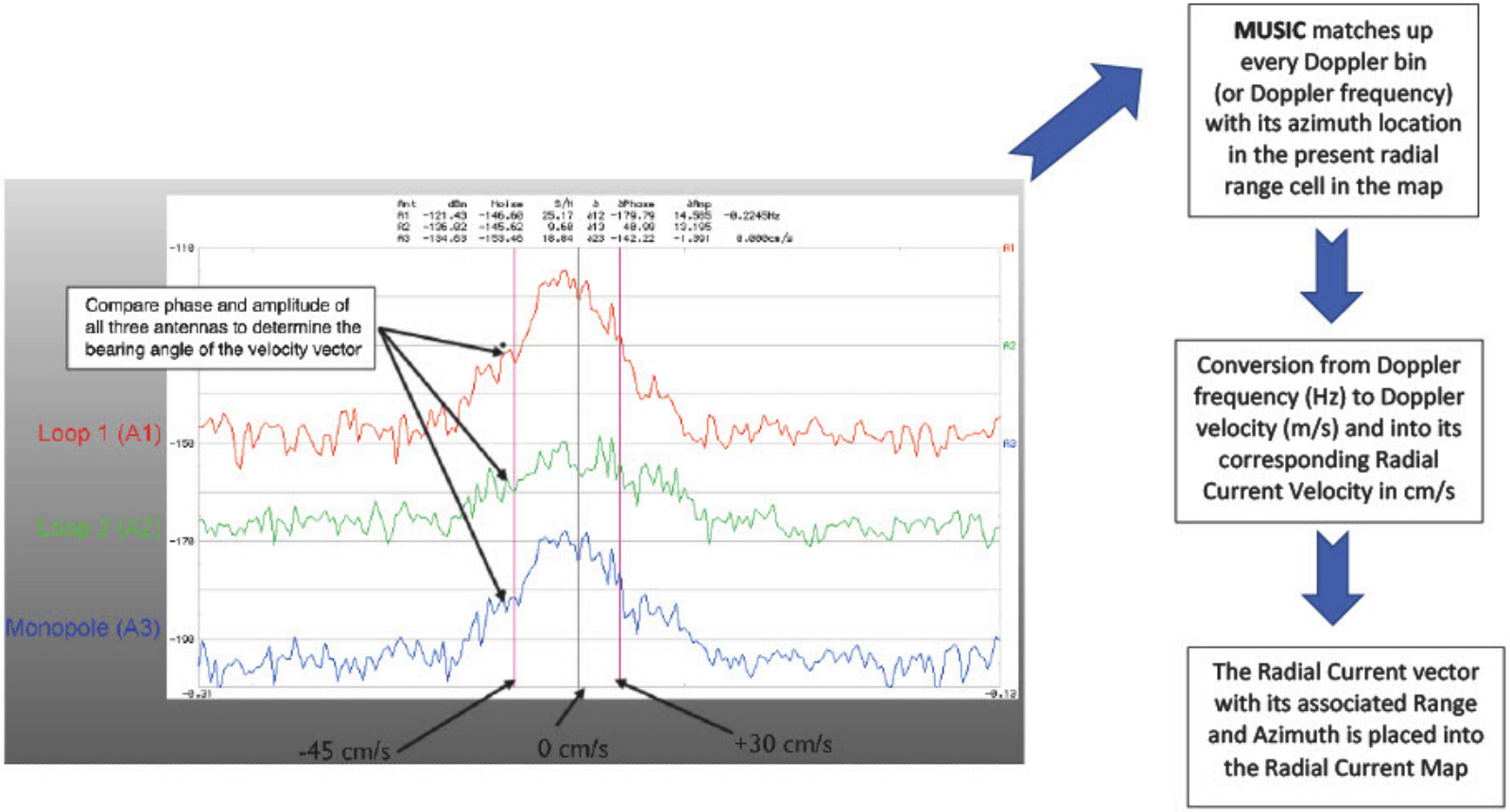

Figure 2.10: From the ocean backscatter of the three antennas for every range cell, we process each of the Doppler bins within the first-order region separately through MUSIC Direction Finding. The information about the ocean surface current they represent can thus be extracted. The system compares phase and amplitude received at the three antennas for that Range-Doppler bin and determines the bearing angle of the radial velocity vector using the MUSIC algorithm.

The radial range cells can be easily identified, but the azimuth of the targets needs to be determined in order for the radial currents to be placed in the map. The incident, azimuth, angle of a received signal can be inferred from phase differences between multiple antennas using the MUSIC algorithm as we will study in Section 2.3.2.

As summarized in Figure 2.10:

- From the ocean backscatter of the three antennas for every range cell, we process each of the Doppler bins within the first-order region separately through MUSIC Direction Finding.
- The information about the ocean surface current they represent can thus be extracted.
- The system compares phase and amplitude received at the three antennas for that Range-Doppler bin and determines the bearing angle of the radial velocity vector.

2.3.1 Beam-Forming and Direction-Finding

Currently, in oceanographic HF radars, there are two major architectures for the receive antennas and they are described next.

Beam-Forming Antenna Arrays: A linear (or phased) array, typically with 8 to 16 antennas spaced one-half wavelength apart. Following the theory of conventional beam-forming processing techniques, the angular resolution of an antenna array of width L, emitting radiation of wavelength λ, is about λ/L radians.

Therefore, the large number of antennas and physical length of the array provides good angular resolution. The disadvantages of a beam-forming array are their high costs, large required installation space and other logistical requirements.

When we work with a phased array of antennas we assume that we know the direction where our beam should be pointing at. We steer the antenna beam to this direction by applying phase shifts to the receive antenna element signals and summing them. We then take the signal representing the echo from that desired beam direction and extract the information, namely, the radial velocity in case of ocean current mapping. The beam-forming method allows for estimating signal from any specific direction chosen, since the incidence angle is an independent variable.

The size of an antenna array required to obtain a good angular resolution for azimuthal direction can be extremely large (more than 100 m) when we study ocean surface gravity waves that have wavelengths in the range of a few to tens of meters or more. For example, in the Strait of Georgia, Bragg scattering from surface gravity waves with $\lambda = 6$ m (period, T = 2 s) requires electromagnetic radiation with a wavelength of 12 m, and to obtain an angular resolution of 6° the antenna array required is about 120 m across, which is in most cases impractical.

Direction-Finding systems: Like a phased array, a direction-finding system employs multiple antenna elements and it also collects the signals received at each individual element and processed them in software to extract the desired information.

However, in a direction-finding system we consider each range bin separately and assume we know the radial velocity of the current: it is the output of the digital spectral processing for each Doppler-shift bin. What we search for is the bearings of each of the incoming signals associated with their respective radial velocities.

While in a beam-forming system the bearing is limited by the resolution, i.e. the beamwidth (as seen in Section 2.2.4); the bearing accuracy in a direction-finding architecture is limited by the signal-to-noise ratio of the system (refer to Section 2.2.7) and we obtain a much better signal bearing accuracy when using direction finding[7].

Barrick[7] compares the two systems with the following example: "A strong ocean current variation within a given radar range cell will produce many separate Doppler-shifted signals, originating from different directions. However, some current patterns could have the same radial velocity (Doppler shift) coming from two or more bearings.

- In beam scanning, two signals at the same frequency (radial velocity) coming from within the same antenna beam can not be separated, unless they are displaced more than one beamwidth.
- In direction-finding, the number of signals at the same frequency that can be uniquely determined is N–1, where N is the number of receiving antenna elements used.

Therefore, an important difference between the two methods is how they handle signals at the same frequency but coming from different directions".

When a single target is generating a return signal, then direction-finding techniques can be used with a much more compact antenna. Radar pulse timing separates targets by range, so the requirement is for *only a single target at a given range*, over all the azimuths returning signals.

Halverson et al.[25] present the following example:
"If an ocean surface current is constant, moving eastward at 10 cm/s, over the radar field of view, which spans 180° of azimuth on a straight east-west coastline, then the radial component of this current at a fixed range would vary as 10 $\cos(\theta)$ where θ is the azimuthal direction, between a value of –10 cm/s in the up-current direction to +10 cm/s in the down-current direction. A particular radial speed, like 2 cm/s, would therefore appear only once in a 180° arc and would be a "single target" in the Doppler bin associated with speeds of 2 cm/s. If the current was moving southwards towards the coast, then all radial speeds would appear as two targets, equally spaced to the left and right of the radar, so that a direction-finding algorithm would need to separate two targets of variable relative power."

In the compact CODAR SeaSonde radar systems, the technique used is indeed direction-finding. The radar return directions are determined by combining the signals from one omnidirectional monopole and two dipole antennas (oriented at 90° to each other). These dipole antennas are co-located with the omnidirectional antenna, all three are mounted in the same vertical axis.

Because of their symmetric and orthogonal design, the three elements do not interact with each other, i.e. there is no mutual coupling between them. Due to its design, the ground footprint of this antenna system can be only a few tens of centimeters across.

The first robust direction-finding algorithm applied to the SeaSonde three antenna system for high-frequency oceanographic current mapping radar was Least Squares, as described by Lipa and Barrick[41]. However, as described next, it was later found that the MUSIC algorithm is more accurate and robust.

2.3.2 The MUSIC Algorithm

As we introduced in Section 2.3.1, the SeaSonde high-frequency radars designed by CODAR use the MUSIC algorithm (MUltiple SIgnal Classification) – developed by Schmidt[61] in 1986 – to provide the most likely direction for a maximum of two targets.

MUSIC is a direction finding algorithm that considers each of the range bins separately. Each spectral point in each range bin where Bragg energy is detected in the Fourier spectrum of the received signal is processed. The MUSIC algorithm then determines the direction of the areas of the ocean surface that are reflecting the radar signal at that Doppler shift.

In other words, MUSIC chooses a radial surface current magnitude and then it determines the bearings of patches of the ocean surface where that particular current magnitude is present (See Figure 2.10).

The number of directions that can be resolved for a single current magnitude is limited to the minimum between (N–1) and M. N is the number of receiving antenna elements and M is the number of data segments averaged in the process of making the measurement. Since a SeaSonde radar has three receiving antennas (N=3), as described in Section 2.3.1, we can infer that MUSIC is indeed able to provide the direction of a maximum of two simultaneous targets. In addition, the performance of the MUSIC algorithm has been found to be very good in examples where the requirements mentioned above are met, i.e. there are a maximum of two targets in a given range and Doppler bin.

The MUSIC algorithm analyses the sample-averaged antenna covariance matrix and determines the locations of the targets that are backscattering the signal. The largest eigenvalues of this covariance matrix are associated with signals and the others are associated with noise. The ability to determine the direction of arrival (DOA) for each of the signal emitters is based on the fact that the signal eigenvectors are mutually perpendicular to each other and to the noise eigenvectors. Provided that the above limit on the number of signal emitters detected is not exceeded and the signal to noise ratio is high enough, the directions of arrival for each of the targets can, in principle, be completely resolved.

The spectrum of Doppler shifts received by the antennas are binned into a discrete set, each associated with a particular radial velocity. By repeating the calculation for all Doppler bins at a particular range, a plot of radial velocity as a function of azimuthal direction can be created, with one (or two) points for each velocity associated with each Doppler bin plotted at their most likely azimuth.

This set of points, irregularly spaced in angular direction, is then smoothed and interpolated to find radial velocities at a fixed set of azimuths for that range, normally every 5°. The whole set of calculations is then repeated for all ranges to build up a radial velocity field like the one we see in the radial currents map in Figure 2.9.

Finally, we obtain a full two-dimensional current velocity vector when we combine the radial velocities from two separated radar systems. The error in this combination is dependent on the system geometry. MUSIC is most accurate when the two radial velocities are in perpendicular directions, and becomes less accurate as the two radials become co-linear. Furthermore, the area of the scattering patch increases with distance from the antenna, and the degree to which the variability is resolved depends on their radial distance[7][25].

Chapter 3

First-Order Boundaries and Summary of Research Approach

The Doppler spectrum of the radar echo –backscattered signal from the sea– at the range of high-frequencies (3 to 30 MHz) presents two main peaks – one from approaching waves and one from receding waves.

Previously, in Section 2.3, we provided an introduction to the radar backscatter spectra in the range of high-frequencies, HF, (3 to 30 MHz). In this Chapter, we first describe in detail the Bragg scattering from ocean waves in Section 3.1. Later, we discuss what are the first and the second order regions of the backscattered spectrum in Section 3.2. Finally, we present our research approach, which is related to the aforementioned first and second order regions and why is it important to be able to separate them.

3.1 Bragg Scattering at the Ocean Surface

3.1.1 Historical Background

Lawrence Bragg and his father William Henry Bragg first proposed what we know as Bragg diffraction in 1913, due to their discovery that crystalline solids produced surprising patterns of reflected X-rays. It was four decades later, in 1955, that Crombie examined the temporal power spectrum of 13.56 MHz radio waves backscattered from the sea and correctly identified Bragg scattering as the simple physical interaction mechanism responsible for the dominant first-order peaks appearing in the HF radar backscatter spectrum. Crombie[19] suggested that "grazing incidence radio waves were coherently backscattered by sea surface waves with a wavelength one half the wavelength of the incident radio wave, propagating radially toward, or away from, the radar."

In 1966, Wait[68] studied the scatter of electromagnetic radiation from a single sinusoidal wave train and found that a principal resonance occurs when the electromagnetic wavelength is double the wavelength of the sea wave. From 1967 until 1972, Barrick and Peake[12][13] and Barrick[2][3] employed a boundary perturbation approach to quantitatively explain first-order sea echo from a random sea surface. Barrick provided a quantitative link between the backscattered energy recorded by radar and the ocean wave-height spectrum of the sea surface from which the energy was scattered. By extending the boundary perturbation approach to second order he showed that a double interaction was responsible for the scattered energy evident in recorded sea echo spectra at frequencies other than those predicted by first-order scattering theory.

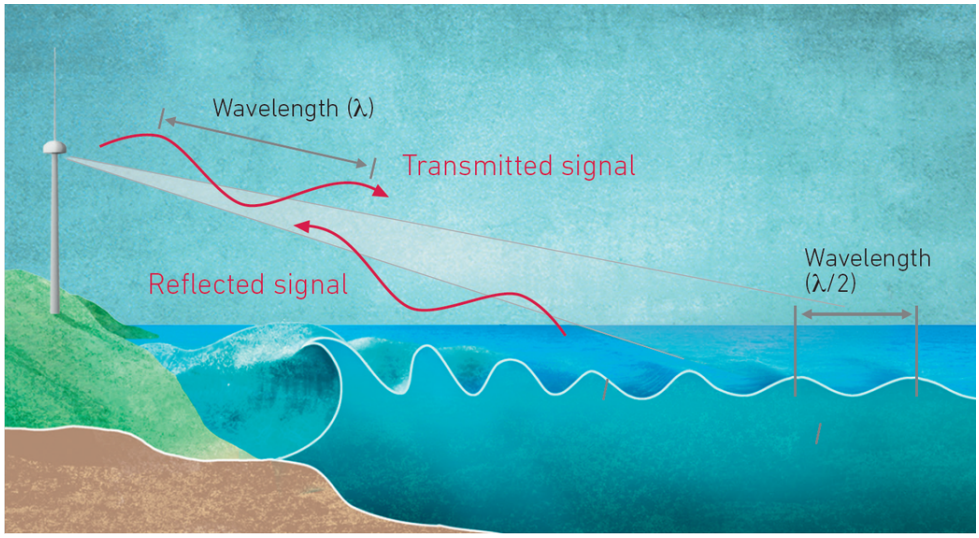

Figure 3.1: (Credit: SOCIB - Coastal Ocean Observing and Forecasting System from the Balearic Islands in Spain) Illustration of Bragg scattering from ocean waves. The transmitted signal from the SeaSonde HF radar travels at wavelength λ and the ocean waves traveling directly towards the radar at half this wavelength $\lambda/2$ will be reflected and added coherently to form a strong echo known as first-order Bragg scattering.

Barrick[2][3][4][5][6][8][9] identified two separate mechanisms which contribute to the second-order backscattered energy: an electromagnetic contribution that takes account of double scatter of the incident radiation by two sea surface waves which results in coherent backscatter, and a hydrodynamic contribution which results from the backscatter of the incident radiation from second-order ocean waves which are the product of the nonlinear interaction of two sea surface wave components. Barrick developed HF radar techniques for extracting r.m.s. wave height, mean wave period and the non-directional wave-height spectrum, which involved a weighting function to take account of the second-order coupling coefficient.

3.1.2 The Backscatter of Ocean Waves

As described by Kingsley and Quegan[34], waves on the ocean are generated by local wind (wind waves) or distant storms (swell). The motion of an ocean wave is carried by particles of water exhibiting, roughly circular motion as the wave travels passes. Surface gravity waves, both wind waves and swell, propagate along the interface between water and air under the predominating influence of gravity. At the surface of the sea, these circular motions of the water particles can have quite large amplitudes but they quickly fade away with depth.

Any of the waves travelling perpendicular to the radar beam will produce backscatter and part of the signal will be reflected back to the radar. But only those having a length equal to half the radar wavelength ($\lambda/2$) will generate a much stronger echo, as illustrated in Figure 3.1. This is because their reflected radio waves add together coherently, on what is known as Bragg resonant scatter.

For surface radio waves Bragg resonance occurs when the distance between the echoes from two ocean wave crests is a multiple of $\lambda/2$, half the radar wavelength. As we see in Figure 3.2, and by simple trigonometry, we can infer that

$$L_n = \frac{n\lambda/2}{\sin\theta} \qquad \text{(m)} \tag{3.1}$$

and when the radar is at a far enough distance $\theta = 90$ degrees, thus $\sin\theta = 1$.

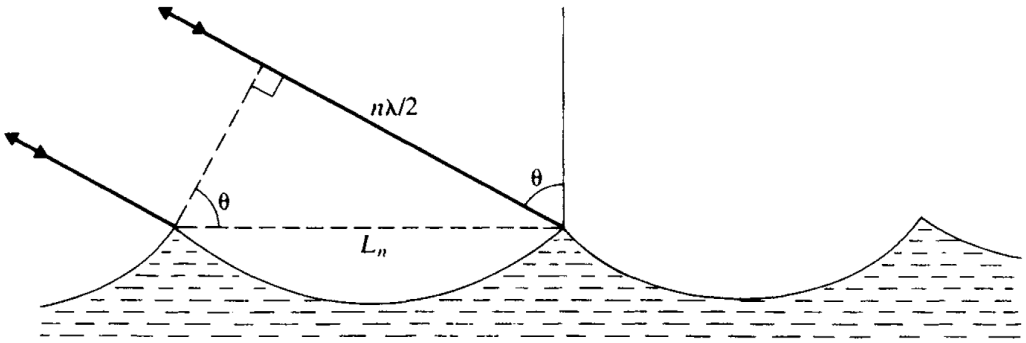

Figure 3.2: (Credit: Kingsley and Quegan[34]) The distance between two wave crests is L_n, the radar signal incidence angle is θ and the path difference between their two echoes is $n\lambda/2$ i.e. a multiple of half the radar wavelength. When this occurs in a periodic medium as the ocean surface, the backscatter will be added coherently in what we know as Bragg resonance.

For an ocean gravity wave of wavelength L_n, we find the phase velocity V_n through the dispersion relation in Eq.(3.2), where g is the acceleration due to gravity.

$$V_n = \sqrt{g\frac{L_n}{2\pi}} \qquad (\mathrm{m\,s^{-1}}) \tag{3.2}$$

The Doppler formula describes how the radar Doppler shift f_d relates to the radial ocean wave speed v_r and the radar wavelength λ by

$$f_d = \frac{2v_r}{\lambda} \qquad (\mathrm{Hz}) \tag{3.3}$$

and thus, the Doppler frequency shift that corresponds to the Bragg resonant ocean wave of wavelength $n\lambda/2$ will be:

$$f_n = \pm\frac{2(V_n \sin\theta)}{\lambda} \qquad (\mathrm{Hz}) \tag{3.4}$$

Combining Equations (3.1), (3.2) and (3.4) we obtain Eq.(3.5), where we observe that the Doppler shift from a wave of wavelength $n\lambda/2$ has a square-root dependence on the radar frequency $f_{radar} = c/\lambda$.

$$f_n = \pm\sqrt{ng\frac{f_{radar}\sin\theta}{\pi c}} \qquad \text{(Hz)} \tag{3.5}$$

For the HF frequency band, the corresponding wavelengths of the Bragg resonant ocean waves are approximately between 5 and 50 meters. The associated phase velocities, following Eq.(3.2), are about 3 to 10 m/s. The range of frequencies of these Bragg resonant ocean waves is about 0.2 to 0.6 Hz, and their periods roughly 2 to 6 seconds.

Figure 3.3 presents a typical sea echo spectrum for a radar working at 25 MHz frequency, and we observe two strong peaks.

- Peak located in the positive Doppler region, at about +0.45 Hz. This peak is due to Bragg resonant waves propagating towards the radar.

- Stronger peak found at about -0.55 Hz. This negative Doppler peak is due to Bragg resonant waves traveling away from the radar.

The large ratio of the negative peak height in Figure 3.3 respect to the positive peak height is indicative that the receding waves – propagating away from the radar – have larger wave heights than the advancing waves traveling towards the radar. This is happening most probably due to wind blowing in the direction away from the radar.

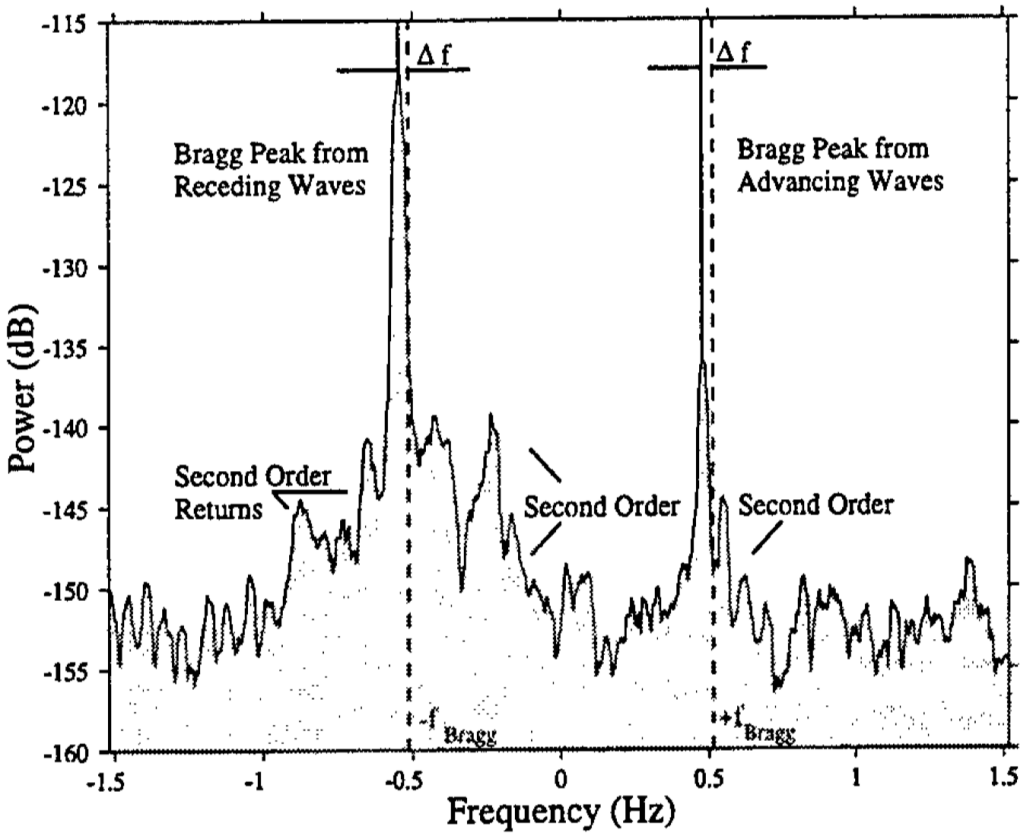

Figure 3.3: (Credit: Paduan and Graber[52]) Sample radar backscatter spectrum at a $f_{radar} = 25$ MHz and thus $\lambda_{radar} = 12$ m. In this section we follow the calculations for the theoretical first-order Bragg frequencies, which correspond to wavelength $\lambda_{Bragg} = 6$ m and thus their Doppler shift $f_{\text{Bragg}} = \pm\ 0.5098$ Hz.

3.1.3 Example: Bragg Calculations for a 25 MHz Radar

Next, similar to what we observe in Figure 3.3, we present an example of the first-order Bragg calculations to determine f_1 i.e. the theoretical Doppler frequency shift of the first-order Bragg wave (f_{Bragg} in the Figure). The theoretical f_1 frequency is a calculation for zero surface current. As shown in Figure 3.3, a surface current will shift the actual first order Bragg region away from the theoretical f_1 location for zero current (dashed line).

$$f_{radar}\text{=25 MHz, } c\text{=3x10}^8 \ m/s, \ g\text{=9.8 } m/s^2$$

$$\lambda_{radar} = \frac{c}{f_{radar}} = \frac{3 \times 10^8}{25 \times 10^6} = 12 \text{ m}$$

Thus, the first-order Bragg wavelength λ_{Bragg} will be L_1 as we saw in Equation (3.1) and with the order n=1.

$$L_1 = \lambda_{radar}/2 = 6 \text{ m}$$

Continuing with Equations (3.1), (3.2), (3.4) and (3.5), we obtain:

$$V_1 = \sqrt{g\frac{L_1}{2\pi}} = \sqrt{g\frac{6}{2\pi}} = 3.059 \text{ m/s}$$

and the Doppler shift for the first-order Bragg wave:

$$f_1 = \pm\frac{2(V_1 \sin\theta)}{\lambda_{radar}} = \frac{\pm V_1 \sin\theta}{L_1}$$

$$f_1 = \frac{\pm 3.059 \cdot \sin(90^\circ)}{6} = \pm 0.5098 \text{ Hz}$$

3.2 First and Second Order Regions in the Bragg Backscatter

The spectra presented in Figure 2.8 corresponds to the backscattered energy received by Antenna 3 (the monopole antenna of the SeaSonde radar system) for all of its ranges, as shown in the vertical y-axis of the figure. During this research, we study **the monopole backscatter energy of each of these radial range arcs separately to extract the corresponding first-order region boundaries at that distance from the radar**[9][10][40][41][43].

As an example, Figures 3.4, 3.5, 3.6 and 3.7 have a Doppler resolution (`fftlen`) of 1024 Doppler bins and they all correspond to the 16th constant-range arc, usually known as *"Range Cell"* of that radar site (Refer to Figure 2.9 for a description of those range cells). The x-axis on this type of spectra graphs can be centered around the four different aspects described in the figures and we describe their conversions in Algorithms 7 through 11.

3.2.1 First-Order Bragg Region: Measuring Ocean Currents

The first-order region of the Doppler frequency spectra corresponds to the backscattered energy due to the radar electromagnetic wave coherently reflecting off those sea surface waves having a wavelength equal to half the incident radar wavelength and thus satisfying the Bragg resonant condition[31][67].

The location of the Bragg peaks we observe in the backscatter spectrum in Figure 3.3 is determined by the phase speed of the resonant waves relative to the radar.

The effect of ocean currents present is to shift the frequency spectrum by an amount equal to the radial component of the surface current. That is why we see the first-order Bragg peaks with a frequency shift Δf respect to the zero-current first-order advancing and receding Bragg resonant waves i.e. $\pm f_0$ or $\pm f_{\text{Bragg}}$ shown in the figure as dashed lines.

As developed in the example above, the theoretical frequency of the first-order Bragg waves ($\pm f_0$ or $\pm f_{\text{Bragg}}$) is due to the phase speed of the Bragg resonant wave traveling in still water, predicted by the dispersion relation for gravity waves on the ocean surface. Deviations from this expected Doppler shift can be attributed to the radial component of the surface current relative to the radar. Therefore, the surface current measurements are made by estimating the difference between the measured frequency shift of the Bragg peaks and the theoretical $\pm f_{\text{Bragg}}$, i.e. the frequency difference Δf that we observe in Figure 3.3.

In addition, we see how the Doppler shifts Δf appear to the left of the theoretical Bragg waves, which means that the ocean current is moving away from the receiver, which agrees with our deduction of the wind blowing away from the receiver as well. If the wind was blowing predominantly towards the radar, then the region of the advancing Bragg waves (positive Doppler frequency) would be stronger than the receding one (negative Doppler).

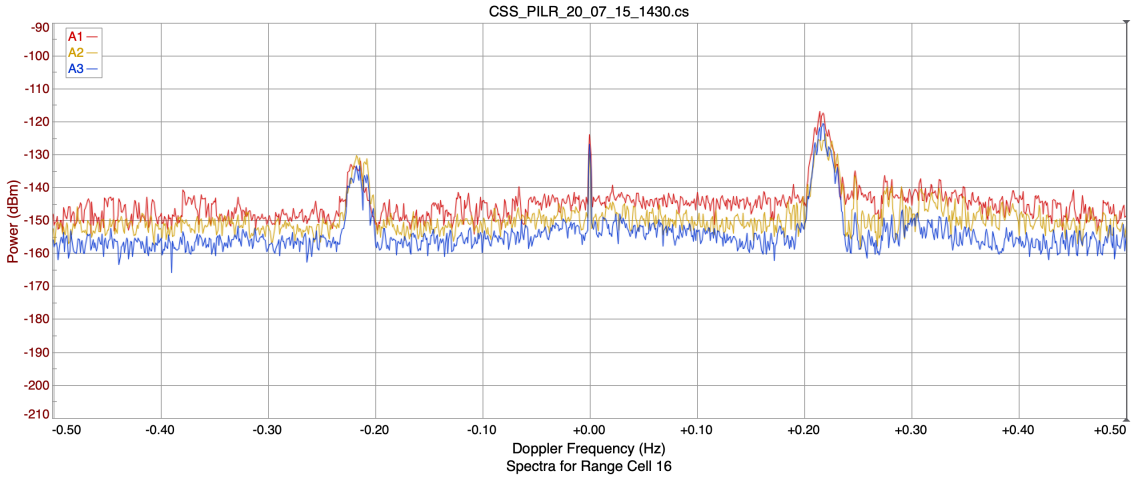

Figure 3.4: Example of the backscattered received by Antenna 3 (monopole) at the 16th range of the SeaSonde PILR site in Pillar Point, CA. In this first figure, the x-axis corresponds to the Doppler frequency shift in Hz.

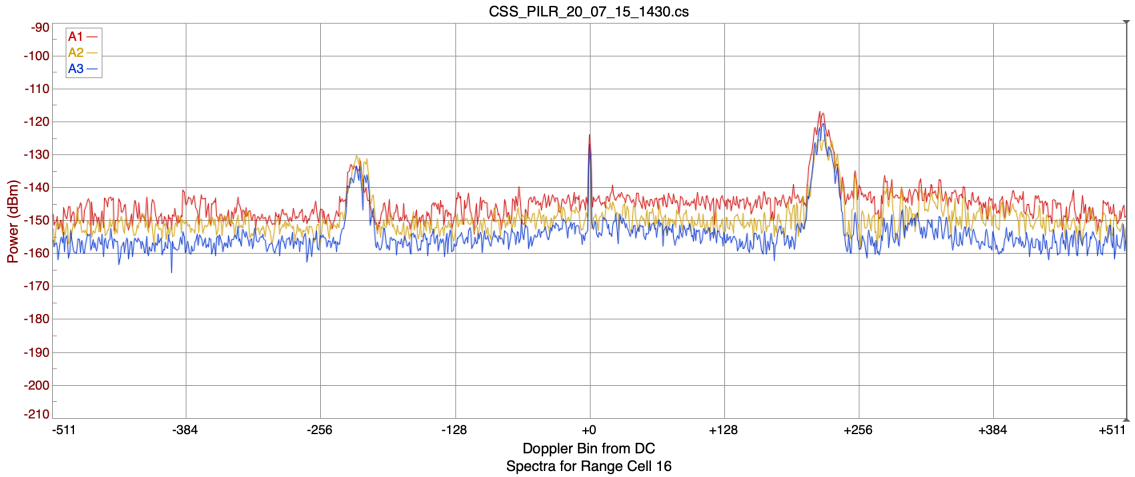

Figure 3.5: Similarly to Figure 3.4, this is the backscattered signal received by Antenna 3 (monopole) at the 16th range of the PILR SeaSonde site in Pillar Point, CA. In this second figure, however, the x-axis is set as the equivalent Doppler bin location from the 0 at DC.

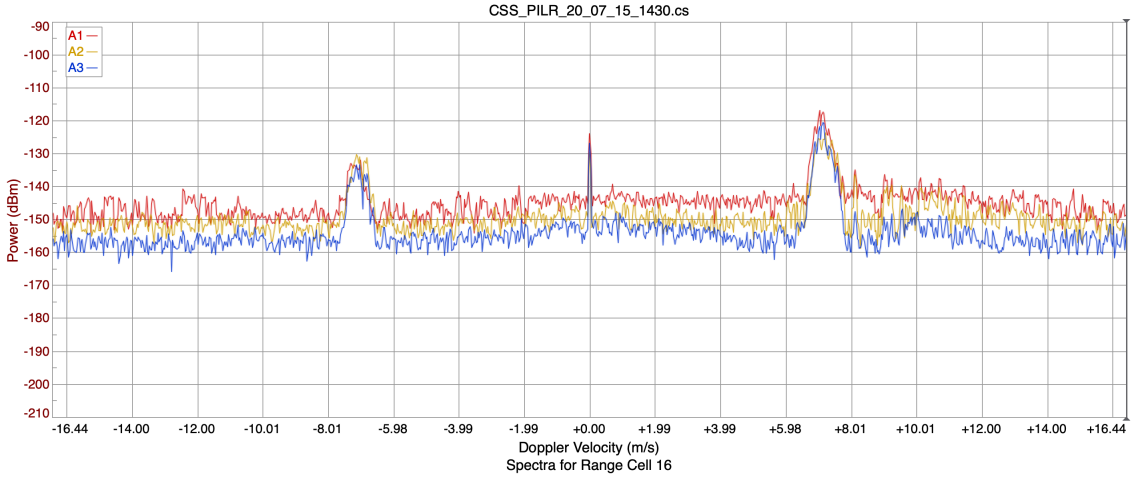

Figure 3.6: Similarly to Figure 3.4, this is the backscattered signal received by Antenna 3 (monopole) at the 16th range of the PILR SeaSonde site in Pillar Point, CA. In this third figure, however, the x-axis is set to be the equivalent Doppler Velocity in m/s.

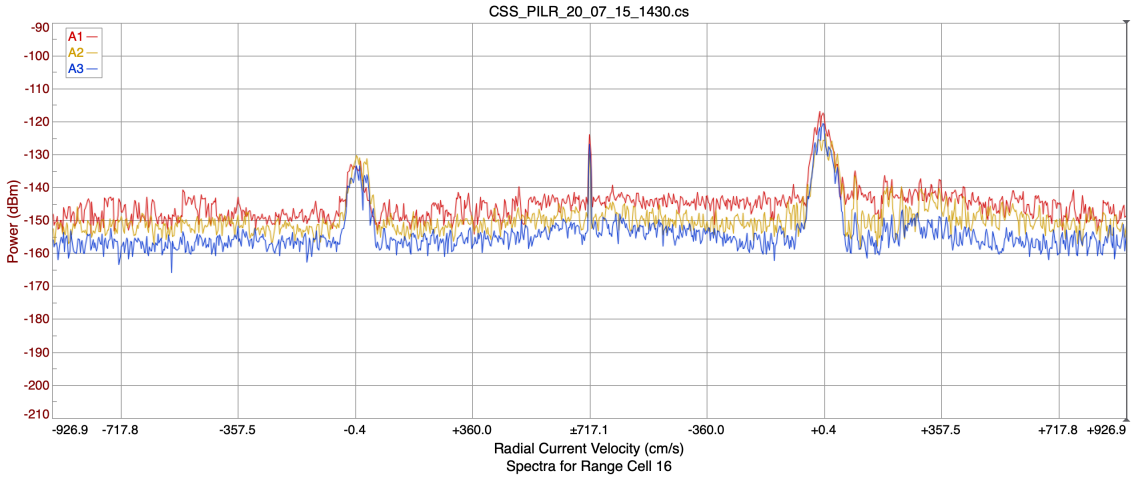

Figure 3.7: Similarly to Figure 3.4, this is the backscattered signal received by Antenna 3 (monopole) at the 16th range of the PILR SeaSonde site in Pillar Point, CA. In this fourth figure, however, the x-axis is set to be the equivalent Radial Current Velocity in cm/s.

3.2.2 Second-Order Bragg Region: Measuring Waves

A second region of stronger backscattered energy is visible in the spectrum when the transmit signal is strong enough and the sea state is high enough – often the case during storms or strong winds. These second-order region will present peaks as well, which are due to multiple reflections from ocean waves of all wavelengths traveling in all directions. The power of these second-order peaks are typically 20 to 40 dB below the first-order peaks in the spectrum, like we observe in Figure 3.3.

Starting in the 1970s, Barrick and Lipa studied the connection between this second- order region of the sea echo Doppler spectrum and the directional spectrum of the ocean waves[2][3][4][5][6][1][8][9], and also proposed methods of inversion to extract the ocean wave directional spectrum from the second-order region of the sea echo[10][40][41][43][45]. In addition, Wyatt[71][72][73][74][75][76][77] also performed extensive research in the 1980s and 1990s about measurements of the ocean wave directional spectrum from HF radar Doppler spectra.

One of our objectives in this thesis is to develop methods to exclude second-order peaks from consideration in determining surface currents. This becomes important as wave height increases and for situations where there are multiple peaks of first-order Bragg scattering, as discussed next.

As an example for second-order regions describing ocean waves, Figure 3.8 below (credit to Lipa[43]) presents an instance of SeaSonde backscatter spectra where we observe the first and second- order regions for the three antennas.

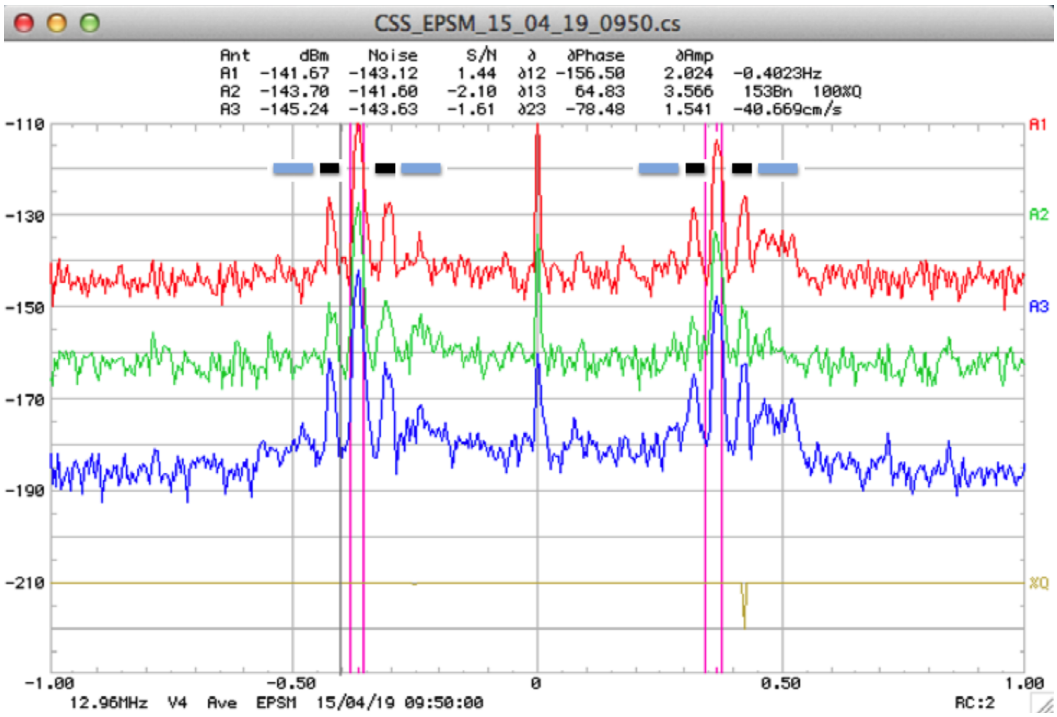

Figure 3.8: (Credit to Lipa[43]) Example of SeaSonde cross spectra at the EPSM site in Portugal for April 19, 2015, 9:50 am. Antenna Loop 1 is shown in red, while Antenna Loop 2 is presented in green, and finally the Monopole Antenna 3 in blue. The first-order regions are delimited by the first-order boundaries shown as magenta vertical lines. The spectra exhibit narrow spectral peaks close to the aforementioned first-order Bragg region produced by swell in addition to the broader echo from wind-waves further from the first-order Bragg region. The color bars at the top indicate the regions in the second-order spectrum dominated by swell (black) and wind-waves (blue).

The first-order regions in Figure 3.8 are delimited by the first-order boundaries shown as magenta vertical lines. In addition, the three spectra exhibit narrow spectral peaks close to the first-order Bragg region produced by swell in addition to the broader echo from wind-waves further from the first-order Bragg region.

3.3 First-Order Boundaries and Split Bragg

3.3.1 Separating the First and Second Order Bragg Regions

The frequency shift of the first-order scatter from its theoretical still-water value in deep water, known as Doppler shift, is related directly to the vector component of the sea surface current moving toward or away from the radar site. In consequence, studying the Doppler shift present at the first-order region allows multiple oceanographic radars to generate two-dimensional maps of ocean surface currents.

The second and higher-order scattering, manifesting as side-bands of the first-order echo, are caused by coherent reflections from all sea surface waves and provide, after analysis, more information on the sea surface properties and the directional wave-height spectra.

Estimation of the first-order line boundaries – that sometimes are referred to as first-order lines or FOL – shown as pink vertical lines in Figure 3.9, is therefore key in the task of mapping ocean currents correctly.

Setting a first-order boundary from the monopole antenna spectrum implies setting both positive and negative thresholds, surrounding a Bragg peak (for still water, zero surface current), that limit the largest Doppler shift (hence the maximum corresponding radial surface current) accepted for processing to retrieve surface currents. It is important to bound the first-order region as accurately as possible so that second-order echoes that could lead to erroneous ocean current estimates are excluded properly.

The radial components of surface currents corresponding to Doppler shifts larger than the threshold set by that first-order boundary will therefore be rejected. In turn, accepting erroneous radials would be directly including incorrect and strong radial currents, due to their Doppler shift being outside the thresholds.

Once we have determined the location of these boundaries for a specific range bin, we select each Doppler bin within them that produces an echo above the noise floor. From the backscatter spectral level from each of the Doppler bins we obtain an estimate of the radial surface current for some azimuth location within the range bin arc of the range cell we are working with.

Next, the MUSIC algorithm will determine the azimuth location of all of the retrieved surface current radial component using a state-space method of direction finding. More details about the MUSIC algorithm can be found on Section 2.3.2, or the original paper by Schmidt[61] and on the CODAR Ocean Sensors Ltd. U.S. Patent by Barrick and Lipa[11].

3.3.2 Bragg Splitting: a New Challenge

The dominant peaks of the backscattered Doppler spectrum (the "first-order Bragg region") are generated, as we learned in this chapter, by the ocean wave spectrum components with wavelength $\lambda/2$ moving forward or receding from the radar. Under normal conditions, the effect of a surface current on the position of the Bragg peaks is stable, and the dominant Bragg peak appears as a narrow-band single spike. Figure 3.10 presents what we know as a typical Doppler backscattered spectrum, where the two first-order Bragg regions are easily identifiable around ± 0.25 Hz.

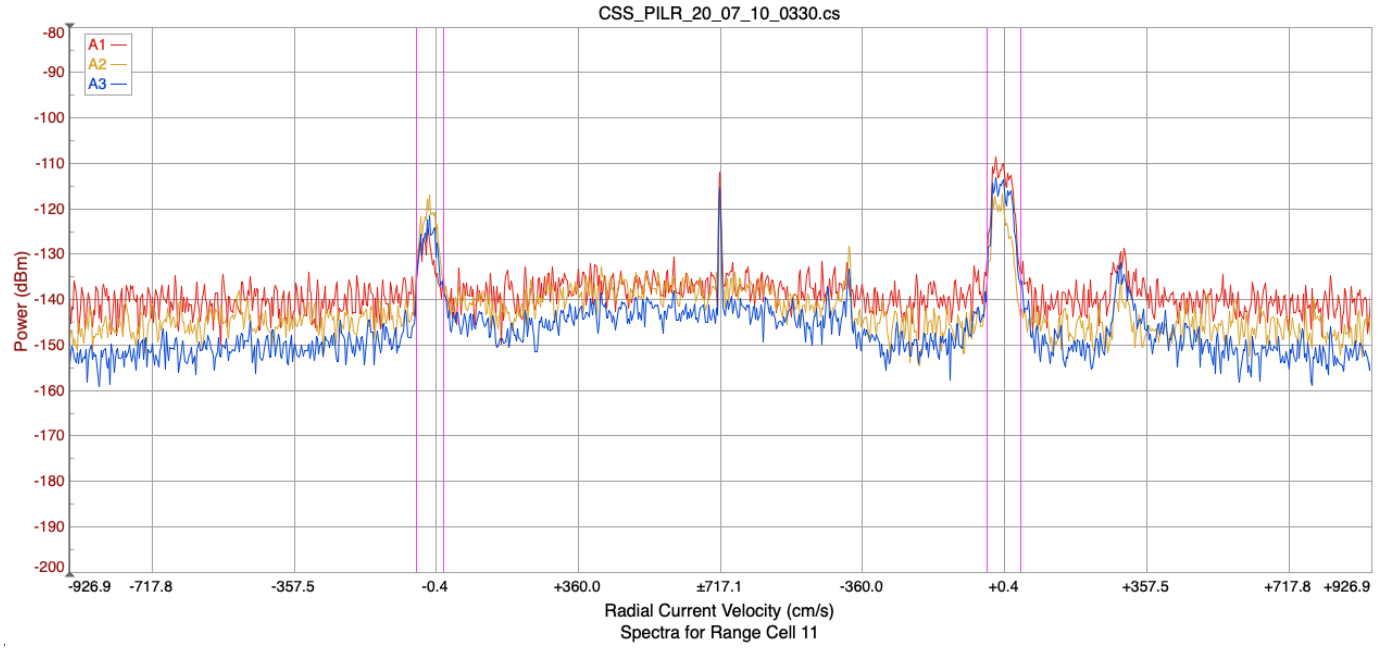

Figure 3.9: Example of the 11th range cell Doppler spectrum at the PILR radar site at Pillar Point, CA. The backscatter from the three antennas is shown as red, yellow and blue lines and the determined first-order boundaries are the vertical pink lines.

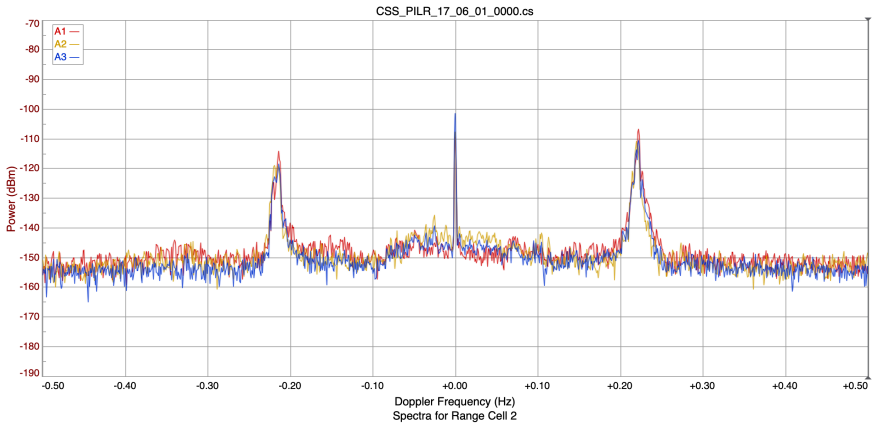

Figure 3.10: Example of the Doppler spectrum of the 2nd range cell at the PILR radar site in Northern California where we can see the first-order Bragg regions as almost symmetrical peaks on both sides of the Doppler spectrum. The backscatter from the three antennas is shown as red, yellow and blue lines and the first-order Bragg peaks appear around the ± 0.25 Hz Doppler frequency, 40 to 50 dB above the noise floor.

However, these first-order Bragg regions of the spectrum will migrate, spread or even split into double-peaked or multiple peaks in the presence of a strong horizontal surface current shear, eddy or some other non-uniform surface current fields across the radar field of view.

In addition, the first-order region splitting can also be caused by geographical features at the location where the radar was installed, and how the ocean currents behave under those conditions as opposed to their usual circulation in the open ocean. For example, the presence of the Sangiang island in front of the radar at the Sunda Strait in Indonesia causes the first-order region to split in certain ranges.

Another site that is particularly affected by the Bragg regions splitting is the VCOL site located at the Straits of Georgia in British Columbia, Canada. Figure

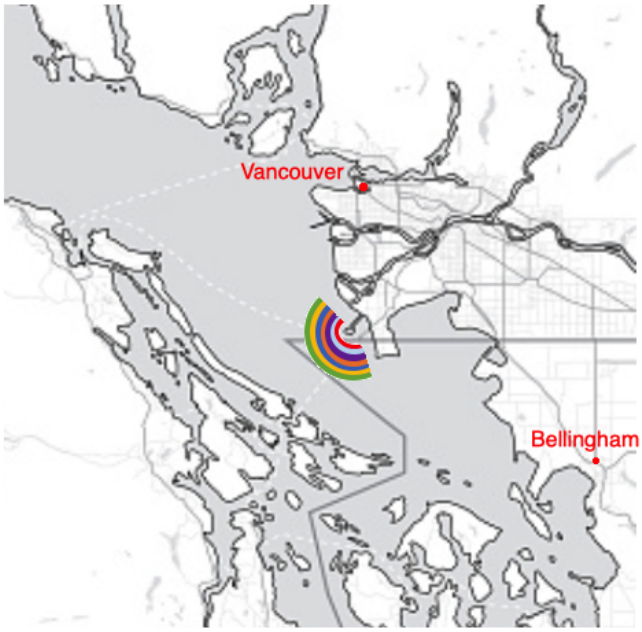

Figure 3.11: Map of the VCOL SeaSonde radar site location in the Straits of Georgia in British Columbia, Canada. The figure also presents semicircular arcs in different colors that represent some of the range cells regions that reflect the backscatter energy received back at the radar site.

3.11 presents a map of the VCOL radar site location in the straits and in addition we have added some range arcs in different colors that represent some of the range cells regions that reflect the radar backscatter.

The phenomenon of a Bragg peak splitting into a multiple peaked Bragg is presented in Figure 3.12 corresponding to the VCOL site. As we observe in the figure, it imposes a problem for surface current mapping since it increases the difficulty of determining the correct Doppler location of the first-order Bragg boundaries –described in the previous section– that separate this region from the second-order spectrum, and therefore the radial components of the currents present in the field of view of the radar.

The backscattered spectra shown in Figure 3.12 corresponds to "Range Cell 18" of the VCOL radar site at the Straits of Georgia. For this radar configuration, the width of a range cell is approximately 500 meters, therefore the 18th cell corresponds to a range of about 9 km from the radar.

Note how the positive (approaching waves) Doppler spectrum splits into two components. The question arises as to whether the lower peak (left) is first-order and hence used to retrieve surface currents or not.

Figure 3.13 presents the whole backscattered energy spectrum received on Antenna 3 (the monopole) for all range bins out to range 44.5 km for the same day and time as Figure 3.12, July 1st 2020 at 07:50 UTC.

We observe that the splits of the energy on the right (advancing) first-order Bragg region start to manifest around 7 km range from the site. In addition, multiple splits can be found further in range, above 14 km from the radar.

On the other hand, the split of the left (receding) first-order Bragg region starts around the 12 km range from the radar's site. Furthermore, we can see how the corresponding receding radial currents for the left side of this left Bragg (receding Bragg wave) are weaker than the right portion of the split, which is stronger. This means that ocean currents present underneath the receding Bragg wave are predominantly advancing towards the radar, and their current velocity is therefore positive.

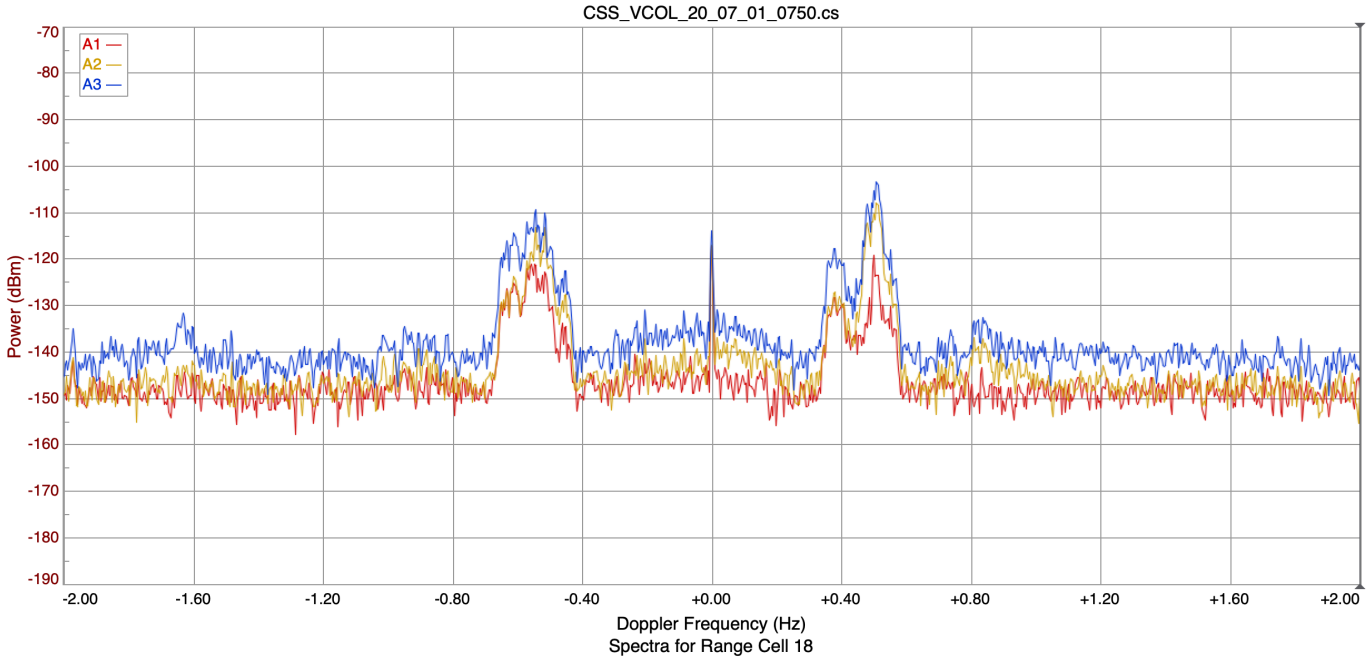

Figure 3.12: Example of the 18th range Doppler spectrum of the VCOL radar site in the Straits of Georgia where we can see the right (advancing) first-order Bragg region splitting. The backscatter from all the three antennas is shown as red, yellow and blue lines and the splitting of the Bragg into a double-peaked Bragg occurs around the positive 0.4 Hz Doppler frequency.

In these figures where we can see the whole backscatter spectrum of an antenna, the red tick marks (± 0.5 Hz) on the horizontal axis mark the Doppler shift of the waves for zero surface current, i.e. still water.

As we will study in Chapter 4, it will be of particular interest to study the radial currents present around the Bragg regions splits since in fact, this is where the *conventional* method is not be able to track them properly.

3.4 Research Approach

As we have described in this chapter, setting the first-order boundaries correctly is key for the surface current retrieval process. The *conventional* method that has been used to determine the first-order boundaries has a series of limitations, which lead to some erroneous boundary determinations in challenging scenarios like storms, strong shear, tides, or the presence of an island across the radar field of view. In addition, there are radar sites where the backscattered Bragg regions present splits (discussed in Section 3.3.2), which further complicate the determination of the first-order boundaries by the *conventional* method.

Improving the determination of the boundaries between the first and second order region is therefore the heart of the problem investigated in this thesis. We develop an *alternative* method in order to avoid some of the known shortcomings of the *conventional* method and consequently improve the surface current estimation results.

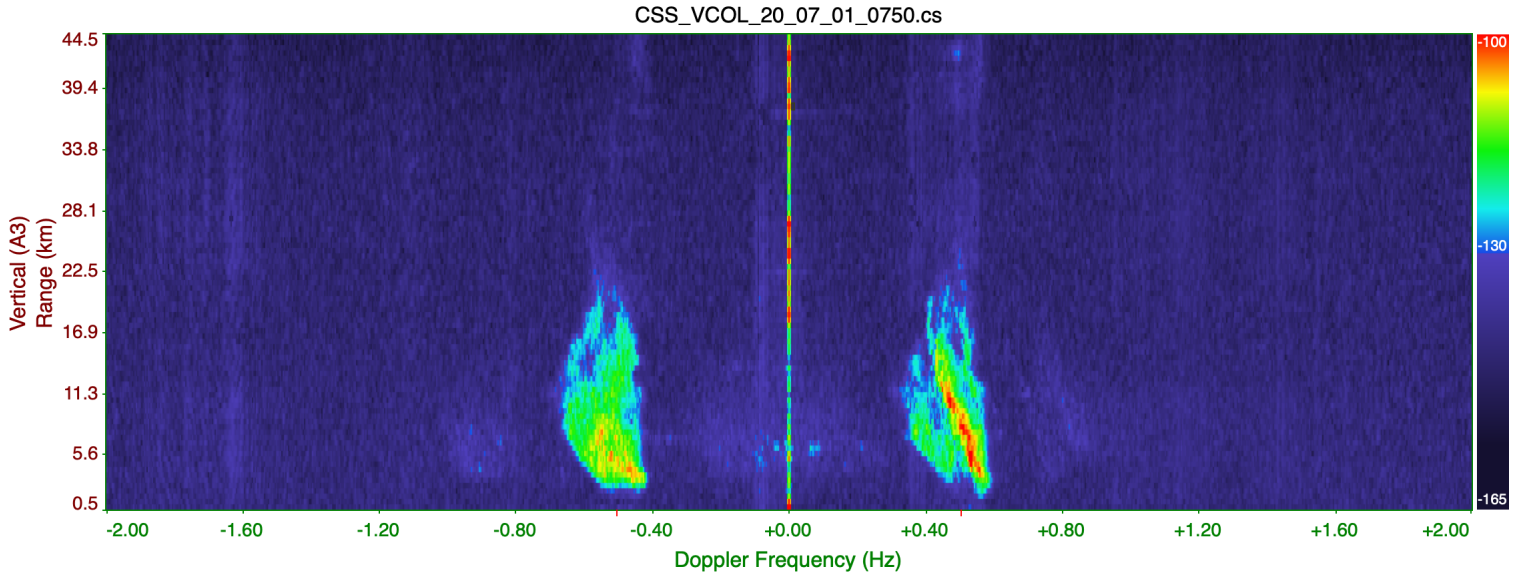

Figure 3.13: Backscatter spectrum at the monopole Antenna 3 of the VCOL radar site in the Straits of Georgia. Starting around the 7km range, we observe the right (advancing) first-order Bragg region splitting. On the other hand, the split of the left (receding) first-order Bragg region starts around the 12 km range from the radar's site. The red tick marks (+/- 0.5 Hz) on the horizontal axis mark the Doppler shift of the waves for zero surface current, i.e. still water.

Finally, a summarized list of the research steps conducted in this thesis and their corresponding chapters is presented next.

List of Research Steps:

1. Develop an algorithm to find the initial estimates of the first-order boundaries by using a more complex null detection method. *(Chapter 4)*
2. Description of the Bayesian estimation techniques studied. *(Chapter 5)*
3. Combining the null-finding algorithm and the diverse Bayesian estimation techniques into a Python script that determines the first-order boundaries. *(Chapter 6)*
4. Bayesian tracking application for our first-order boundaries problem. *(Chapter 7)*
   (a) Probabilistic Data Association Filter (PDAF) to resolve multiple options for a boundary
   (b) Bayesian tracking in time and range to provide the final estimates for the boundaries
5. Demonstration of the validity of the method: *(Chapters 8 and 9)*
   (a) Statistical comparisons against conventional method results.
   (b) Comparisons of radial velocity current results against some drifter measurements obtained in the Straits of Georgia.

Chapter 4

First-Order Boundaries Estimation Algorithms

Improving the determination of the boundaries between the first and second order regions of the Bragg backscatter from the ocean is the heart of the problem investigated in this thesis. Some of the known shortcomings of the *conventional* method discussed in Section 3.4 complicate the determination of the first-order boundaries, and on that account we developed an *alternative* method described in the present chapter with the objective of overcoming those limitations.

Previously in Chapter 3, we described in detail the Bragg scattering from ocean waves in Section 3.1. In addition, we discussed the first and the second order regions of the backscattered Bragg spectrum in Section 3.2 and concluded that determining the first-order boundaries locations correctly is key for a successful surface current retrieval process.

Our improved technique to find the *nulls* (local minimums) that separate the first and second order regions of the Bragg backscatter is described in this chapter. It follows some fundamental theoretical principles that distinguish the regions; those principles were introduced in Section 3.2 and will be outlined again in Section 4.2 as the basis of the null-finding algorithm.

The aforementioned principles are the foundations for the *conventional* method as well, and on that account the core of the *alternative* method is definitely based on its predecessor developed by CODAR Ocean Sensors Ltd. As a reminder, the *conventional* method that is currently being used on SeaSonde software does succeed at most of the radar sites, but it has been observed to provide erroneous boundary determinations in challenging scenarios, like the ones listed below.

- Radar sites where the backscattered Bragg regions present either splits or has unusual inclined shapes.
- Radar sites where there is an island on its field of view and consequently the water does not circulate as usual.
- Radar sites when the following occur:
    - storms,
    - strong shear,
    - tides, etc.

Summarizing, the *alternative* method described in this Chapter

- ◇ is based on the same theoretical principles producing the Bragg regions as the *conventional* one;
- ◇ applies a characteristic normalization of the signal leading to a more consistent determination between ranges;
- ◇ determines the noise floor differently, resulting in better determination at longer ranges;
- ◇ uses a generally more complex search of the typical nulls in the signal between the regions that allows temporary multiple locations to be considered;
- ◇ searches for split Bragg regions and stores the whole set of possible boundary locations;
- ◇ includes Bayesian estimation techniques that will weight the possible locations and decide the most probable ones;
- ◇ and takes into consideration temporal and spatial previous conditions of the first-order boundaries by applying Bayesian tracking in both time and range;

which all leads to determining more accurate estimates of boundaries separating first and second order Bragg backscatter and consequently improves the surface current estimation results.

4.1 Overview of the Null-Finding Algorithm

Taking into consideration the three fundamental features that we describe in Section 4.2 and that distinguish the first and second-order regions of the backscatter spectrum, we decide that the null-finding algorithm will search for those *nulls* (local minimums) in the backscatter spectrum of "SeaSonde Antenna 3" which is the monopole antenna. We use the monopole antenna since it observes echoes from all azimuth directions equally. Therefore, the null-finding algorithm steps described below are applied to the entire monopole self-spectrum[9][10][40].

In addition, we need to extract a series of parameters specific to the radar site, as we will describe in detail in Section 6.2.

– From the `Header.txt` file in the radar data file we extract:

- **fftlen:** Number of Doppler frequency bins, each bin represents the amount of energy that the signal has at that particular Doppler frequency.
- **noisefact:** Noise level factor used to threshold the spectra.
- **nsm:** Amount of smearing points chosen to smooth the data.
- **currmax:** Absolute maximum ocean current velocity expected at the radar site.
- **fdown:** Factor to protect first-order region surrounding the Bragg peaks.
- **flim:** Factor to define minimum power considered first-order which allows rejection of weaker surrounding side-bands.

– From the `FirstOrderConfigs.plist` file in the radar data collection file we extract:

- **fdown**, **flim** and **currmax**: If different, this values will overwrite the ones on the `Header.txt`.
- **freject:** Minimum threshold of power difference in dB for the backscatter spectrum of a range to be taken into consideration.
- **minBinsBragg:** Minimum amount of Doppler bins for a section to be considered a split Bragg region.
- **minSplitBragg:** Minimum power magnitude below the Bragg peak for a section to be considered a split Bragg region.
- **Bayesian estimation parameters** (Described in Chapter 7 Section 1): For both **Time Tracking** and **Range Tracking** we extract
  - **Mode Selection**
  - **Initial State Mean Value**
  - **Initial State Covariance**
  - **Initial Model Behavior**
  - **Initial Model Covariance**
  - **New Measurement Covariance**
  - **State to Measurement**

The complete algorithm to obtain the final first-order boundaries is divided into the following steps described in the List of Algorithm Steps below.

List of Algorithm Steps:

1. Extract the backscatter spectrum of the monopole antenna
2. Normalize the spectrum
3. Apply a noise threshold
4. Smooth the spectrum
5. Limiting the spectral range on both sides
6. Find the possible nulls (*initial* estimates of the first-order boundaries)
7. Search for *split Bragg* regions
8. Apply Bayesian tracking in range and time
9. Extract the *final* estimates of the first-order boundaries

Next, we describe the null-finding algorithm steps in detail, giving the associated setup parameters, which are set in the `Header.txt` and `FirstOrderConfigs.plist` files in the setup files folder. While this chapter focuses on describing Steps 2 through Step 7, the Bayesian estimation tracking in range and time performed on Step 8 will be discussed in Chapters 6 and 7.

Finally, it is important to note that the radar site parameters mentioned above should be calibrated periodically. Not only the values of those parameters will differ for every radar site, but also they should be adjusted during the event of a strong storm or high tides being predicted. For example, the maximum current expected at a radar site (`currmax` in cm/s) might be increased for a period of time if the site is predicted to experience an unusual meteorological incident.

4.2 Premise of the Nulls Between the First and the Second-Order Regions

One of the most critical steps in finding the boundaries that separate the first and the second-order Bragg regions is the empirical determination of the *nulls* (local minimums) between their Doppler shift frequencies, as illustrated in Figure 3.3.

Even the visual identification of the nulls between those regions in frequency can be challenging at times, either due to

- the spectrum becoming saturated with noise or strong waves
- difficult radar sites like the scenarios listed at the beginning of this chapter.

In consequence, the computational method used to find the nulls has to include solutions that succeed when the identification is particularly difficult.

The following **three fundamental features that distinguish the first and second-order regions of the backscatter spectrum** are validated in prior literature[1][3][4][27][40], and are the necessary *theoretical basis* when developing an algorithm to determine the nulls (local minimums) between their frequencies.

1. The second-order region peaks surround the first-order region on both sides.
2. Second-order spectrum is typically an order of magnitude below first-order, and the first-order peaks are typically two orders of magnitude above the second-order continuum.
3. Second-order frequencies are roughly symmetric about the first-order Bragg peak frequency.

4.2.1 Typical Backscatter Evaluation for a 13 MHz Radar

Next, we are going to evaluate an example of a typical radar backscatter signal from the ocean to confirm the three features distinguishing first- and second-order spectrum listed above.

The ocean backscatter cross spectra received by a SeaSonde HF radar has a characteristic appearance. As introduced in Section 3.2, the dominant first-order peaks occur due to scatter from the ocean waves moving directly toward or away from the radar with wavelength equal to one half the radar wavelength and thus satisfying the Bragg resonant condition.

– If there were no ocean surface currents present, the first-order peaks would be narrow – impulse (delta) functions – located at the Doppler shift frequency f_1 corresponding to the phase speed of the resonant waves.

– If a surface current is present, the Bragg resonant echo is shifted from the zero-current Bragg position f_1 by an amount Δf proportional to the radial component of the current velocity, and we find the displaced peaks at the Doppler frequency f_{peak}, as the example below shows.

Due to varying ocean currents present in the radar scatter area, as shown in Figure 2.9, there is typically a different Doppler shift for each azimuth angle of return, causing the Doppler spectrum to be spread out in frequency and therefore the first-order region is broadened and not a delta function. More details can be found in Section 3.2.1.

A typical radar backscatter signal from the ocean, as the one presented in Figure 4.1, exhibits two relatively stronger (first-order) regions in its spectra:

- Left first-order region, located in the negative Doppler shift frequency region, that corresponds to the receding Bragg ocean wave (moving away from the radar).
- Right first-order region, located in the positive Doppler shift frequency region, that corresponds to the advancing Bragg ocean wave (moving towards the radar).

Similar to the example presented in Section 3.1.3 for a 25 MHz radar, we show next some calculations for a 13 MHz radar like the one in Figure 4.1.

$$f_{radar}\text{=13 MHz, } c\text{=3x10}^8 \ m/s, \ g\text{=9.8 } m/s^2$$

$$\lambda_{radar} = \frac{c}{f_{radar}}$$

$$\lambda_{radar} = \frac{3 \times 10^8}{13 \times 10^6} = 23 \text{ m}$$

Thus, the first-order Bragg wavelength λ_{Bragg} will be L_1 as we saw in Eq.(3.1) and with the order n=1.

$$L_1 = \lambda_{radar}/2 = 11.5 \text{ m}$$

Continuing with Equations (3.1), (3.2), (3.4) and (3.5), we obtain the velocity of the Bragg wave – `Doppler_Velocity`:

$$V_1 = \sqrt{g\frac{L_1}{2\pi}} = \sqrt{g\frac{11.5}{2\pi}} = 4.192 \text{ m/s}$$

Therefore, we show below the theoretical frequency shift location of the first-order Bragg wave for this 13 MHz radar site if there was zero ocean current and when the radar is at a far enough distance so that $\theta = 90$ degrees, thus $\sin\theta = 1$.

$$f_1 = \pm\frac{2(V_1 \sin\theta)}{\lambda_{radar}} = \frac{\pm V_1 \sin\theta}{L_1}$$

$$f_1 = \frac{\pm 4.192 \cdot \sin(90^\circ)}{11.5} = \pm 0.3645 \text{ Hz}$$

We find that the peak of the advancing Bragg wave in Figure 4.1 is located around +0.34 Hz, which corresponds to the following Δf displacement from the theoretical Bragg wave:

$$\Delta f = f_1 - f_{\text{peak}} = 0.3645 - 0.34 = 0.0245\text{Hz}$$

Therefore, as we described in Section 3.2.1, this frequency shift is attributed to the radial component of the surface current relative to the radar. Since the displacement of the first-order Bragg peak (located at f_{peak}) is $\Delta f = 0.0245$ Hz we find the corresponding radial ocean current through the calculation below.

$$V_{\text{peak}} = \frac{f_{\text{peak}} L_1}{\sin\theta} = \frac{0.34 \cdot 11.5}{1} = 3.91 \text{ m/s}$$

$$\texttt{Velocity\_Current} = V_1 - V_{\text{peak}} = 28.2 \text{ cm/s}$$

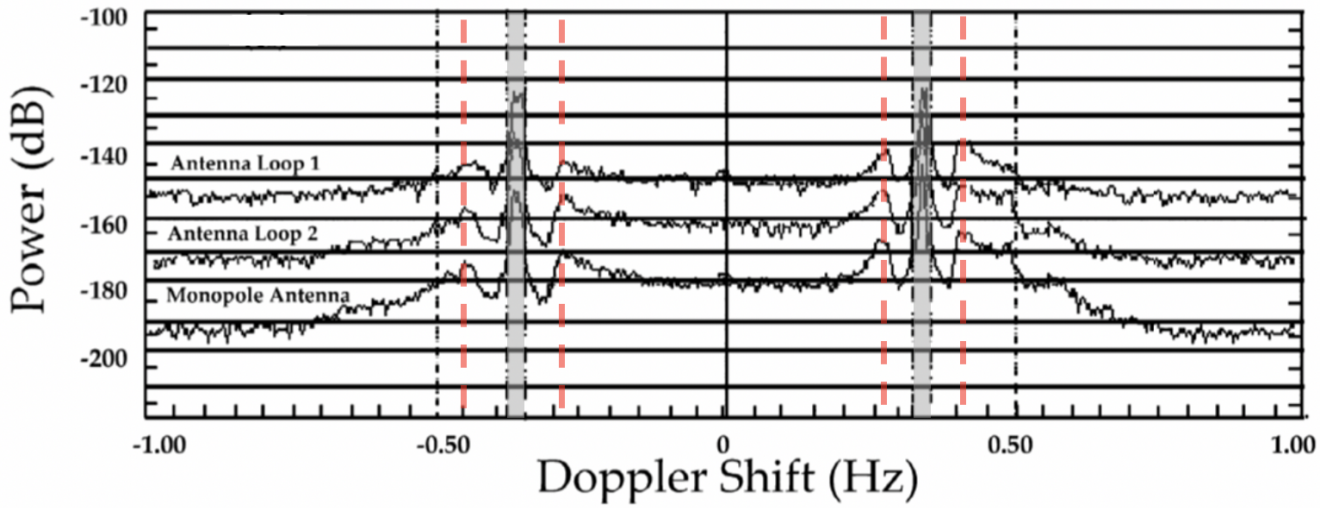

Figure 4.1: Backscatter spectra measured by the standard-range SeaSonde 13 MHz radar at Bodega Marine Laboratory on November 30, 2001 at 4 a.m.[44] As labeled, the upper two curves are from the loop antennas while the lower curve corresponds to the Antenna 3 (the monopole antenna). The curves have been displaced vertically by 20 dB to allow ease of viewing. The dashed-black vertical lines surrounding the dominant Bragg peaks indicate the boundaries that separate the first- and second-order structure. The first-order is then shaded in grey, and the red dashed lines are located at the peaks of the second-order regions. The dotted-black vertical lines simply mark the ± 0.5Hz Doppler shift.

The first-order peaks are typically two orders of magnitude higher than the surrounding continuum, from which they are separated by well-defined *nulls*. The continuum emerges from higher-order scatter, the greater part of which arises from scatter from pairs of ocean waves[1][8]. Ocean wave information is obtained from interpretation of this higher order spectrum, normalized by the first-order energy[44][77].

For the 13 MHz example in Figure 4.1 in particular: the peak of the left first-order region is found around the Doppler shift frequency $f_L = -0.34$ Hz, while the peak of the right first-order region is located around the Doppler shift frequency $f_R = +0.34$ Hz. Each of these two first-order regions – shaded in grey in Figure 4.1 – are surrounded by higher order continuum regions (about 20 dB below the first-order peaks located at $f_{\text{peak}} = \pm 0.34$ Hz), that correspond indeed to what we know as second-order.

As detailed in Section 3.2.2 and presented in Figure 3.8, the second-order spectrum is predominantly caused by the interaction of the radar electromagnetic wave with pairs of ocean waves present on the ocean surface. The maximums of the four second-order regions in Figure 4.1 are marked with red dashed lines. The two second-order regions on the left half are surrounding the left first-order Bragg – the receding Bragg ocean wave (moving away from the radar). On the other hand, the other two second-order regions surround the right first-order Bragg that corresponds to the advancing Bragg ocean wave (traveling towards the radar). In addition, we note that both the first and second order echoes are well above the noise level indicated by the power level at the wings of the spectrum near ± 1 Hz on Figure 4.1.

Summarizing, the presence of second-order regions surrounding the first-order on both sides satisfy "Feature 1" on our list of fundamental features that distinguish the first and second-order regions of the backscatter spectrum. We observe how the second-order spectrum energy is indeed at least an order of magnitude below the first-order peaks (specifically around 15 to 20 dB below in the Figure 4.1 example) and the second-order peaks are roughly symmetric to the first-order Bragg frequencies (about ± 0.1 Hz in this example). Therefore, the characteristics of the second-order regions do satisfy "Feature 2" and "Feature 3" on our list too.

4.3 Backscatter Spectra of the Monopole Antenna

For the specific cross-spectra file that we are working with, namely `css_file`, we extract the self spectra of the monopole antenna (SeaSonde's Antenna 3), stored as `Antenna3Self`.

Figure 4.2 shows the backscatter spectra of the monopole antenna: its relative signal magnitude in voltage-squared as extracted from the cross-spectra file. In addition and for clarity, we present its equivalent in decibels in Figure 4.3, converted following the calculation in Algorithm 1. Note that the decibels unit dBm express the absolute power level relative to 1 mW (milliwatt).

In Algorithm 1 we find a `-40 dB` factor which corresponds to the conversion loss in the receiver, and a `+5.8 dB` factor which responds to the processing computational gain[46].

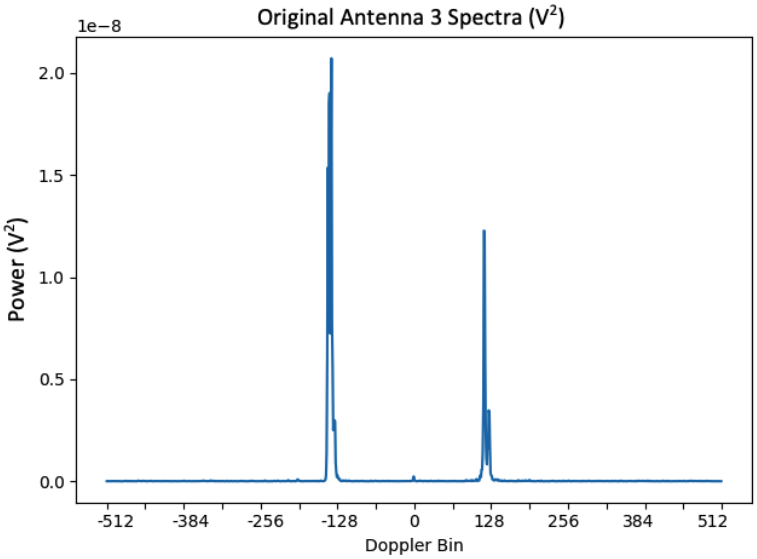

Figure 4.2: Backscatter spectra example from the monopole antenna (SeaSonde Antenna 3). The horizontal axis has been chosen to be Doppler bins, the total `fftlen` is 1024 Doppler bins. The vertical axis corresponds to its relative signal magnitude in voltage-squared as extracted from the cross-spectra file.

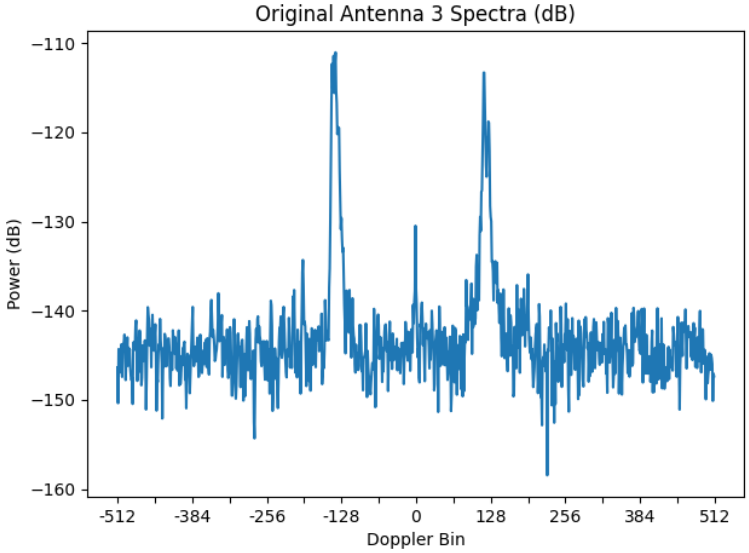

Figure 4.3: Backscatter spectra example from the monopole antenna (SeaSonde Antenna 3). The horizontal axis has been chosen to be Doppler bins, the total `fftlen` is 1024 Doppler bins. As opposed to Figure 4.2, the vertical axis on this figure corresponds to the equivalent signal amplitude in dBm received at the monopole antenna, which provides more clarity about the changes in amplitude.

Algorithm 1 Conversion Self Spectra in Volts to Decibels (dB)

To convert self spectra `Antenna3Self` in Volts to decibels:

```
Antenna3Self_dBm = 10 * log10(abs(Antenna3Self)) - (-40 + 5.8)
```

4.4 Apply a Noise Threshold to the Spectra

First, before searching for the first-order region boundaries in the monopole antenna spectra, we apply a threshold to eliminate data that is considered too noisy. To do so, we need the noise factor `noisefact` extracted from `Header.txt`. For each range in the spectra, and as described in Algorithm 2, we first find the corresponding noise baseline `noise_baseline` at that range. The noise baseline is based on an estimation of the noise in the wings of the spectrum (near ± 1 Hz in Figure 4.1), specifically the median value on the backscatter energy in the wings. Later on, we multiply the baseline of that range by the noise factor to obtain what we call `noise_threshold`. Finally, we assign the threshold value `noise_threshold` in Volts to all data points in that range that fall below that threshold.

Algorithm 2 Applying a Noise Threshold

Load the monopole antenna spectrum `Antenna3Self`
for each RANGE **do**
  `left_wing` = the left wing portion (about 1/6 of the total Doppler spectra)
  `noise_baseline = findSNR(side_wing)`
  `noise_threshold = (noisefact` $\times$ `noise_baseline)`
  `indices_below_threshold = Antenna3Self < noise_threshold`
  `Antenna3[indices_below_threshold] = noisefact` $\times$ `noise_baseline`
end for

Figure 4.4 shows the backscatter spectra of the monopole antenna in dB after applying the noise threshold with `noisefact = 3.98` extracted from the Header file. We observe how the original spectra in dB shown in Figure 4.3 has indeed all its values falling below `noise_threshold = (noisefact` $\times$ `noise_baseline)` filtered out.

Algorithm 3 Smoothing the Monopole Antenna Spectra

Load the cross-spectra array for the current range `css_array`[RANGE]
Extract the `nsm` factor for the chosen radar site from the `Header.txt`
`f_smooth`[RANGE] is the smoothed cross-spectra array for the current range
for each RANGE **do**
1) Create a python *pandas* dataframe
`f_smooth`[RANGE] = `pd.Series(css_array`[RANGE])
2) Apply the rolling mean function with a smoothing factor of `nsm` points
`f_smooth`[RANGE]`.rolling(nsm, center=True).mean()`
3) Fill the NaN values on the wings (appearing due to smoothing `nsm` points at a time) with the minimum value of the array i.e. the noise floor value
`f_smooth`[RANGE]`.fillna(min(css_array`[RANGE]`))`
end for
Return the smoothed cross-spectra array `f_smooth`

4.5 Smooth the Spectrum

After applying the noise threshold and before determining the null positions, the spectrum is smoothed through a rolling mean. This smoothing will remove non-relevant nulls which consequently will be avoided in the determination, and thus produce a more stable estimate of the first-order boundary locations.

A rolling mean is simply the mean of a certain number of previous samples in a time series. That number of previous periods is determined by the `nsm` factor extracted from the `Header.txt`. In Python, to calculate the rolling mean of a *pandas* data structure, we simply use the `a_array.rolling(nsm).mean()` function, as detailed in Algorithm 3.

Once the spectra has been thresheld by `noise_fact` $\times$ `noise_level`, we apply the running mean described in Algorithm 3: it performs a smearing over `nsm` points, applies a shift to center the spectra back to 0 Doppler frequency, and fills any empty bins on the wings with the minimum value of the array. The resulting smoothed spectra after the noise threshold seen in Figure 4.4 is presented in Figure 4.5. We observe how some of the sidelobes – specially around the right Bragg region in Figure 4.4 – have indeed been smoothed in Figure 4.5, which definitely aids in the null-detection steps described in Section 4.8.

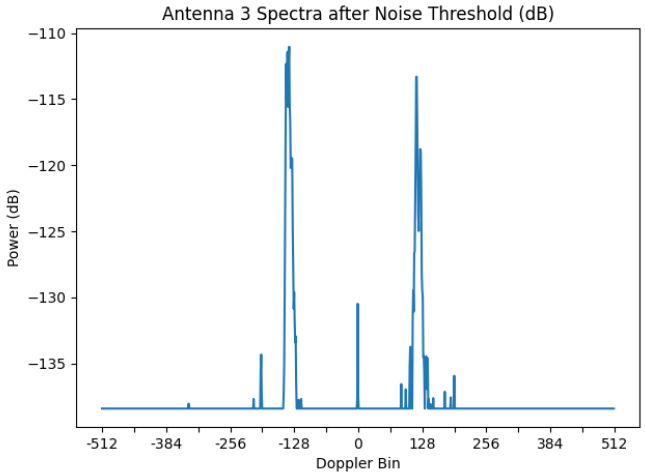

Figure 4.4: Backscatter spectra example from the monopole antenna (SeaSonde Antenna 3) after applying the noise threshold to the original power signal in Figure 4.3. The horizontal axis is Doppler bins (total `fftlen` is 1024 bins). The vertical axis is the equivalent signal amplitude in dBm received at the antenna, and as opposed to Figure 4.3 we see how the noise threshold has been applied and the minimum power is now homogeneous.

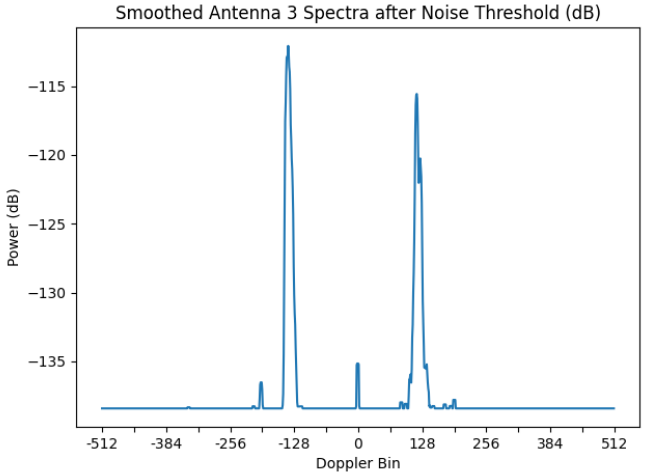

Figure 4.5: Backscatter spectra example from the monopole antenna (SeaSonde Antenna 3) after applying the noise threshold and corresponding smoothing to the original power signal in Figure 4.3. The horizontal axis is Doppler bins (total `fftlen` is 1024 bins). The vertical axis is the equivalent signal amplitude in dBm received at the antenna, and as opposed to Figure 4.3 and Figure 4.4 we see how the smoothed signal presents less local minima (*nulls*).

4.6 Normalize the Spectrum

Besides the original and the smoothed versions of the backscatter spectra, we generate a "normalized" version as well. As described in Algorithm 4, given the range-dependent maximum power extracted from the Bragg peaks at each range, we normalize the spectra into a proportional function spanning from 0 (the bins where the power is at a minimum) to 1 (the bins where the power is at a maximum).

Algorithm 4 Normalizing the Monopole Antenna Spectrum

Load the smoothed cross-spectra array `f_smooth`
for each RANGE **do**
  1) Find `a_max`, the maximum peak value for the current range
     `a_max = max(f_smooth`[RANGE])
  2) Find `a_min`, the current range minimum value (the noise floor value)
     `a_min = min(f_smooth`[RANGE])
  3) Create the normalized spectra array for current range
     `f_normalized`[RANGE]`= [(i - a_min) / (a_max - a_min)]`
     `for` each bin value `"i"` in `f_smooth`[RANGE]
end for
Return the normalized cross-spectra array `f_normalized`

The process of normalization helps the distinction between Bragg and non-Bragg energy at higher range cells where the overall signal is weaker, and therefore leads to a more consistent null detection between ranges because (as opposed to the *conventional* method) we apply the power thresholds described in Section 4.7, that set limiting boundaries for valid current retrieval, to the normalized spectra.

In addition we also use the factor `freject` to discard far ranges where the signal strength is negligible because the overall ratio between maximum power and minimum power in dB is below the `freject` requirement.

4.7 Limit the Spectral Range on Both Sides of the Peak

The first-order region is separated from the second-order by some well defined energy minimums, which are known as frequency *nulls*. As we describe in Section 4.8, we want to avoid searching for the nulls within the region that we know is the stronger first-order, since this region has its own local minimums, as shown in Figure 4.1. In other words, the null detection should only consider the Doppler frequencies on the periphery of the first-order region, and using the methods described in this section we will set a series of limits that restrict the area where first- and second-order boundaries should fall and avoids searching for nulls outside of those limits.

On the first hand, there are two sets of limits that we denominate "Power Threshold Limits":

– The *higher* energy set depends on the factor `fdown` and are closer to the Bragg region peaks, therefore stronger energy; they protect the higher Bragg energy that we definitely want to classify as first-order spectra.

– The *lower* energy set of "Power Threshold Limits" depend on the factor `flim` and are further in Doppler frequency, closer to the second-order region. In fact, this set of limits define the minimum ratio of power respect to the peak that any Doppler bins is required to have to be considered a bin within the first-order region.

On the other hand, there is a third set of limits that are generally further in Doppler frequency than the ones set by the `flim` parameter.

These *outer* limits are the higher Doppler frequency set of limits associated with the maximum ocean current estimated possible within the geographical location of the radar site, which we denominate "Current Velocity Limits".

Next we describe the algorithms necessary for calculating both the Current Velocity Limits and the Power Threshold Limits.

4.7.1 Velocity of the Bragg Wave in m/s

For those algorithms we first need to calculate the wavelength of the first-order Bragg wave in Algorithm 5 and its velocity in Algorithm 6. These two algorithms follow the ocean backscatter descriptions in Section 3.1.2, and specifically Equations (3.1), (3.2), (3.4) and (3.5).

Algorithm 5 Calculate the Wavelength of the Bragg Wave

· Load `Radar_Wavelength` as λ
· The Bragg waves wavelengths – Eq.(3.1) – are multiples of λ:
$L_n = n\lambda/2$
· The wavelength of the first-order Bragg wave is then:
$\lambda_{\text{Bragg}} = \lambda/2$

Algorithm 6 Calculate the Velocity of the Bragg Wave

· Load the Bragg wave wavelength (Algorithm 5) as λ_{Bragg}
· Load the gravitational acceleration $g = 9.8$ m/s^2
· The velocity of the Bragg wave – Eq.(3.2) – is then:
`Velocity_BraggWave` $= \sqrt{g \cdot \lambda_{\text{Bragg}}/(2\pi)}$

4.7.2 Doppler Frequency and Doppler Bin Locations

Like we mentioned in Chapter 3, and showed in Figures 3.4, 3.5, 3.6 and 3.7, the x-axis of the radar backscatter spectra can be one of the following options:

- Doppler Frequency: the frequency shift of the ocean echo signals scattered from the moving target. In our case, the moving targets are the Bragg waves. (Refer to Section 3.1.2 for more information about the Bragg scattering of the ocean waves) Figure 3.4 shows how the 0 Doppler frequency corresponds to DC, where there is no frequency shift.

- Doppler Bins: the number of bins that we can separate our Doppler frequency into, depending on our frequency resolution. The horizontal axis can either span between 0 and `fftlen`, or most usually be centered at 0 and have `fftlen/2` negative Doppler bins on the left side (receding Bragg wave) and `fftlen/2` positive Doppler bins on the right side (advancing Bragg wave) as seen in Figure 3.5.

- Doppler Velocity: the equivalent velocity of the Doppler shift in m/s. The measured Doppler shift is proportional to the sum of the known velocity of the Bragg ocean wave and the unknown velocity of the surface current. As we see in Figure 3.6, the `Velocity_Doppler = 0` corresponds indeed to DC.

- Doppler Current Velocity: the equivalent velocity of the ocean currents underlying below the ocean waves. As we studied in Section 3.2.1 the two `Velocity_Current = 0` correspond to the theoretical Bragg wave frequencies f_0 like calculated in the example above for a 13 MHz radar.

The currents moving along in the same direction as the Bragg wave will appear to the right of that Bragg peak (positive current velocity), while the currents moving opposite to the Bragg wave will appear on the left side of the corresponding Bragg peak (negative current velocity).

The calculations performed in Algorithms 7 through 11 will be used in many instances of our code. For example to easily convert between the four aforementioned units depending on the requirement or application e.g. when we are given a maximum current velocity `currmax` in cm/s but we need to find its Doppler bin location in the corresponding spectra.

Algorithm 7 Transform between Doppler Velocity and Current Velocity

1) Load the velocity of the Bragg wave as `Velocity_BraggWave` (Algorithm 6)

2) To obtain a `Doppler_Velocity`:
· Load the current velocity we want to convert as `Velocity_Current`
if Right Bragg wave (Advancing):
`Doppler_Velocity` = `Velocity_Current` + `Velocity_BraggWave`
else if Left Bragg wave (Receding):
`Doppler_Velocity` = `Velocity_Current` – `Velocity_BraggWave`
end if

3) To obtain a `Velocity_Current`:
· Load the Doppler velocity we want to convert as `Doppler_Velocity`
if Right Bragg wave (Advancing):
`Velocity_Current` = `Doppler_Velocity` – `Velocity_BraggWave`
else if Left Bragg wave (Receding):
`Velocity_Current` = `Doppler_Velocity` + `Velocity_BraggWave`
end if

Algorithm 8 Convert Current Velocity (cm/s) to Doppler Frequency (Hz)

· Load the Bragg wave wavelength (Algorithm 5) as λ_{Bragg}
· Load the velocity of the Bragg wave as `Velocity_BraggWave` (Algorithm 6)
· Transform `Velocity_Current` into `Doppler_Velocity`: (Algorithm 7)
$\texttt{Doppler\_Velocity} = \texttt{Velocity\_Current} \pm \texttt{Velocity\_BraggWave}$
· Calculate the `Doppler_Frequency` as:
$\texttt{Doppler\_Frequency} = 2 \cdot \texttt{Doppler\_Velocity}/\lambda_{\text{Bragg}}$

Algorithm 9 Convert between Doppler Bin and Doppler Frequency

From the `Header.txt` file:
· Load `fftlen`, the amount of Doppler frequency resolution bins
· Load `sweep_period` in seconds as `T`.
· Calculate the `Doppler_Frequency` as:
$\texttt{Doppler\_Frequency} = \texttt{Doppler\_Bin}/(\texttt{T} \cdot \texttt{fftlen}) - 1/(2\texttt{T})$
· Calculate the `Doppler_Bin` as:
$\texttt{Doppler\_Bin} = \texttt{T} \cdot \texttt{fftlen} \cdot (\texttt{Doppler\_Frequency} + 1/(2\texttt{T}))$

Algorithm 10 Convert Current Velocity to Doppler Bin

From the `Header.txt` file:
· Load `fftlen`, the amount of Doppler frequency resolution bins
· Load `sweep_period` in seconds as `T`.
· Transform `Velocity_Current` into `Doppler_Frequency` (Algorithm 8)
· Calculate the `Doppler_Bin` location:
$\texttt{Doppler\_Bin} = \texttt{T} \cdot \texttt{fftlen} \cdot (\texttt{Doppler\_Frequency} + 1/(2\texttt{T}))$

Algorithm 11 Convert Doppler Bin to Current Velocity

· Load `Radar_Wavelength`
· Load the gravitational acceleration $g = 9.8$ m/s^2
· Transform `Doppler_Bin` into `Doppler_Frequency`
and then `Doppler_Frequency` into `Velocity_Current`
by following Algorithms 8 and 9

4.7.3 Current Velocity Limits

To find the locations of each of the four **current velocity limits**, we consider the absolute maximum ocean current velocity expected at the site given by the parameter `currmax` (in absolute cm/s) from the `Header.txt` file. Then, as described in Algorithm 12 we calculate the Doppler frequencies (or equivalent Doppler bin locations) of the positive maximum current allowed on the site `+freq_MaxCurrent` and its negative `-freq_MaxCurrent` for each of the two halves of the spectra. The location of these maximum current boundaries are independent in range since they are based on the factor `currmax` and radar parameters such as wavelength and sweep rate, and thus the current velocity limits will only be calculated once.

Figure 4.6 presents an example of the 19th range cell (or bin) of the original monopole antenna spectra in dB shown in Figure 4.3. For clarity, the x-axis in Figure 4.6 and Figure 4.7 are set to current velocity in cm/s. Assuming `currmax=150 cm/s` for this site, the four magenta vertical lines in the figure represent the **current velocity limits** that would be placed due to the maximum current allowed in the site. If comparing with Figure 4.3, the maximum current limits would correspond to the Doppler bins ± 87 and ± 168 approximately.

Algorithm 12 Limiting the Spectrum - Current Velocity Limits

- Load `currmax`: the maximum expected current in the site
- To determine the **current velocity limits**:
  Find the Doppler bins locations that correspond to the maximum current (in cm/s) allowed in the site (Algorithm 10)
  ```
  +freq_MaxCurrent = Doppler_Bin(+currmax)
  -freq_MaxCurrent = Doppler_Bin(-currmax)
  ```
- Apply those same limits in frequency to all the ranges available in the cross-spectra, shown as magenta vertical lines in Figure 4.6.

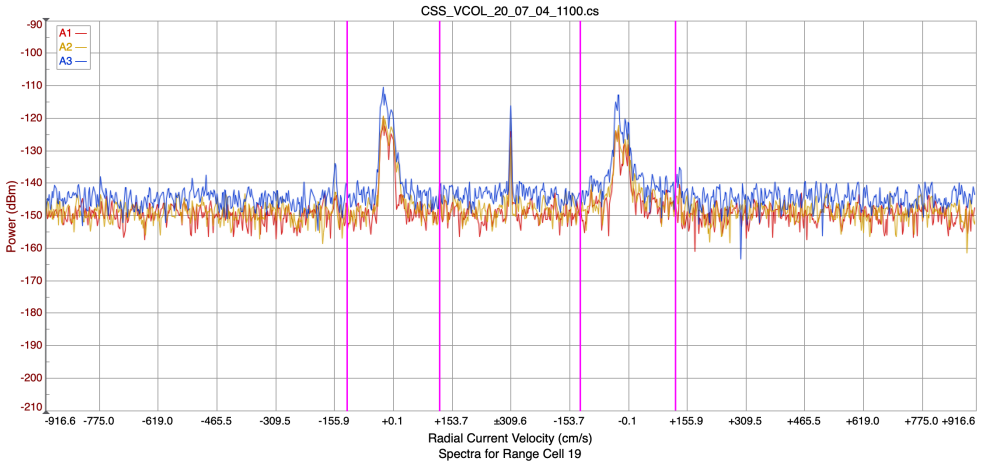

Figure 4.6: Backscatter spectra from the monopole antenna (SeaSonde Antenna 3) at the VCOL radar site in British Columbia. The first-order boundaries have to be found within the "Current Velocity Limits" corresponding to the `currmax` factor in the Header and marked in the figure by four magenta vertical lines.

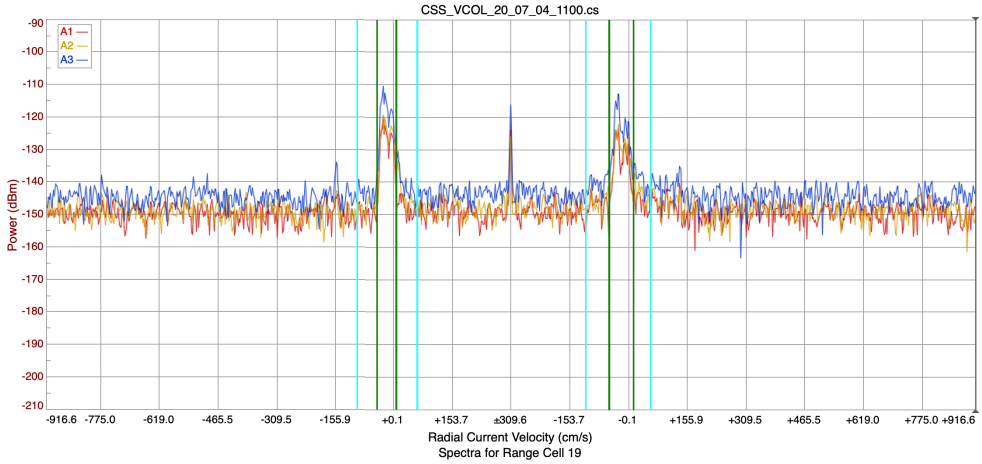

Figure 4.7: Backscatter spectra from the monopole antenna (SeaSonde Antenna 3) at the VCOL radar site. The first-order boundaries have to be found within the higher and lower "Power Threshold Limits" corresponding to the factors `fdown` and `flim` respectively. The higher "Power Threshold Limits" are shown as four green vertical lines, while the lower ones are presented in light blue.

4.7.4 Power Threshold Limits

Higher Power Limits: To find the higher power limits for each of the backscatter spectra halves, we first find the points with the maximum power `a_max_L` and `a_max_R` as described in Algorithm 13. Next we locate the surrounding points with voltage-squared equal to `a_max/fdown`, like in Algorithm 14. The set of higher power limits are shown as four green vertical lines in Figure 4.7. The factor `fdown` in the Header has to be chosen carefully. If it is set too small then we could erroneously select a null around a strong current as the boundary between first and second order and therefore all the ocean currents below that one would be incorrectly discarded. On the other hand, if `fdown` is chosen too large, the nulls between first and second-order may be missed altogether, and second-order data included erroneously as ocean currents resulting in wild radial vectors.

Lower Power Limits: The lower power limits for each of the backscatter spectra halves are set to eliminate spectral points that are too far below the peak energy `a_max`. To set the lower power limits we use the factor `flim` in the Header, and the limits are set at the points where the voltage-squared is below `a_max/flim`, as described in Algorithm 15. The set of lower power limits are shown as four light blue vertical lines in Figure 4.7.

Algorithm 13 Set the Values of the Bragg Peaks

· Load the cross-spectra array `css_array`
· Load `midfft` – the Doppler bin equivalent to 0 Hz Doppler frequency
for each RANGE **do**
  1) Find `a_max_L`, the maximum left peak value for the current range
    `a_max_L = max(css_array`[RANGE, `0:midfft`])
  2) Find `a_max_R`, the maximum right peak value for the current range
    `a_max_R = max(css_array`[RANGE, `midfft::`])
end for

Algorithm 14 Limiting the Spectrum - Higher Power Limits

· Load the cross-spectra array `css_array`
· Load `midfft` – the Doppler bin equivalent to 0 Hz Doppler frequency
· Load `fdown` – the factor down in power to protect first-order energy
for each RANGE **do**
1) Load `a_max_L` and `a_max_R`,the Bragg peaks for the current range
2) Find the power threshold limits relative to `fdown`
± `freq_PowerHi_L`[RANGE] = ± `Doppler_Bin(a_max_L/fdown)`
± `freq_PowerHi_R`[RANGE] = ± `Doppler_Bin(a_max_R/fdown)`
end for

Algorithm 15 Limiting the Spectrum - Lower Power Limits

· Load the cross-spectra array `css_array`
· Load `midfft` – the Doppler bin equivalent to 0 Hz Doppler frequency
· Load `flim` - factor setting the minimum power requisite for first-order
for each RANGE **do**
1) Load `a_max_L` and `a_max_R`,the Bragg peaks for the current range
2) Find the power threshold limits relative to `flim`
± `freq_PowerLo_L`[RANGE] = ± `Doppler_Bin(a_max_L/flim)`
± `freq_PowerLo_R`[RANGE] = ± `Doppler_Bin(a_max_R/flim)`
end for

4.8 Locating the Possible Nulls

After the steps to generate our Current Velocity Limits and Power Threshold Limits described in Section 4.7, we are able to restrict the area where to look for the boundaries between first and second order:

- The upper threshold at every range will be set by the respective "Higher Power Limits", ±`freq_PowerHi_L` and ±`freq_PowerHi_R`.
- At each range, we find the lower power limits for each of the backscatter spectra halves, ±`freq_PowerLo_L` and ±`freq_PowerLo_R`. These are set to eliminate spectral points that are too far below the peak energy.
- On the other hand, `+freq_MaxCurrent` and `-freq_MaxCurrent` are the current velocity limits that focus on eliminating all of those spectral points that correspond to higher ocean current velocities that are not considered realistically possible and they are equal across all ranges.
- Finally, for every range the lower threshold will be set as to the highest of limits between the "Current Velocity Limits" and the "Lower Power Limits".

Following the examples set in Figure 4.6 and Figure 4.7, the area where we will be searching for local minimums (*nulls*) will be the non-shaded regions of frequency shown in Figure 4.8. Therefore, we will search for *nulls* in those four regions as it is described in Algorithm 16 and later assign those nulls as the first-order region boundaries from where the ocean current information will be derived and the radial ocean current vectors determined (Chapter 6).

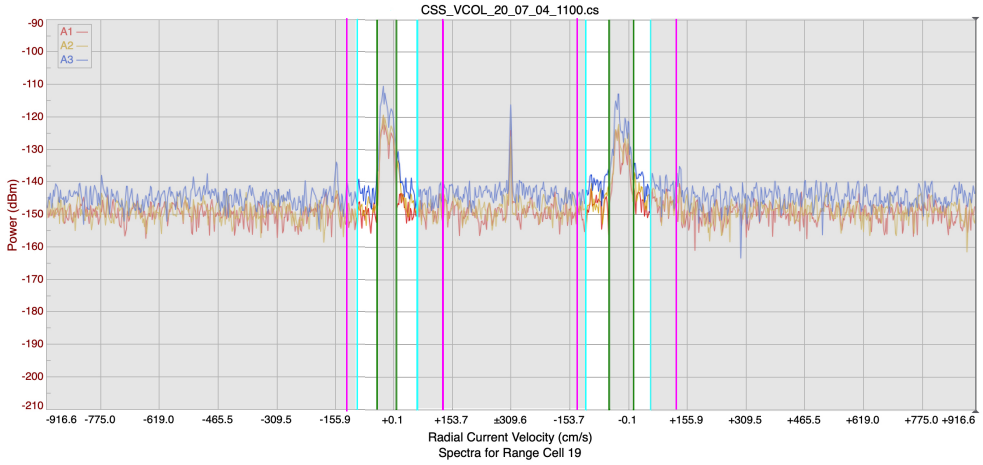

Figure 4.8: Backscatter spectra from the monopole antenna (SeaSonde Antenna 3) at the VCOL radar site. The first-order boundaries have to be found within the non-shaded regions. The lower energy thresholds correspond to the "Lower Power Threshold Limits" because they are more restrictive than the "Current Velocity Limits" in this instance.

Algorithm 16 Finding the Nulls within the Frequency Thresholds

For each RANGE, `data` is every array from every REGION within our *limits*.
for each RANGE **do**
  for each REGION between limits **do**
    – Apply the Python function `scipy.signal.argrelmin(data)`
    – `argrelmin(data)` calculates the relative minima of `data` and extracts the locations of the nulls.
  end for
end for
If no local minima are found in the REGION, then the output boundary is set as the "Higher Power Limit" location, which is the most restrictive.

4.9 Search for Split Bragg Regions

Even though the strongest *nulls* (local minima within our set of limits) are found using Algorithm 16, we designed an extra step to further improve the determination of the first-order boundaries that is not available in the *conventional method*. We apply a recursive search for split Bragg regions –as they were presented in Section 3.3.2– in every range bin (cell) of our radar backscatter spectra.

The full Recursive Split Bragg Search as described in Algorithm 17:

- We start by defining a Doppler window for each half of the spectra.
  Each window starts as the interval of Doppler bins from the negative current velocity limit to the positive current velocity limit, shown as magenta vertical lines in Figures 4.6 and 4.8.

- Window L1, surrounding the receding (negative) Bragg wave, is placed in the left half of the spectra, (See Figure 4.9) $f <$ `fftlen/2`, between [$-$`freqMaxCurrent` : $+$`freqMaxCurrent`].

- Window R1, surrounding the advancing (positive) Bragg wave, is placed in the right half of the spectra, (See Figure 4.9) $f >$ `fftlen/2`, between [$-$`freqMaxCurrent` : $+$`freqMaxCurrent`].

- Apply Algorithm 13 to find the Bragg peaks `a_max_L` and `a_max_R` in each of the Doppler windows – Doppler bins within a window where the voltage hits a maximum. These are labeled as `Vpeak_L1` and `Vpeak_R1` in Figure 4.9 to match the windows labels.

- Apply Algorithm 14 to find the Doppler bin locations going outwards in frequency from the Bragg peaks where their voltage first crosses the threshold `a_max/fdown` in the normalized spectra, using the same scale factoring as when normalizing the Antenna 3 spectra in Algorithm 4.

- Apply Algorithm 15 to find the Doppler bin locations going outwards in frequency from the Bragg peaks at where the voltage first crosses the threshold `a_max/flim` in the normalized spectra as well, and label them as `Vmin`.

- `Vmin` defines the minimum power allowed for that bin to be considered first-order, and therefore we can reject all the frequencies below the frequency where `Vmin` is found and below `freqMaxCurrent` as well, depending on the range either of them could be the most restrictive frequency.

- We search for any frequencies spanning between `Vmin` frequency on each side of each Bragg up until the `freqMaxCurrent` further on that side where their voltage-squared value is above `Vmin` too.

- If `Vmin` is found to be further in frequency than `freqMaxCurrent`, then the `Vmin` frequency will be replaced by `freqMaxCurrent`.

- Apply Algorithm 16 to find the local minimums, *nulls*, within Window L1 and Window R1 and label them as FOL:

  - `FOL_L1_L` and `FOL_L1_R` are the first-order boundaries found in L1
  - `FOL_R1_L` and `FOL_R1_R` are the first-order boundaries found in R1
  - See the four primary first-order Bragg boundaries in Figure 4.9

- **"Split Bragg test"**: We search for secondary Bragg peaks where their voltage-squared value is also above `amax/minSplitBragg`. In addition we want at least `minBinsBragg` number of Doppler bins to be above the `Vmin` threshold. Note how `minSplitBragg` and `minBinsBragg` are parameters that we collected from `FirstOrderConfigs.plist` (Section 4.1).

- Each time that the spectra portion passes the **"Split Bragg test"** we create a new window that spans between `Vmin` frequency on that side of the Bragg up until the corresponding `freqMaxCurrent`.

- See the new windows in Figure 4.9:
  - Window L2L is created on the left side of window L1, while Window R2R is found at the right side of window R2.
  - No additional windows are shown in Figure 4.9 because the frequencies between `Vmin` and `freqMaxCurrent` do not pass the **"Split Bragg test"** and thus no further windows are generated.

- Find the secondary Bragg peaks by applying Algorithm 13 into the new windows L2L and R2R.
  - `Vpeak_L2L` is the secondary Bragg peak found in Window L2L
  - `Vpeak_R2R` is the secondary Bragg peak found in Window R2R

- Repeat the null finding Algorithm 16 in the new windows L2L and R2R to find secondary first-order boundaries.
  - `FOL_L2L_L` and `FOL_L2L_R` are the secondary first-order boundaries found around `Vpeak_L2L`

  - `FOL_R2R_L` and `FOL_R2R_R` are the secondary first-order boundaries found around `Vpeak_R2R`
  - See the four secondary first-order Bragg boundaries in Figure 4.9

- Recursively continue trying to find more windows if the Bragg is split into more sections and thus more regions pass the **"Split Bragg test"**. For our example in Figure 4.9 there are no tertiary or further windows.

Algorithm 17 Recursive Split Bragg Search

· Apply Algorithm 14 and Algorithm 15 to find the threshold limits and set the initial Doppler windows.
· Apply Algorithm 13 to find the Bragg peaks `a_max_L` and `a_max_R` in each of the Doppler windows.
· Apply Algorithm 16 to find the local minimums, *nulls*, within the windows and around the Bragg peaks.
· Run the **"Split Bragg test"**:
  1) Search for secondary Bragg peaks where their voltage-squared value is above `amax/minSplitBragg`.
  2) We require at least `minBinsBragg` number of Doppler bins to be above the `Vmin` threshold.
· If a spectra portion within a window passes the **"Split Bragg test"** we create a new window with that portion.
· Recursively repeat the steps at each of the new windows and find the corresponding boundaries around the secondary Bragg region peaks. (See secondary Bragg regions in Figure 4.9)
· Repeat again if a tertiary Bragg region is present, etc.

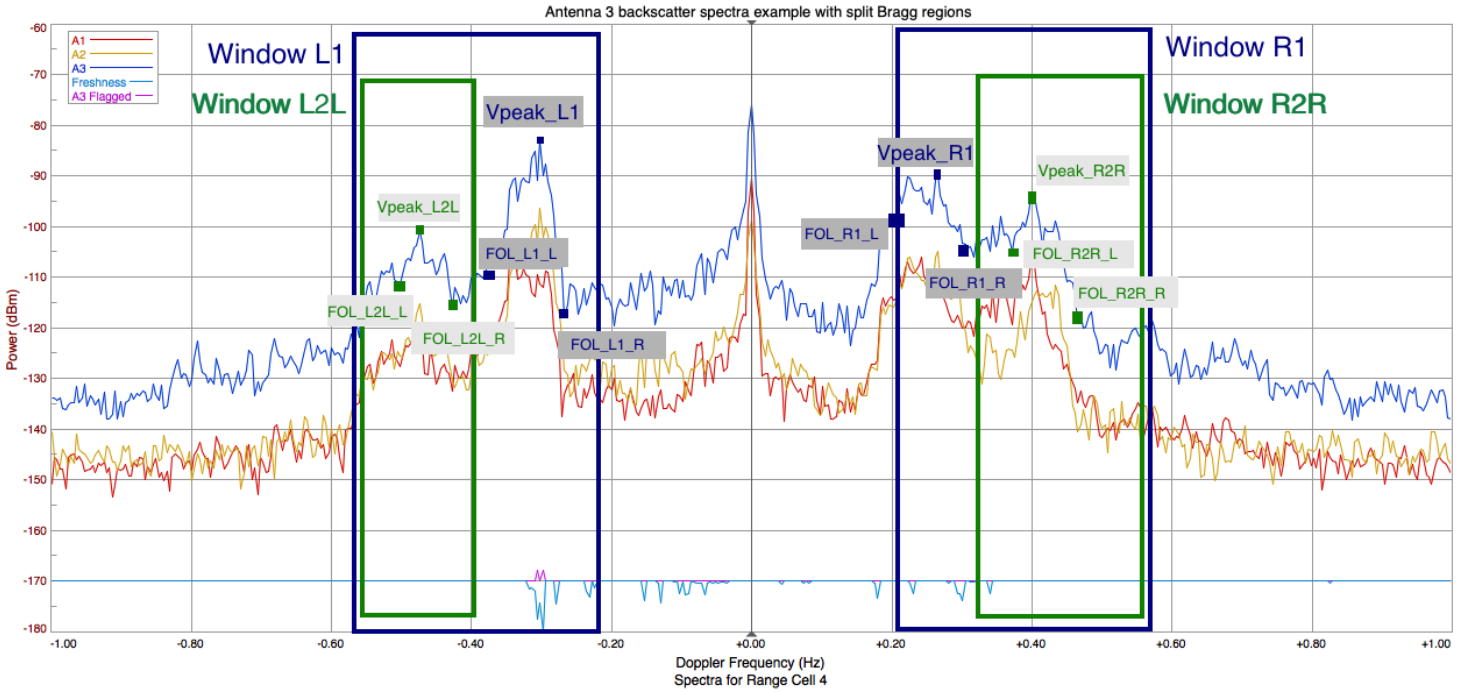

Figure 4.9: Backscatter spectra plot where the primary Doppler windows, the `Window L1` and `Window R1`, are shown as blue frames while the secondary (`Window L2L` and `Window R2R`) in green. In addition we observe the (receding Bragg wave) left peak `Vpeak_L1` and its two first order boundary nulls `FOL_L1_L` and `FOL_L1_R` as blue squared dots.
On the advancing Bragg wave we show the peak `Vpeak_R1` and its two first order boundary nulls `FOL_R1_L` and `FOL_R1_R`. Finally the secondary peaks `Vpeak_L2L` and `Vpeak_R2R` also have their own boundaries shown as green squared dots (`FOL_L2L_L` and `FOL_L2L_R`, and `FOL_R2R_L` and `FOL_R2R_R`)

4.10 Conventional Method Algorithm

In this section we present Algorithm 18, where we list which of the steps described in this chapter are actually executed in the *conventional* method to find the nulls in the spectra and therefore the first-order boundaries.

The *conventional* method steps include some of the algorithms we have presented in this section:

1. Smooth the spectra,
2. Find the Current Velocity Limits,
3. Find the Power Threshold Limits,
4. Search the nulls within those limits.

On the other hand, it differs from the method we have designed because:

- The *conventional* method applies its own calculations for the noise threshold to apply to the the spectra, by only averaging the first few Doppler bins on the wings of the spectra.
- The *conventional* method does not use a normalized spectra to calculate the power ratio thresholds.
- The *conventional* method does not include the recursive search for split Braggs.

Algorithm 18 Finding the Nulls using the Conventional Method

· Load the cross-spectra array `css_array`
· Place the calculated Doppler frequency *limits* in the `css_array`
· Discard all the regions that do not fall within those thresholds, shown shaded in Figure 4.8.
· Search for nulls in the frequencies where the first-order boundaries are allowed to be selected.

for each RANGE **do**
  for each REGION between limits **do**
    – Apply the Python function `scipy.signal.argrelmin(data)`
    – `argrelmin(data)` calculates the relative minima of `data` and extracts the locations of the nulls.
  end for
end for

For each RANGE, `data` is every array from every REGION within our *limits*.

The complete *alternative* multi-probabilistic null-finding algorithm we designed includes all the phases described in this chapter and its steps are summarized in the next section.

4.11 Multi-Probabilistic Null-Finding Algorithm

Lastly, the complete *alternative* multi-probabilistic null-finding algorithm we designed to determine the first-order boundaries includes all the aforementioned steps and its script is transcribed in Algorithm 19.

The Bayesian tracking mentioned in Step 10 of the complete Algorithm 19 is detailed in Chapter 5, while the steps that describe its application to observational radar data can be found in Chapters 6 and 7.

Algorithm 19 Complete Multi-Probabilistic Null-Finding Algorithm to Determine the First-Order Boundaries

1. Load the cross-spectra array `css_array`
2. Apply noise threshold
3. Calculate the normalized version of the spectra
4. Find the locations of the current velocity limits and the power threshold limits
5. Place all our *limits* in the `css_array`: the Doppler frequency thresholds as described in Algorithms 12, 14 and 15.
6. Discard all the regions that do not fall within those thresholds, shown shaded in Figure 4.8.
7. Apply the **Recursive Split Bragg Search** (Described in Section 4.9)
8. Search for nulls in the frequencies where the first-order boundaries are allowed to be selected.
   for each RANGE **do**
   for each REGION between limits **do**
   – Apply the Python function `scipy.signal.argrelmin(data)`
   – `argrelmin(data)` calculates the relative minima of `data` and extracts the locations of the nulls.
   end for
   end for
   For each RANGE, `data` is every array from every REGION within our *limits*.
9. If no local minima are found in the REGION, then the output boundary is set as the "Higher Power Limit" location, which is the most restrictive.
10. Apply Bayesian tracking, as introduced in Chapter 5, to choose the final decision for the first-order boundaries using Algorithms 31 through 34 in Chapter 7.

Chapter 5

Object Tracking and Bayesian Estimation

In this Chapter we first present an introduction to Object Tracking in Section 5.1. The objects we will be tracking are the Doppler bin locations of the first-order Bragg boundaries (see Chapter 4) that vary with both range and with time of observation. We provide a description of how to track a single object in clutter (Section 5.1.2) as well as using a simple Gauss - Hermite or g-h filter (Section 5.1.3). In Section 5.1.4, we describe our problem: tracking the limits of the first-order boundaries interpreted as four individual single-object tracking problems.

Next, in Section 5.2, we introduce Bayesian estimation and describe the most relevant estimation concepts: Bayes theorem (Section 5.2.1), Gaussian probability distributions (Section 5.2.2), and at the end the Bayesian estimation solution for recursive state estimation (Section 5.2.3).

In Section 5.3, we study the Kalman Filter that we use for our problem, and describe all of the terms required in the Kalman Filter estimation algorithm solution. While Section 5.4 focuses on the Bayesian estimation filters needed when we receive more than one measurement associated with a single first-order boundary location at a given time and range from the radar. First, the general Bayesian solution for simultaneous measurements is presented in Section 5.4.1. This approach is necessary for the times when we are processing simultaneous multiple measurements received for a single location of a first-order boundary when they all correspond to a unique time and range.

Since the general Bayesian solution for simultaneous measurements is complex, two approximations to this general solution are presented:

1. The Nearest Neighbor Filter (Section 5.4.2):

   The NNF filter is an approximation of the optimal Bayesian filter recursive solution for multiple simultaneous measurements that will only choose the measurement whose statistical distance to the predicted measurement of the filter is the smallest.

2. The Probabilistic Data Association Filter (Section 5.4.3):

   The PDAF is the most robust approximation of the optimal Bayesian filter recursive solution for multiple simultaneous measurements. It is one of the most successful algorithms in the class of Bayesian all-neighbors filters, and it performs by updating the object state estimate using all of the validated measurements and their respective posterior probability weightings as described in the corresponding section.

5.1 Object Tracking

5.1.1 Introduction

The concept of object tracking relates to the technique of using measurements from a sensor to determine the location, path and characteristics of an object of interest (*target*). A tracking filter is generally used to provide an accurate and precise estimation of the position, velocity, and acceleration of the target in the presence of uncertainty. This section provides an overview of Object Tracking, before describing our tracking problem in Section 5.1.4.

Object Tracking: Position, Velocity and Acceleration.
When calculating the predicted position of a target we use Newton's equations of motion as:

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{v}_0 \Delta t + \frac{1}{2}\mathbf{a}\Delta t^2, \tag{5.1}$$

where

$\mathbf{x}$ = future target position (m),

$\mathbf{x}_0$ = target initial position (m),

$\mathbf{v}_0$ = target initial velocity (m/s),

$\mathbf{a}$ = target acceleration (m/s^2),

Δt = track update interval (s).

In a three-dimensional Cartesian coordinate system the set of target position *(x, y, z)*, velocity (v_x, v_y, v_z) and acceleration (a_x, a_y, a_z) variables are often referred to as the *system state* $(x, y, z, v_x, v_y, v_z, a_x, a_y, a_z)$.

We combine the system state variables on what is known as a *state space model*:

$$x = x_0 + v_{x0}\Delta t + \frac{1}{2}a_x\Delta t^2$$
$$y = y_0 + v_{y0}\Delta t + \frac{1}{2}a_y\Delta t^2$$
$$z = z_0 + v_{z0}\Delta t + \frac{1}{2}a_z\Delta t^2$$

In a tracking algorithm, the input is the current measurement of the state (position). This target position could be inaccurate due to measurement noise, random errors and uncertainty. One of the outputs of the tracking algorithm is the predicted future state, which can be a difficult calculation due to the aforementioned factors. Therefore, it is necessary for a tracking algorithm to take these uncertainties into account.

5.1.2 Single-Object Tracking in Clutter

Our problem here is tracking a single object state in the presence of noise, using a combination of false alarms and true target observations.
For this problem we will take into account the following assumptions[17]:

- There is only a single object in the surveillance area, and its position is unknown.

- The possible trajectory models of the object are assumed known, and the models are assumed to propagate as Markov chains. (A Markov chain or Markov process is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.)

- In each iteration we may receive one measurement per object. The probability of detection is determined under P_D.

- The clutter is either known or estimated from the received measurements.

- The measurements are produced by the sensor with infinite resolution. Each measurement can only have one source, either from an actual detection of the object or from clutter.

5.1.3 Simple g-h Filter for Object Tracking

Before introducing the concept of Bayesian Estimation and the Kalman Filter in particular, that we do apply for our research problem, we first present a simpler filter known as the g-h filter (or alpha-beta filter). The g-h filter is related to the Kalman filters (Section 5.3) but does not require a detailed system model as we present in the Bayesian Estimation Section 5.2.

The g-h filter is often used for object tracking since it is a simplified version for parameter estimation and smoothing. It approximates by assuming that the system presents only two internal states and that the first state is determined by integrating the second state over time. Therefore, this is a low-order approximation that can be used on simple object tracking problems.

The α and β terms (or g and h), shown in Equation 5.4, are usually found though experiment, and refer to the two scaling factors: α is the scaling we use for the reliability of measurements we receive, while β is the scaling we apply to how the measurement changes over time, in other words, how reliable is our model.

Generally, larger values of α and β result in a quicker response to transient changes, while smaller values reduce the noise level in the estimated state. Thus, the values of α and β should be positive and small for good convergence and stability[26][49][38].

Example: Design of a g-h filter to track the position of a target

We can use the simplified g-h filter for this problem because the position of the target can be determined from the time integral of a series of velocity measurements. Our state variable would be $\mathbf{x}_k$, the target's position predicted at sampling time k. This position equals to the previous target position $\mathbf{x}_{k-1}$ plus the displacement Δx. Since we are only receiving measurements of the velocity, every Δt sampling time, we can predict $\mathbf{x}_k$ as:

$$\mathbf{x}_{k,k-1} = \mathbf{x}_{k-1,k-1} + \mathbf{v}_{k-1,k-1}\Delta t \qquad \text{(m)}, \tag{5.2}$$

where:

$\mathbf{x}_{k,k-1}$ = predicted position at k, given measurements up to k–1 (m),

$\mathbf{x}_{k-1,k-1}$ = filtered position at k-1 given measurements up to k–1 (m),

$\mathbf{v}_{k-1,k-1}$ = filtered velocity at k-1 given measurements up to k–1 (m/s),

Δt = measurement update interval (s).

Since the velocity is assumed to be constant, $\mathbf{v}_{k,k-1} = \mathbf{v}_{k-1,k-1}$.

The measurement at sampling time k, $\mathbf{y}_k$, is expected to deviate from $\mathbf{x}_{k,k-1}$ (the predicted value at time k given all measurements up to k – 1) due to noise and other effects not included in this simplified model.

This error, often referred as residual or innovation, is simply expressed as

$$\mathbf{r}_k = \mathbf{y}_k - \mathbf{x}_{k,k-1}. \tag{5.3}$$

Finally the position and velocity estimates for the g-h filter, using the constants α and β, can be expressed as:

$$\begin{aligned} \mathbf{x}_{k,k} &= \mathbf{x}_{k,k-1} + \alpha \mathbf{r}_k \quad \text{(m)} \\ \mathbf{v}_{k,k} &= \mathbf{v}_{k,k-1} + \frac{\beta}{\Delta t}\mathbf{r}_k \quad \text{(m/s)} \end{aligned} \tag{5.4}$$

where

α, β = filter constants,

$\mathbf{r}_k$ = residual at update k (m),

$\mathbf{x}_{k,k-1}$ = predicted position at update k (m),

$\mathbf{v}_{k,k-1}$ = predicted velocity at update k (m/s),

$\mathbf{x}_{k,k}$ = filtered position at update k given measurements up to k (m),

$\mathbf{v}_{k,k}$ = filtered velocity at update k given measurements up to k (m/s).

Following the equations, the algorithm for the g-h filter proceeds as follows:

1. Select values for α and β.
2. Initialize $\mathbf{x}_{0,0}$ and $\mathbf{v}_{0,0}$ with initial measurement values, or set to zero.
3. Predict state $\mathbf{x}_{k,k-1}$ and $\mathbf{v}_{k,k-1}$ – Eq. (5.4).
4. Calculate the residual $\mathbf{r}_k$ and adjust the predicted states – Eq. (5.3).
5. Repeat steps (3)–(4) for each new measurement with $\mathbf{x}_{k,k}$ and $\mathbf{v}_{k,k}$ as the filter output.

5.1.4 Our Object tracking Problem: First-Order Boundaries

The heart of the problem we are investigating is to track the four first- order boundaries, which correspond to the maximum ocean current velocity present in a specific location at a specific time.

In our case we are tracking the position of the first-order boundary limits separately, each one being an *object* in the object tracking problem we perform. As in any object tracking problem, there will be uncertainties in the measurements and they will be taken into consideration.

Overview of Our First-Order Boundaries Tracking Problem:

1. What are we tracking?

   The location of the first-order boundaries i.e. what is the velocity of the maximum ocean current present at that time and at that rangebin (Section 3.2). The four first-order boundaries are tracked independently, thus we are dealing with a *single-object tracking in clutter* problem (Section 5.1.2) that we will run four independent times, using data from radar observations of the Doppler spectra from each single range bin.

2. How do we track them?

   We face a state estimation problem where the object states - first order boundaries - are to be estimated from noisy measurements.

   The tracking system consists of the *object* to be tracked, a *sensor* (the SeaSonde radar) which measures some aspect of the object, and also provides us with some *noisy measurements* (i.e. the velocity of maximum ocean current present at that time and at that rangebin). Therefore we will use a Kalman Filter (Section 5.3) for this tracking problem.

3. What are our observations or measurements?

   From our *sensor* (the SeaSonde radar) we will be able to process a series of backscatter spectra files, which correspond to the specific time we are studying. We will go range bin by range bin using their respective spectra files, and find the initial estimates for each of the four first-order boundaries for each range bin spectra file. Those initial estimates are the input *noisy measurements* we will be sending to our Kalman Filter (Section 5.3).

4. What if we have multiple initial estimates for a single boundary location?

   Sometimes we will have multiple estimates for a single boundary location at a specific time and range. As we will see, we will apply the concepts of a Bayesian filter receiving *multiple simultaneous measurements* (Section 5.4) to choose between them:

   - The Nearest Neighbor Filter (first introduced in Section 5.4.2 and described in detail in Algorithm 31 in Section 6.5) would be a simple option to decide between multiple single boundary location possibilities.
   - The Probabilistic Data Association Filter (Section 5.4.3 and Algorithm 32 in Section 7.5) is a more complex filter applicable to our problem which will statistically weight and estimate a boundary location somewhere in between the multiple single boundary location possibilities we receive as measurements.

5. How do we filter the four boundaries once we have chosen the complete set of initial estimates for them?

   We realized that tracking in time and range made sense for the first order line problem, physically the maximum ocean current present should not change many cm/s (or Doppler bins) when advancing in range (about 5 km step) or moving forward in time (about 30 minutes time). This two-dimensional tracking in range and in time is fully described in Chapter 7 and it is a novel application of Bayesian tracking for this first-order boundaries determination problem.

5.2 Bayesian Estimation

The Bayesian method is a probabilistic and statistical theory that can be applied to the modeling and solution of object tracking problems. In most object tracking problems, measurements from sensors are received sequentially over time. At each of the measurement cycles, new estimates of the object state are made by combining new information (*measurement*) and estimates of the object state at the current time. At the next stage, the latest estimates are updated in view of the new measurement coming in.

When developing Bayesian estimation algorithms, we realize how including all possible sources of information (models, estimates, measurements) will help in obtaining the best estimate possible, never throwing any information away, no matter how poor it is. If we only form estimates from the measurement then the prediction will not affect the result. If we only form estimates from the prediction then the measurement will be ignored. Therefore, for a successful estimation we need to take some kind of blend of the prediction and measurement.

5.2.1 Bayes Theorem

The recursive approach used in Bayesian estimation is based on the Bayes theorem, which goes back to the work of Thomas Bayes and Pierre-Simon de Laplace in the 18th and early 19th century, and is now the theoretical framework for reasoning under uncertainty.

The Bayes Theorem states that

$$P(X|Y) = \frac{P(X)\ P(Y|X)}{P(Y)} \tag{5.5}$$

where

X: Random variable of interest, in object tracking X is the *state of the object*.

Y: Measurement related to X. Given Y we want to update the existing knowledge of the variable of interest X.

P(X): The *prior*, attributes probability values to X.

P(X|Y): The *posterior*, is the conditional probability distribution of X given the measurement Y.

$\mathcal{L}$(X): Also known as the *Likelihood* or P(Y|X) is the probability that each X would result in the measurement Y.

P(Y): The *normalization factor*, ensures that the resultant probability distribution P(X|Y) satisfies the axioms of probability[17] and sums to 1.

Let X be the random variable of interest, thus the *state* of the object under consideration. Let Y be a *measurement* of the state of X. We can interpret the theorem as a rule for updating our knowledge of the event X, given a measurement labeled as Y.

A description of Eq.(5.5) would be that "the probability of X given the data Y is proportional to the probability that the data Y was observed given X and multiplied by the prior probability of that event X to be true".

$$\texttt{posterior} = \frac{\texttt{likelihood} \times \texttt{prior}}{\texttt{normalization factor}} \tag{5.6}$$

The term $\mathcal{L}(\mathrm{X}) = \mathrm{P}(\mathrm{Y}|\mathrm{X})$ is known as the *likelihood function* as a function of X. The term P(Y) is a *normalization factor* since it does not depend on the event X that we are evaluating. On the other hand, the *prior probability distribution* of X, the term P(X), is our best estimate of the probability of the event before the observations are made.

In other words, the conditional probability of event X given an observation of event Y is equal to P(X,Y) normalized by P(Y). Where P(X,Y) is the joint probability of events X and Y, and P(Y) is the unconditional probability of event Y. The resulting distribution of X, P(X|Y), is called the *posterior probability distribution.*

5.2.2 Gaussian Distributions

In Kalman filters we express all probabilities as Gaussian distributions, thus $\mathrm{P}(\mathrm{X}) \sim \mathcal{N}(\mu, \sigma^2)$, where μ is the mean and the variance is σ^2.
Equation (5.7) represents a Gaussian probability distribution function:

$$P(X) \sim \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.7}$$

As described previously, our objective is to combine prior information with observations (measurements).

The Bayes Theorem, Equation (5.5), describes how to obtain the probability of a hypothesis given an observation. As we have seen on Equations (5.5) and (5.6), the posterior P(X|Y) is proportional to the multiplication of P(Y|X) and P(X), which are two Gaussian distributions. Thus the posterior P(X|Y) is a Gaussian as well.

In addition, we can derive the mean and variance of the posterior P(X|Y), from the formulas of the mean and variance of the product of two Gaussian distributions presented below.

Gaussian probability density function 1: $P(X_1) \sim \mathcal{N}(\mu_1, \sigma_1^2)$
Gaussian probability density function 2: $P(X_2) \sim \mathcal{N}(\mu_2, \sigma_2^2)$

And by multiplying $P(X_1)$ and $P(X_2)$, we obtain a third Gaussian[38][17] which has the mean μ and variance σ^2 presented below in Eq.(5.8).

Resulting Gaussian probability density function: $P(X) \sim \mathcal{N}(\mu, \sigma^2)$, where the mean μ and variance σ^2 of P(X) are:

$$\mu = \frac{\sigma_1^2 \mu_2 + \sigma_2^2 \mu_1}{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad \sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{5.8}$$

Similarly, for the Gaussian posterior density P(X|Y) described in Eq.(5.5), we can derive its mean and variance as shown below.

Gaussian probability density function of the prior:

$$\mathrm{P(X)} \sim \mathcal{N}(\mu_x, \sigma_x^2)$$

Gaussian probability density function of the likelihood $\mathcal{L}$(X):

$$\mathcal{L}(\mathrm{X}) = \mathrm{P(Y|X)} \sim \mathcal{N}(\mu_y,\, \sigma_y^2)$$

Resulting Gaussian probability function of the posterior and its mean and variance:

$$\mathrm{P(X|Y)} \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mu = \frac{\sigma_x^2 \mu_y + \sigma_y^2 \mu_x}{\sigma_x^2 + \sigma_y^2} \quad \text{and} \quad \sigma^2 = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2} \tag{5.9}$$

Next, we are going to show two examples of the Gaussian result for a posterior probability density of our variable of interest X.

In the first example, we assume that our Gaussian prior has a mean of μ=35 and standard deviation σ=2, thus P(X) $\sim \mathcal{N}(\mu$=35, σ^2=4) as presented in Figure 5.1. Having that prior at the present time we would think that the value of X will be around 35, but we receive a measurement quite far from the prior, Y=15, shown in red on Figure 5.2. The observation brings some uncertainty, in this case a large error, which results in a likelihood probability P(Y|X) $\sim \mathcal{N}(\mu$=15, σ^2=9). After combining these two Gaussians –following Eq.(5.9)– we obtain the resulting *posterior* P(X|Y)$\sim \mathcal{N}(\mu$=28.85, σ^2=2.77), as we see in Figure 5.3.

As expected, the variance of our result is better than the variances of the prior and the likelihood probability functions, because we are combining two sets of information and therefore our uncertainty decreases. As for the mean, our resulting estimation is centered around 28.85 which is closer to the prior because its variance was smaller, and therefore the system model gives more reliability to the prior than to the measurement.

For the second example, and as opposed to the first, we assume that our prior is accurate respect of the true value of X, and that we receive a measurement which is closer to the mean set by the prior as well.

In this case, our Gaussian prior has a mean of μ=20 and standard deviation σ=1, thus P(X) $\sim \mathcal{N}(\mu$=20, σ^2=1) as presented in Figure 5.4. We receive a pretty reliable measurement P(Y|X) $\sim \mathcal{N}(\mu$=22, σ^2=1.25^2) with a standard deviation of σ=1.25 and mean μ=22, shown in red on Figure 5.5.

The result of combining those two Gaussians is the posterior probability density function shown in orange in Figure 5.6. In this case, the posterior has a mean of μ=20.78 and its variance has been reduced to σ^2=0.61.

Interpretation of the Bayes Theorem for our problem: We are tracking the First Order Boundaries, which correspond to the maximum ocean current velocity present in a specific location at a specific time. We track the four boundaries separately, each one being the *object* in the object tracking problem we study. Example: the location of the first order line is at bin 300, and the variance is 2 bins, thus the Gaussian describing the state of the object would be

$$\text{P(X)} \sim N\ (\mu=300,\ \sigma^2=2).$$

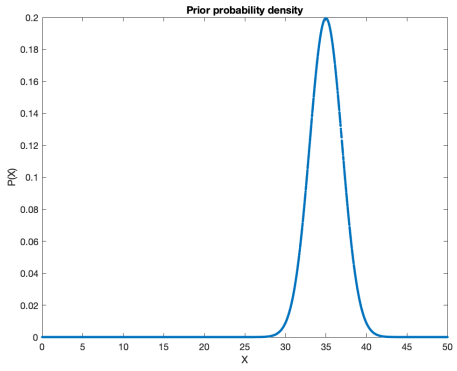

Figure 5.1: The probability density function of the *prior* $\mathrm{P(X)} \sim \mathcal{N}(\mu{=}35,\ \sigma^2{=}4)$ is shown in blue.

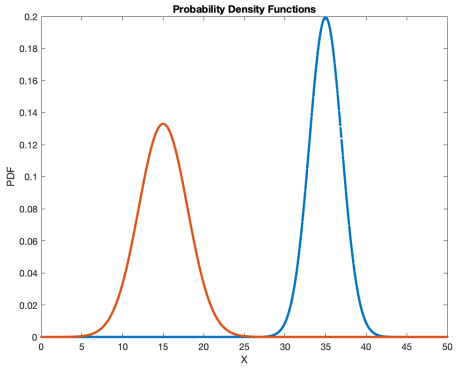

Figure 5.2: While the probability function of the *prior* is $\mathrm{P(X)} \sim \mathcal{N}(\mu{=}35,\ \sigma^2{=}4)$ is shown in blue, we receive a measurement Y with its *likelihood* $\mathrm{P(Y|X)} \sim \mathcal{N}(\mu{=}15,\ \sigma^2{=}9)$ shown in red.

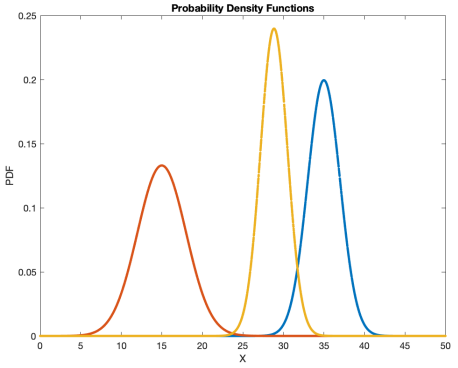

Figure 5.3: After combining the *prior* $\mathrm{P(X)} \sim \mathcal{N}(\mu{=}35,\ \sigma^2{=}4)$ in blue and the *likelihood* of the measurement $\mathrm{P(Y|X)} \sim \mathcal{N}(\mu{=}15,\ \sigma^2{=}9)$ in red we obtain a third Gaussian, $\mathrm{P(X|Y)} \sim \mathcal{N}(\mu{=}28.85,\ \sigma^2{=}2.77)$ which is our *posterior* and thus the predicted mean value of X at the current time would be x=28.85.

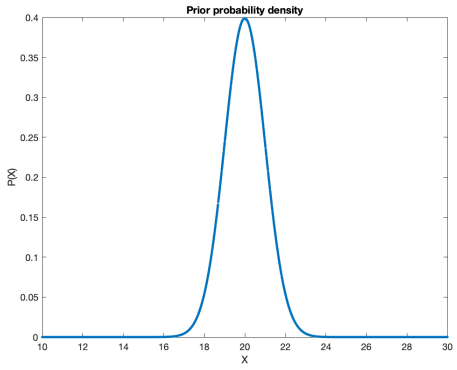

Figure 5.4: The probability density function of the *prior* $\mathrm{P(X)} \sim \mathcal{N}(\mu{=}20,\ \sigma^2{=}1)$ is shown in blue.

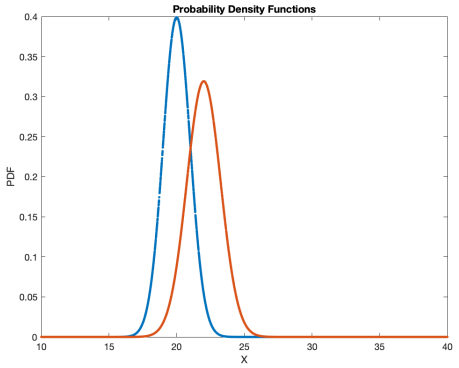

Figure 5.5: While the probability function of the *prior* is $\mathrm{P(X)} \sim \mathcal{N}(\mu{=}20,\ \sigma^2{=}1)$ is shown in blue, we receive a measurement Y with its *likelihood* $\mathrm{P(Y|X)} \sim \mathcal{N}(\mu{=}22,\ \sigma^2{=}1.25^2)$ shown in red.

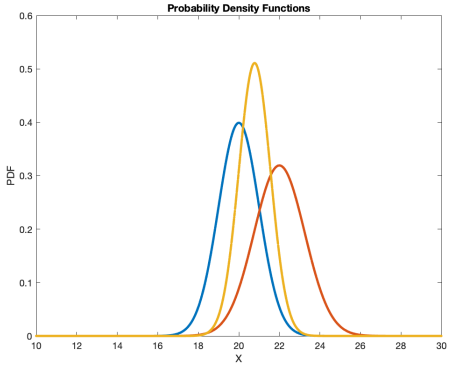

Figure 5.6: After combining the *prior* $\mathrm{P(X)} \sim \mathcal{N}(\mu{=}20,\ \sigma^2{=}1)$ in blue and the *likelihood* of the measurement $\mathrm{P(Y|X)} \sim \mathcal{N}(\mu{=}22,\ \sigma^2{=}1.25^2)$ in red we obtain a third Gaussian, $\mathrm{P(X|Y)} \sim \mathcal{N}(\mu{=}20.78,\ \sigma^2{=}0.61)$ which is our *posterior* and thus the predicted mean value of X at the current time would be x=20.78.

5.2.3 Bayesian Solution for Recursive Measurements

When receiving measurements at several times $k = 1, 2...$ we apply the recursive Bayesian solution.

To compute the Bayesian solution at time k, we take the posterior distribution solution at time $k-1$, which can be written as $\mathrm{P(X|Y_1, Y_2, ...Y_{k-1})}$ and use it as the prior distribution to compute the posterior at time k.

$$\mathrm{P(X \mid Y_1, ...Y_k)} = \frac{\texttt{likelihood} \times \mathrm{P(X \mid Y_1, Y_2, ...Y_{k-1})}}{\texttt{normalization factor}} \tag{5.10}$$

The complete **general Bayesian recursive equation** for the state conditional probability density is shown in Eq.(5.11) below and its components are described next.

$$\mathrm{P(X_k \mid Y^k)} = \frac{\mathrm{P(Y_k \mid X_k)}}{\mathrm{P(Y_k \mid Y^{k-1})}} \int_{\mathrm{X_{k-1}}} \mathrm{P(X_k \mid X_{k-1}) P(X_{k-1} \mid Y^{k-1}) dX_{k-1}} \tag{5.11}$$

Note the following terms describing relevant probability densities present at the general Bayesian recursive equation (5.11). For reference see Challa[17]:

- $\mathrm{P(Y_k \mid X_k)}$ is the *Likelihood* $\mathcal{L}(\mathrm{X})$, or probability that each X would result in the measurement Y.
- $\mathrm{P(X_k \mid X_{k-1})}$ is the probability density of the transition from time $k-1$ to time k.
- As seen above, $\mathrm{P(X_{k-1} \mid Y^{k-1})}$ is the posterior probability density function at the previous time $k-1$, which now (at time k) is multiplying the transition density.

- The integral of the last two terms mentioned, known as the predicted density at time k, is

$$\int_{X_{k-1}} P(X_k \mid X_{k-1}) P(X_{k-1} \mid Y^{k-1}) dX_{k-1} \tag{5.12}$$

  and can also be expressed as simply $\mathrm{P(X_k \mid Y^{k-1})}$, as we will see below in Eq.(5.13). This is what we use as our *prior* for time k, as described above in Eq.(5.10).

- $\mathrm{P(Y_k \mid Y^{k-1})}$ is the *normalization factor*.

Finally, the four **recursive Bayesian algorithm steps** are *Prediction* of the state of X, *Update* the prediction with the measurement Y, *Normalize* the update and determine the estimate through the *Posterior*, and they are described below.

1. *Prediction step*:

   The Chapman-Kolmogorov equation[17]

$$\mathrm{P(X_k \mid Y^{k-1}) = \int_{X_{k-1}} P(X_k \mid X_{k-1}) P(X_{k-1} \mid Y^{k-1}) dX_{k-1}} \tag{5.13}$$

   produces $\mathrm{P(X_k \mid Y^{k-1})}$ which is the predicted state of the elements of X_k given the previous state X_{k-1} and all measurements up to $k-1$, i.e. Y^{k-1}. Within the Chapman-Kolmogorov equation integral we find the posterior probability density function for the previous time, which is $\mathrm{P(X_{k-1} \mid Y^{k-1})}$. We also see the probability function $\mathrm{P(X_k \mid X_{k-1})}$, which is known as the transition density between the states of our variable.

2. *Update step*:

   Upon receiving the newest measurement Y_k we proceed to the *correction* through multiplying by the likelihood $\mathcal{L}(X)$=$P(Y_k \mid X_k)$.

3. *Normalize the update*:

   Obtain the posterior probability density by dividing the normalization factor $P(Y_k \mid Y^{k-1})$ and thus normalizing the updated state.

4. *Posterior density*:

   Finally, the full equation for $P(X_k \mid Y^k)$ –Eq.(5.11)– is the posterior probability density function of the object state X_k conditioned on the available measurements up to Y_k.

At this point, we introduce the **additive noise** and list a series of assumptions for optimal Bayesian tracking and Kalman filters.

1. Given a current state of the variable of interest $\mathbf{x}_k$, its previous state $\mathbf{x}_{k-1}$ and a measurement at the current time $\mathbf{y}_k$; we refer to the Additive White Gaussian Noise (AWGN) in the state variable at the current time k as $\mathbf{v}_k$, while $\mathbf{w}_k$ refers to the noise associated with the measurement.

2. The object dynamics and measurement equations in Eq.(5.14) that include the AWGN are linear. The matrices $\mathbf{F}$ and $\mathbf{H}$ are assumed to be the functions containing the transitions from one state to the next.

3. $\mathbf{v}_\mathrm{k}$ and $\mathbf{w}_\mathrm{k}$ are white, uncorrelated, Gaussian noise sequences with zero mean and their corresponding covariances are the matrices $\mathbf{Q}_\mathrm{k}$ and $\mathbf{R}_\mathrm{k}$.

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k \qquad \mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k \tag{5.14}$$

5.3 The Kalman Filter

The most important and most often used estimation algorithm is the Kalman Filter, named after Rudolph E. Kalman. The original paper[32] published in 1960, presented a recursive solution to a linear filtering problem with discrete data. The Kalman filter is optimal in the sense that the estimated error uncertainty is minimized under the specific conditions listed above, and it follows the concept of predict and correct described on Section 5.1.3 for the g-h filter. Even when the minimal conditions are not met, the Kalman filter is still chosen for many applications since it still performs quite well.

Hence we introduce the Kalman Filter (KF) as a Bayesian Optimal filter that also assumes the three assumptions[17] we listed in Section 5.2.3 above related to Gaussian additive noise and linearity of its state equations.

From the object dynamics and measurement equations in Eq. (5.14) and following the form $\mathrm{P(X)} \sim \mathcal{N}(\mu, \sigma^2)$ to define a Gaussian probability density function, we present the subsequent functions that fully define a Kalman Filter[49]:

- The *transition* probability density function of the object state from time $k-1$ to the current time k, $\mathrm{P}(\mathrm{X_k} \mid \mathrm{X_{k-1}})$, is a Gaussian with mean $\mathbf{F}\mathbf{x}_{\mathrm{k-1}}$ and covariance $\mathbf{Q}_{\mathrm{k}}$.

$$\mathrm{P}(\mathrm{X_k} \mid \mathrm{X_{k-1}}) \sim \mathcal{N}(\mathbf{F}\mathbf{x}_{\mathrm{k-1}}, \mathbf{Q}_{\mathrm{k}})$$

- The *posterior* probability density function of the object state at time $k-1$, $\mathrm{P}(\mathrm{X_{k-1}} \mid \mathrm{Y^{k-1}})$, is a Gaussian with mean $\hat{\mathbf{x}}_{\mathrm{k-1|k-1}}$ and covariance $\mathbf{P}_{\mathrm{k-1|k-1}}$.

$$\mathrm{P}(\mathrm{X_{k-1}} \mid \mathrm{Y^{k-1}}) \sim \mathcal{N}(\hat{\mathbf{x}}_{\mathrm{k-1|k-1}}, \mathbf{P}_{\mathrm{k-1|k-1}})$$

- The *prior* or predicted probability density function of the object state at time k (before receiving the measurement $\mathbf{y}_{\mathrm{k}}$), $\mathrm{P}(\mathrm{X_k} \mid \mathrm{Y^{k-1}})$, is a Gaussian with mean $\hat{\mathbf{x}}_{\mathrm{k|k-1}}$ and covariance $\mathbf{P}_{\mathrm{k|k-1}}$.

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}$$
$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q}_k$$

$$\begin{aligned} P(X_k \mid Y^{k-1}) &= \int_{X_{k-1}} P(X_k \mid X_{k-1}) P(X_{k-1} \mid Y^{k-1}) dX_{k-1} \\ &= \int \mathcal{N}(\mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_k) \cdot \mathcal{N}(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \\ P(X_k \mid Y^{k-1}) &= \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \end{aligned} \tag{5.15}$$

- The *likelihood* probability density function of the measurement or observation received at time k, $\mathrm{P}(\mathrm{Y_k} \mid \mathrm{X_k})$, is a Gaussian with mean $\mathbf{H}\mathbf{x}_{\mathrm{k}}$ and covariance $\mathbf{R}_{\mathrm{k}}$.

$$\mathrm{P}(\mathrm{Y_k} \mid \mathrm{X_k}) \sim \mathcal{N}(\mathbf{H}\mathbf{x}_{\mathrm{k}}, \mathbf{R}_{\mathrm{k}})$$

- The *normalization factor* probability density function of the measurement received at time k given all the measurements received up to time $k-1$, $\mathrm{P}(\mathrm{Y_k} \mid \mathrm{Y^{k-1}})$, is a Gaussian with mean $\hat{\mathbf{y}}_{\mathrm{k|k-1}}$ and covariance $\mathbf{S}_{\mathrm{k}}$.

$$\begin{aligned}
\hat{\mathbf{y}}_{k|k-1} &= \mathbf{H}\hat{\mathbf{x}}_{k|k-1} \\
\mathbf{S}_k &= \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k \\
P(Y_k \mid Y^{k-1}) &= \int_{X_k} P(Y_k \mid X_k) P(X_k \mid Y^{k-1}) dX_k \\
&= \int \mathcal{N}(\mathbf{H}\mathbf{x}_k, \mathbf{R}_k) \cdot \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) = \mathcal{N}(\hat{\mathbf{y}}_{k|k-1}, \mathbf{S}_k)
\end{aligned}$$

- The *posterior* probability density function –Eq.(5.11)– at time k given all the measurements received up to time k, $\mathrm{P}(\mathrm{X_k} \mid \mathrm{Y^k})$, is a Gaussian with mean $\hat{\mathbf{x}}_{\mathrm{k|k}}$ and covariance $\mathbf{P}_{\mathrm{k|k}}$.

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1}(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}) \\
\mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1}\mathbf{H}\mathbf{P}_{k|k-1} \\
P(X_k \mid Y^k) &= \frac{\mathcal{N}(\mathbf{H}\mathbf{x}_k, \mathbf{R}_k) \cdot \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})}{\mathcal{N}(\mathbf{H}\hat{\mathbf{x}}_{k|k-1}, \mathbf{S}_k)} \\
&= \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})
\end{aligned} \tag{5.16}$$

5.4 Solutions for Object Tracking with Simultaneous Measurements

For the problem under study in this document, we consider each of the four first-order boundaries as a separate object tracking problem, and therefore we are estimating four boundaries independently.

Every time a measurement is received, we want to recursively estimate the current state of the variable under study – the location of each of those first-order boundaries. In this section, we take into consideration that multiple measurements can be received simultaneously for each of the boundaries, e.g. a few different possible locations for the first-order boundary and we want to estimate a final location considering those sensor measurements.

Since the optimal Bayesian solution presented in 5.4.1 is the most complex option, two approximations for this optimal recursion have been successfully used and implemented:

- the Nearest Neighbor Filter described in section 5.4.2,
- the Probabilistic Data Association Filter presented in section 5.4.3.

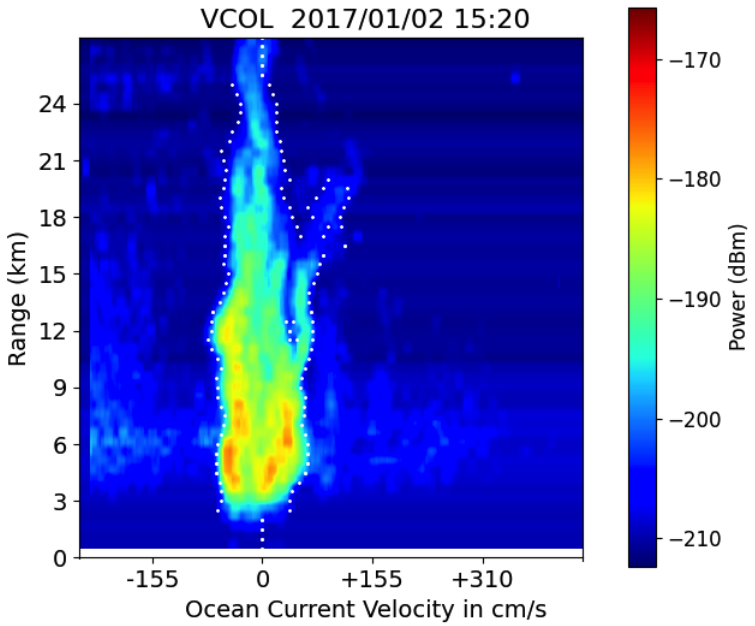

Figure 5.7: Example showing multiple simultaneous measurements received for the determination of the right first-order boundary, particularly they start at 15 km range from the radar. In this section we will see the possible algorithms to decide the final estimated location of the boundary when multiple options are possible. In the figure, we see multiple simultaneous possible locations for the right first-order boundary, due to the higher power backscattered energy to the immediate right of the Bragg region between 15 and 21 km range. The null-finding algorithm described in Chapter 4 accepts multiple possibilities and the algorithms described in this section will have to determine the most probable location for the boundary at each range.

5.4.1 Bayesian Solution for Simultaneous Measurements

While the object dynamics remain the same as for the recursive solution for single object tracking in clutter that we saw on Section 5.2.3,

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k, \tag{5.17}$$

the difference in this case relates to the sensor observations.

The *recursive Bayesian solution* is used when we have a sensor that receives a single measurement for every epoch k. On the other hand, in the present Bayesian solution for *simultaneous measurements* our sensor observations at each time $k = 1, 2, 3, ...$ result in a set of simultaneous measurements instead of a single one. Thus, for each time k we receive

$$\mathbf{y}_k = \{\mathbf{y}^k(1), \mathbf{y}^k(2), \mathbf{y}^k(3), ..., \mathbf{y}^k(m_k)\}, \tag{5.18}$$

where m_k denotes the number of received measurements at time k. The whole set of measurements for all epochs is $m^k = \{m_1, m_2, ..., m_k\}$.

The posterior density of the target state will not only be conditioned by the received measurements $\mathbf{y}_k$ like before, but also on the number of measurements m_k received at each time k, as described next.

- The *prior* or predicted probability density function of the object state at time k (before receiving the set of measurements $\mathbf{y}_\mathrm{k}$), is $\mathrm{P}(\mathrm{X_k} \mid \mathrm{Y^{k-1}}, \mathrm{m^{k-1}})$.

$$\text{From Eq.(5.17)} \Rightarrow P(X_k \mid X_{k-1}) = P_{v_k}(\mathbf{x}_k - \mathbf{F}(\mathbf{x}_{k-1}))$$

$$P(X_k \mid Y^{k-1}, m^{k-1}) = \int_{X_{k-1}} P(X_k \mid X_{k-1}) P(X_{k-1} \mid Y^{k-1}, m^{k-1}) dX_{k-1}$$
$$= \int_{X_{k-1}} P_{v_k}(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1}) P(X_{k-1} \mid Y^{k-1}, m^{k-1}) dX_{k-1} \tag{5.19}$$

- The *joint likelihood* probability density function of the measurements or observations received at time k is $\mathrm{P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1})}$.

$$P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) =$$
$$= P(\mathbf{y}_k(1), \mathbf{y}_k(2), \mathbf{y}_k(3), ..., \mathbf{y}_k(m_k), m_k \mid \mathbf{x}_k, \mathbf{y}^{k-1}, \mathbf{m}^{k-1}).$$

- The *normalization factor* density function of the set of measurements received at time k given all the measurements up to time $k-1$ is $\mathrm{P(Y_k, m_k \mid Y^{k-1}, m^{k-1})}$.

$$P(Y_k, m_k \mid Y^{k-1}, m^{k-1}) =$$
$$= \int_{X_k} P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) P(X_k \mid Y^{k-1}, m^{k-1}) dX_k.$$

- Finally, the *conditional* probability density function (*posterior*) at time k given all the measurements received up to time k is $\mathrm{P(X_k \mid Y^k, m^k)}$ and follows the Bayes theorems in Equations (5.5) and (5.6).

$$P(X_k \mid Y^k, m^k) =$$
$$= P(X_k \mid Y_k, m_k, Y^{k-1}, m^{k-1}) =$$
$$= \frac{P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) \cdot P(X_k \mid Y^{k-1}, m^{k-1})}{P(Y_k, m_k \mid Y^{k-1}, m^{k-1})}$$

5.4.2 Nearest Neighbor Filter

The Nearest Neighbor Filter, also known as NNF, is an approximation of the optimal Bayesian filter recursive solution for multiple simultaneous measurements we studied in Section 5.4.1. This approximation is based on five assumptions[17]:

1. The true object exists and is always detected.
2. The measurement that is closest to the predicted measurement (in a statistical sense) comes from the object.
3. All other measurements are thus from clutter.
4. The object motion or behavior obeys linear Gaussian statistics.
5. The measurement noise is white Gaussian.

Hence, the main concept of the Nearest Neighbor filter is that only one measurement (among all validated measurements $\mathbf{y}_k = \{ \mathbf{y}^k(1), \mathbf{y}^k(2), ..., \mathbf{y}^k(m_k)\}$ at time k), denoted $\mathbf{y}_k(i)$, is considered to be target-originated. That measurement $\mathbf{y}_k(i)$ is the one whose statistical distance to the predicted measurement of the filter is the smallest.

Similarly to how we described the probability densities in Section 5.2.3, we present the main concepts as well as the equation terms that form the posterior probability density function of the Nearest Neighbor filter next.

- The object dynamics function $\mathbf{F}(\cdot)$ is a linear function of the object state:

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k$$

- The sensor measurements are also a linear function of the object state:

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k$$

- The object dynamics and measurement equations above are linear. The noise terms $\mathbf{v}_\mathrm{k}$ and $\mathbf{w}_\mathrm{k}$ are white, uncorrelated, Gaussian noise sequences with zero mean (AWGN). The noise in the state variable at the current time k is $\mathbf{v}_\mathrm{k}$, while $\mathbf{w}_\mathrm{k}$ refers to the white noise associated with the measurement. The matrices $\mathbf{F}$ and $\mathbf{H}$ are assumed to be the functions containing the transitions from one state to the next. The corresponding covariances for the noise vectors are the matrices $\mathbf{Q}_\mathrm{k}$ and $\mathbf{R}_\mathrm{k}$.

- The *prior* probability density function of the object state from the time $k-1$ is a Gaussian with mean $\hat{\mathbf{x}}_{\mathrm{k-1|k-1}}$ and covariance $\mathbf{P}_{\mathrm{k-1|k-1}}$.

$$\mathrm{P}(\mathrm{X}_{\mathrm{k-1}} \mid \mathrm{Y}^{\mathrm{k-1}}, \mathrm{m}^{\mathrm{k-1}}) \sim \mathcal{N}(\hat{\mathbf{x}}_{\mathrm{k-1|k-1}}, \mathbf{P}_{\mathrm{k-1|k-1}})$$

- The *transition* probability density function of the object state from time $k-1$ to the current time k is $\mathrm{P}(\mathrm{X}_\mathrm{k} \mid \mathrm{X}_{\mathrm{k-1}})$.

$$\text{From Eq.(5.17)} \Rightarrow P(X_k \mid X_{k-1}) = P_{v_k}(\mathbf{x}_k - \mathbf{F}(\mathbf{x}_{k-1}))$$

Since $\mathrm{P}_{v_k}(\cdot)$ is a Gaussian probability density, the *transition* probability density function is given by:

$$P(X_k \mid X_{k-1}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{Q}_k|^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1})^T \mathbf{Q}_k^{-1}(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1})\right\}$$

Summarizing, and like we saw in Section 5.2.3, the *transition* probability density function is a Gaussian with mean $\mathbf{F}\mathbf{x}_{\mathrm{k}-1}$ and covariance $\mathbf{Q}_{\mathrm{k}}$.

$$\mathrm{P}(\mathrm{X_k} \mid \mathrm{X_{k-1}}) \sim \mathcal{N}(\mathbf{F}\mathbf{x}_{\mathrm{k}-1}, \mathbf{Q}_{\mathrm{k}})$$

- The *predicted* probability density function of the object state at time k (before receiving the set of measurements $\mathbf{y}_{\mathrm{k}}$) is $\mathrm{P}(\mathrm{X_k} \mid \mathrm{Y}^{\mathrm{k}-1}, \mathrm{m}^{\mathrm{k}-1})$.
  Following Eq.(5.19) we can derive:

$$P(X_k \mid Y^{k-1}, m^{k-1}) =$$
$$= \int_{X_{k-1}} P_{v_k}(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1}) P(X_{k-1} \mid Y^{k-1}, m^{k-1}) dX_{k-1}$$

  The terms in the integral are:

$$P_{v_k}(\mathbf{x}_k - \mathbf{F}(\mathbf{x}_{k-1})) = P(X_k \mid X_{k-1}) \sim \mathcal{N}(\mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_k)$$
$$P(X_{k-1} \mid Y^{k-1}, m^{k-1}) \sim \mathcal{N}(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$$

  Finally, using the normal distribution theorem, the predicted probability density function can be simplified as presented next.

$$P(X_k \mid Y^{k-1}, m^{k-1}) =$$
$$\int_{X_{k-1}} P(X_k \mid X_{k-1}) P(X_{k-1} \mid Y^{k-1}, m^{k-1}) dX_{k-1} \qquad (5.20)$$
$$= \int \mathcal{N}(\mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_k) \cdot \mathcal{N}(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \sim \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}).$$

- The *likelihood* probability density function in the NNF is approximated by choosing $\mathbf{y}_k(i)$ from the set of $\mathbf{y}_k$ by ordering the measurements based on

their statistical distance to the predicted measurements[17]. Since both the measurement noise and the model process noise are Gaussian, the statistical distance is determined by the chi-square test. Among all measurements received, only one measurement is selected to associate and update the object state, and the selection is based on the following criterion:

$$\mathbf{y}_k(i) = \underset{y_k(j),\ \forall j \in \{1,\ldots,m_k\}}{\operatorname{argmin}} \left[\mathbf{y}_k(j) - \mathbf{H}\hat{x}_{k|k-1}\right]^T \mathbf{S}_{k|k-1}^{-1} \left[\mathbf{y}_k(j) - \mathbf{H}\hat{x}_{k|k-1}\right], \tag{5.21}$$

where $\mathbf{S}_{\mathrm{k|k-1}} = \mathbf{HP}_{\mathrm{k|k-1}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{\mathrm{k}}$.

- The *normalization factor* probability density function of the measurement received at time k given all the sets of measurements received up to time k-1, $\mathrm{P}(\mathrm{Y_k}, \mathrm{m_k} \mid \mathrm{Y^{k-1}}, \mathrm{m^{k-1}})$, is a Gaussian with mean $\mathbf{H}\hat{\mathbf{x}}_{\mathrm{k|k-1}}$ and covariance $\mathbf{S}_{\mathrm{k|k-1}}$.

$$P(Y_k, m_k \mid Y^{k-1},\ m^{k-1}) =$$
$$\int_{X_k} P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) P(X_k \mid Y^{k-1}, m^{k-1}) dX_k$$

where

$$P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) = \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k),$$
$$\text{and } P(X_k \mid Y^{k-1}, m^{k-1}) = \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}).$$

Finally, it all results in the *normalization factor*:

$$P(Y_k, m_k \mid Y^{k-1},\ m^{k-1}) = \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\hat{\mathbf{x}}_{k|k-1}, \mathbf{S}_{k|k-1}).$$

- The conditional *posterior* probability density function of the object state at time k given all the measurements received up to time k results from

combining the above equations as shown below. The derivation results in a Gaussian with mean $\hat{\mathbf{x}}_{\mathrm{k|k}}$ and covariance $\mathbf{P}_{\mathrm{k|k}}$.

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1}(\mathbf{y}_k(i) - \hat{\mathbf{y}}_{k|k-1}(i))$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1}\mathbf{H}\mathbf{P}_{k|k-1}$$

$$P(X_k \mid Y^k,\ m^k) = \frac{\mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k) \cdot \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})}{\mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\hat{\mathbf{x}}_{k|k-1}, \mathbf{S}_{k|k-1})}$$

$$P(X_k \mid Y^k,\ m^k) = \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \tag{5.22}$$

As we have seen in this section, the Nearest Neighbor filter chooses which measurement is nearest to the predicted measurement to update the state. As opposed to "all-neighbors filters", the nearest neighbor filter can lead to very poor performance when the density of clutter or spurious measurements is high. This algorithm does not properly account for the fact that the measurement chosen for the update may be unrelated to the object, thus it could be choosing clutter as the most reliable observation.

Therefore, for our estimation problem, we end up choosing to use the Probabilistic Data Association Filter described next.

5.4.3 Probabilistic Data Association Filter

The Probabilistic Data Association Filter (PDAF) presents one of the most successful algorithms in the class of Bayesian all-neighbors filters, which is why we chose it for the first-order boundaries tracking problem on this research, as detailed in Chapter 7 and particularly in the Data Association step in Section 7.5.

The PDAF algorithm updates the object state estimate using all of the validated measurements and their respective posterior probability weightings that we will describe in this section.

The PDAF is derived by making the following assumptions[17]:

1. The prediction density in the optimal Bayesian recursion given all past observations is a Gaussian density, although strictly speaking it is a Gaussian mixture.

2. The object being tracked exists and no other object exists.

3. The object behavior obeys linear Gaussian statistics.

4. Only one measurement of the multiple measurements received can be originated from the object of interest.

5. The noise of the measurements is white Gaussian.

6. The object may or may not be detected all the time and is detected with probability of detection P_D.

7. All non-object originated measurements are assumed to originate from clutter that is uniformly distributed in space and Poisson distributed in time.

8. Only measurements that fall within a proximity of the expected measurement (within a validation gate) are considered for processing.

The object dynamics, as well as the measurement and noise models are the same as those in the nearest neighbor filter in Section 5.4.2. The transition and prediction densities are also the same, and they are all listed below.

- The object state function and sensor measurements:

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k$$

The noise terms $\mathbf{v}_\mathrm{k}$ and $\mathbf{w}_\mathrm{k}$ are white, uncorrelated, Gaussian noise sequences with zero mean (AWGN). The noise in the state variable at the current time k is $\mathbf{v}_\mathrm{k}$, while $\mathbf{w}_\mathrm{k}$ refers to the white noise associated with the measurement. The matrices $\mathbf{F}$ and $\mathbf{H}$ are the functions containing the transitions from one state to the next. The corresponding covariances for the noise vectors are the matrices $\mathbf{Q}_\mathrm{k}$ and $\mathbf{R}_\mathrm{k}$.

- The *prior* probability density function of the object state:

$$\mathrm{P}(\mathrm{X}_{\mathrm{k}-1} \mid \mathrm{Y}^{\mathrm{k}-1}, \mathrm{m}^{\mathrm{k}-1}) \sim \mathcal{N}(\hat{\mathbf{x}}_{\mathrm{k}-1|\mathrm{k}-1},\, \mathbf{P}_{\mathrm{k}-1|\mathrm{k}-1}).$$

- The *transition* probability density function of the object state from time $k-1$ to the current time k:

$$\mathrm{P}(\mathrm{X}_\mathrm{k} \mid \mathrm{X}_{\mathrm{k}-1}) \sim \mathcal{N}(\mathbf{x}_\mathrm{k};\ \mathbf{F}\mathbf{x}_{\mathrm{k}-1},\, \mathbf{Q}_\mathrm{k}).$$

- The *predicted* probability density function of the object state at time k (before receiving the set of measurements $\mathbf{y}_\mathrm{k}$) is:

$$P(X_k \mid Y^{k-1}, m^{k-1}) \sim \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}). \tag{5.23}$$

- The *likelihood* probability density function:

  The key difference between the NNF and the PDAF is their way to extract information out of the received measurements.

The likelihood $\mathcal{L}$ of the PDAF is approximated by choosing the subset of measurements from the total measurement set $\mathbf{y}_k$ by gating the measurements based on their statistical distance to the predicted measurements[17]. G is the validation gate, and P_G is the probability that a measurement falls into the validation gate.

Since the clutter measurements are uniformly distributed and have equal probability of being at any point within the gate G, we derive the *likelihood* probability density function for the PDAF as:

$$\begin{aligned} P(Y_k \mid X_k, m_k, \theta_k(i)) = \\ = P(\mathbf{y}_k(1), \mathbf{y}_k(2), \mathbf{y}_k(3), ..., \mathbf{y}_k(m_k) \mid X_k, m_k, \theta_k(i)) = \\ = \left(\frac{1}{V_k}\right)^{m_k - 1} P(\mathbf{y}_k(i) \mid X_k) \end{aligned}$$

The term $P(\mathbf{y}_k(i) \mid X_k)$ is the likelihood of the ith measurement being object-originated, and the term $\theta_k(i)$ is the association event that the ith measurement is from the target and the rest are from clutter.

When the complete space is considered, $P(\mathbf{y}_k(i) \mid X_k) \sim \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k)$. However, since the effective measurement space is truncated by the validity gate then the likelihood is effectively truncated as well.

By normalizing it with its area in the gate (the gating probability P_G) it can be rewritten as:

$$P(\mathbf{y}_k(i) \mid X_k) = \frac{1}{P_G} \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k).$$

Finally, the *likelihood* probability function $P(Y_k \mid X_k, m_k, \theta_k(i))$ will be as follows, where $i = 0$ refers to the the first measurement.

$$P(Y_k \mid X_k, m_k, \theta_k(i)) = \begin{cases} \left(\frac{1}{V_k}\right)^{m_k - 1} \frac{1}{P_G} \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k), & \forall i \neq 0, \\ \left(\frac{1}{V_k}\right)^{m_k}, & i = 0. \end{cases} \tag{5.24}$$

- The *normalization factor* and the *joint likelihood*.

  First, we will describe the derivation of the *joint likelihood* probability density function.

  From the Bayes theorem in Eq.(5.5) we can derive:

$$P(X_k \mid Y^k, m^k) = \frac{P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1})\ P(X_k \mid Y^{k-1}, m^{k-1})}{P(Y_k, m_k \mid Y^{k-1}, m^{k-1})} \tag{5.25}$$

  The first term in the equation above is the *joint likelihood*

$$P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}),$$

  which is the density function of the measurements or observations received at time k.

  Therefore the *joint likelihood* can be derived as follows:

$$P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) =$$

$$= \sum_{i=0}^{m_k} P(\mathbf{y}_k(1), \mathbf{y}_k(2), ..., \mathbf{y}_k(m_k) \mid X_k, m_k, \theta_k(i)) P(\theta_k(i) | m_k) P(m_k \mid Y^{k-1}, m^{k-1}) =$$

$$P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) = \left(\frac{1}{V_k}\right)^{m_k} P(\theta_k(0)|m_k)P(m_k \mid Y^{k-1}, m^{k-1}) + \left(\frac{1}{V_k}\right)^{m_k-1} \sum_{i=1}^{m_k} P(\mathbf{y}_k(i) \mid X_k)P(\theta_k(i)|m_k)P(m_k \mid Y^{k-1}, m^{k-1}) \tag{5.26}$$

where:

$$\begin{cases} P(\theta_k(0)|m_k) = (1 - P_D P_G)\dfrac{\mu_F(m_k)}{\mu_F(m_k - 1)} \left[P_D P_G + \dfrac{\mu_F(m_k)}{\mu_F(m_{k-1})}(1 - P_D P_G)\right]^{-1} \\ \\ P(\theta_k(i)|m_k) = \dfrac{1}{m_k} P_D P_G \left[P_D P_G + \dfrac{\mu_F(m_k)}{\mu_F(m_{k-1})}(1 - P_D P_G)\right]^{-1} \end{cases}$$

Note that $\mu_F(m_k)$ is the distribution of m_k clutter measurements that are gated.

The *normalization factor* probability density function of the measurement received at time k given all the sets of measurements received up to time $k-1$ is expressed as $\mathrm{P(Y_k, m_k \mid Y^{k-1}, m^{k-1})}$ and we describe its derivation next.

From the Bayes equation Eq.(5.25), the normalization factor in the denominator can be isolated as:

$$P(Y_k, m_k \mid Y^{k-1},\ m^{k-1}) = \int_{X_k} P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1})P(X_k \mid Y^{k-1}, m^{k-1})dX_k. \tag{5.27}$$

Combining the *joint likelihood* in Eq.(5.26) into the normalization factor integral in Eq.(5.27) we obtain the following:

$$\int_{X_k} \Big(P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) P(X_k \mid Y^{k-1}, m^{k-1}) \Big) \; dX_k =$$

$$\int_{X_k} (\texttt{TERM1} \times \texttt{TERM2} \times \texttt{TERM3}) \; dX_k,$$

where

$$\begin{cases} \texttt{TERM1} = P(m_k \mid Y^{k-1}, m^{k-1}) \\ \texttt{TERM2} = \left[\dfrac{P(\theta_k(0)|m_k)}{V_k^{m_k}} + \dfrac{\sum_{i=1}^{m_k} P(\theta_k(i)|m_k) P(Y_k(i) \mid X_k)}{V_k^{m_k - 1}} \right] \\ \texttt{TERM3} = P(X_k \mid Y^{k-1}, m^{k-1}) \sim \mathcal{N}(\mathbf{x}_k; \; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \end{cases}$$

Next we continue the derivation as:

$$\int_{X_k} \Big(P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) P(X_k \mid Y^{k-1}, m^{k-1}) \Big) \; dX_k =$$

$$\int_{X_k} \Big([\texttt{TERM4} + \texttt{TERM5}] \cdot \mathcal{N}(\mathbf{x}_k; \; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \Big) \, dX_k$$

where

$$\begin{cases} \texttt{TERM4} = \dfrac{P(m_k \mid Y^{k-1}, m^{k-1}) \; P(\theta_k(0)|m_k)}{V_k^{m_k}} \\ \texttt{TERM5} = \dfrac{\sum_{i=1}^{m_k} P(m_k \mid Y^{k-1}, m^{k-1}) \; P(\theta_k(i)|m_k) \; P(Y_k(i) \mid X_k)}{V_k^{m_k - 1}} \end{cases}$$

After a series of derivations described by Challa[17], we can rewrite the integral as a sum of two terms:

$$\int_{X_k} \Big(P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1}) P(X_k \mid Y^{k-1}, m^{k-1}) \Big) \; dX_k = \texttt{TERM4} + \texttt{TERM7}$$

Where

$$\texttt{TERM7} = \sum_{i=1}^{m_k} \frac{P(m_k \mid Y^{k-1}, m^{k-1})\ P(\theta_k(i)|m_k)}{P_G V_k^{m_k-1}}$$
$$\times \int_{X_k} \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})\ dX_k$$

We establish the equation for the mean value $\hat{\mathbf{y}}_k$ and the covariance $\mathbf{S}_k$ and rewrite the integral in TERM7 below:

$$\begin{aligned} \hat{\mathbf{y}}_k &= \mathbf{H}\hat{\mathbf{x}}_{k|k-1} \\ \mathbf{S}_k &= \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k \\ \int_{X_k} \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})\ dX_k &= \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k) \end{aligned} \tag{5.28}$$

Finally the *normalization factor* can be summarized as:

$$\begin{aligned} P(Y_k, m_k \mid Y^{k-1},\ m^{k-1}) = \\ \frac{P(m_k \mid Y^{k-1}, m^{k-1})\ P(\theta_k(0)|m_k)}{V_k^{m_k}} \\ + \sum_{i=1}^{m_k} \frac{P(m_k \mid Y^{k-1}, m^{k-1})\ P(\theta_k(i)|m_k)}{P_G V_k^{m_k-1}} \times\ \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k) \end{aligned} \tag{5.29}$$

- The *data association probabilities* in the Probabilistic Data Association Filter PDAF are denoted by $\beta_k(i)$ (also known as "weights"), which are calculated for the ith validated measurements, i.e. the measurements that fall within the validation gate V_k.

- The *posterior* probability $P(X_k \mid Y^k,\ m^k)$ that we are seeking to find to estimate X_k, in other words the *conditional* probability density function of the object state at time k given all the measurements received up to time k results is studied next. To derive the *posterior* we need to combine:

– The joint likelihood in Eq.(5.26)

$$P(Y_k, m_k \mid X_k, Y^{k-1}, m^{k-1})$$

– The predicted density in Eq.(5.23), we use κ for simplification.

$$\kappa = P(X_k \mid Y^{k-1}, m^{k-1}) = \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$$

– The normalization factor in Eq.(5.29), we use δ for simplification.

$$\delta = P(Y_k, m_k \mid Y^{k-1},\ m^{k-1})$$

– The probability P_G that a measurement which is object-originated will fall into the gate.

– The *data association probabilities* or "weights", denoted by $\beta_k(i)$.

The expression for the posterior density function is derived as follows:

$$P(X_k \mid Y^k,\ m^k) = \frac{1}{P(Y_k, m_k \mid Y^{k-1},\ m^{k-1})} \times$$
$$\sum_{i=0}^{m_k} P(\mathbf{y}_k(1), \mathbf{y}_k(2), ..., \mathbf{y}_k(m_k) \mid X_k, m_k, \theta_k(i)) P(\theta_k(i)|m_k) \times P(X_k \mid Y^{k-1}, m^{k-1})$$

$$P(X_k \mid Y^k,\ m^k) =$$
$$\frac{1}{\delta}\left(\frac{1}{V_k}\right)^{m_k}\ P(X_k \mid Y^{k-1}, m^{k-1})\ P(\theta_k(0)|m_k) P(m_k \mid Y^{k-1}, m^{k-1})$$
$$+\frac{1}{\delta}\left(\frac{1}{V_k}\right)^{m_k-1}\ \sum_{i=1}^{m_k} P(\mathbf{y}_k(i) \mid X_k)\ \kappa\ P(\theta_k(i)|m_k) P(m_k \mid Y^{k-1}, m^{k-1})$$

The term $\kappa = P(X_k \mid Y^{k-1}, m^{k-1})$ is the *predicted density* in Eq.(5.23) and its combination with the *likelihood* of the ith measurement being object-originated $P(\mathbf{y}_k(i) \mid X_k)$ can be re-written as:

$$P(\mathbf{y}_k(i) \mid X_k)\ \kappa =$$

$$P(\mathbf{y}_k(i) \mid X_k) \cdot P(X_k \mid Y^{k-1}, m^{k-1}) = P(\mathbf{y}_k(i) \mid X_k) \cdot \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) =$$

$$= P_G^{-1}\ \mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) =$$

$$= P_G^{-1}\ \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k) \frac{\mathcal{N}(\mathbf{y}_k(i);\ \mathbf{H}\mathbf{x}_k, \mathbf{R}_k)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})}{\mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k)} =$$

$$= P_G^{-1}\ \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}^i_{k|k-1}, \mathbf{P}^i_{k|k-1})$$

where, to simplify, we have multiplied and divided by $\mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k)$, the term that we assigned as `TERM7` in Eq.(5.28).
Therefore $P(\mathbf{y}_k(i) \mid X_k)\ \kappa = P_G^{-1}\ \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}^i_{k|k-1}, \mathbf{P}^i_{k|k-1})$.

Next we go back to the *posterior* equation and substitute:

$$\begin{aligned} P(X_k \mid Y^k,\ m^k) = \\ \frac{1}{\delta}\left(\frac{1}{V_k}\right)^{m_k}\ P(\theta_k(0)|m_k) P(m_k \mid Y^{k-1}, m^{k-1}) \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \\ + \frac{1}{\delta}\left(\frac{1}{V_k}\right)^{m_k-1}\ P_G^{-1}\ \sum_{i=1}^{m_k} P(\theta_k(i)|m_k)\ \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k) \\ \times\ P(m_k \mid Y^{k-1}, m^{k-1})\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}^i_{k|k-1}, \mathbf{P}^i_{k|k-1}) \end{aligned} \tag{5.30}$$

Finally, we do the last set of substitutions below, in order to obtain the final expression of the *posterior* in Eq.(5.31).

As we saw in Eq.(5.29), the normalization factor $\delta = P(Y_k, m_k \mid Y^{k-1},\ m^{k-1})$ could be written as a sum of two terms. Instead of `TERM4` and `TERM7`, we now assign them as follows, since we want to represent the normalization factor δ in the form:

$$\delta = a_0 + a_1 \sum_{i=1}^{m_k} e_i$$

$$\begin{cases} a_0 = \dfrac{P(m_k \mid Y^{k-1}, m^{k-1})\ P(\theta_k(0)|m_k)}{V_k^{m_k}} \\ a_1 = \dfrac{P(m_k \mid Y^{k-1}, m^{k-1})}{P_G V_k^{m_k-1}\sqrt{2\pi \mathbf{S}_k}} \\ e_i = \exp\{-\dfrac{1}{2}[\mathbf{y}_k(i) - \hat{\mathbf{y}}_k]^T\ \mathbf{S}_k^{-1}\ [\mathbf{y}_k(i) - \hat{\mathbf{y}}_k]\} P(\theta_k(i)|m_k) \\ b_k = \dfrac{a_0}{a_1} = P(\theta_k(0)|m_k)\dfrac{P_G|2\pi\mathbf{S}_k|^{\frac{1}{2}}}{V_k} \end{cases}$$

Therefore the terms in the *posterior* probability $P(X_k \mid Y^k,\ m^k)$:

$$\frac{1}{P(Y_k, m_k|Y^{k-1}, m^{k-1})} \left(\frac{1}{V_k}\right)^{m_k} P(\theta_k(0)|m_k) P(m_k \mid Y^{k-1}, m^{k-1}) \Rightarrow \beta_k(0)$$
$$\frac{1}{P(Y_k, m_k|Y^{k-1}, m^{k-1})} \left(\frac{1}{V_k}\right)^{m_k-1} P_G^{-1}\ P(\theta_k(i)|m_k)\ \mathcal{N}(\mathbf{y}_k(i);\ \hat{\mathbf{y}}_k, \mathbf{S}_k) \Rightarrow \beta_k(i)$$

$$\beta_k(0) = \frac{a_0}{a_0 + a_1 \sum_{i=1}^{m_k} e_i} = \frac{b_k}{b_k + \sum_{i=1}^{m_k} e_i}$$
$$\beta_k(i) = \frac{a_1 e_i}{a_0 + a_1 \sum_{i=1}^{m_k} e_i} = \frac{e_i}{b_k + \sum_{i=1}^{m_k} e_i}$$

Finally, the *posterior* probability density function is:

$$\begin{aligned} P(X_k \mid Y^k,\ m^k) = \beta_k(0)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \\ + \sum_{i=1}^{m_k} \beta_k(i)\ \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}^i_{k|k-1}, \mathbf{P}^i_{k|k-1}) \end{aligned} \tag{5.31}$$

After this chapter describing the relevant Bayesian estimation calculations that we use in our study we are able to combine it with the complete description of the proposed multi-probabilistic null-finding algorithm we designed to determine the first-order boundaries that we provided in Chapter 4. This combination to form the final algorithm for finding the first-order Bragg boundaries is presented in Chapter 6 and Chapter 7.

The application of the Bayesian tracking in range and time to our observational radar cross-spectra data, which is the Step 10 in Algorithm 19, will be introduced first in Chapter 6. Included in Section 6.4, Algorithm 21 focuses on obtaining the *initial* first-order boundaries which could be multiple. Later on in Chapter 7, the Probabilistic Data Association recursion Algorithm 32 together with the Range Tracking Algorithm 33 and the Temporal Tracking Algorithm 34 (both presented in Section 7.5) will describe how do we study those multiple options and come up with the *final* first-order boundaries, which are the ultimate objective of this research.

Chapter 6

Combined Algorithm for the Alternative Method in Python

While the combined algorithm was described in Chapter 6 and illustrated in the block diagram of Figure 6.1, Chapter 5 developed the mathematical basis for various steps in the algorithm. The project has been developed in Python, and in this chapter we will be providing Python pieces of code and pseudo-code descriptions to analyze the most important steps developed during our research.

First we discuss the reasons to choose the open-source programming language Python for our project (Section 6.1), and the initial selection of the radar site within the HF radar network as well as which set of cross-spectra files are we going to study (Section 6.2).

Next we present the *conventional* method steps in Section 6.3, which refers to the algorithms we introduced in Section 4.10. Then in Section 6.4, right after, we describe the steps to run the *alternative* "multi-probabilistic null-finding method" presented in Chapter 4.

The output of the algorithms presented in Section 6.4 are the "initial estimates" of the first-order boundaries. With those, we apply our Bayesian estimation techniques described in Chapter 5 in order to achieve the final decisions on first-order line boundary locations. In Section 6.5, we introduce the algorithms to process these "initial estimates" and from there the reader is referred to the full details on the Bayesian tracking in Chapter 7.

Later, once we have obtained the "final estimates" of the first-order boundaries we can proceed using the steps below:

1. Generate the corresponding radial current vector maps, as described in Section 6.6, where we see the status of the current vector fields in the vicinity of our radar site.

2. Study the validity of these radial vectors and compare them against the boundaries obtained through the *conventional* method. These statistical comparisons are executed as described in Section 6.7, and the reader is referred to Chapter 8 for complete details.

3. Perform drifter comparisons to see if indeed our method of determining the Bragg first-order boundaries tends to add radials with Root Mean Squared errors (see Section 8.1) much smaller than the RMS errors of the radials that were determined by the conventional method but are rejected by the proposed method. Section 6.8 introduces the comparisons that are described in detail in Chapter 9.

   The complete algorithm is shown in functional block diagram form in Figure 6.1.

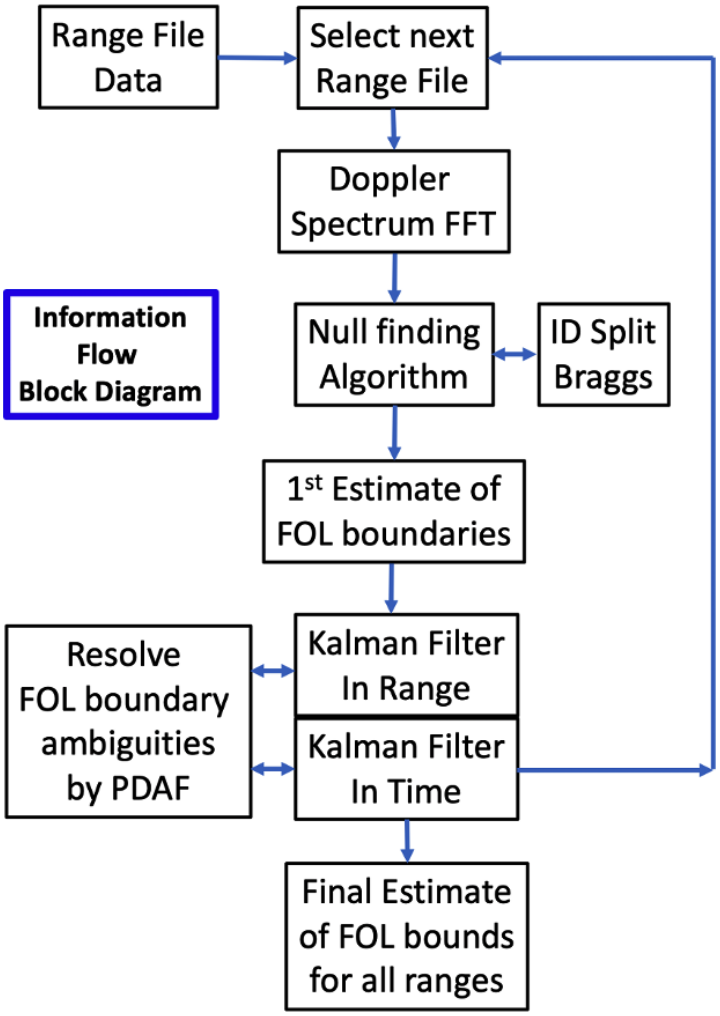

Figure 6.1: Functional block diagram describing the information flow in the new alternative line boundary algorithm. Note the iterative steps through a Probabilistic Data Association Filter that are recursively performed in order to determine the final estimates of the first-order boundary locations.

6.1 Choosing Python

Python is an open-source programming language widely used for scientific computing in both academic and multiple industry fields. Python programming is in great demand for Big Data and supports multiple platforms. It can run on Windows, Mac and Linux environments, and it involves fewer lines of codes as compared to other languages available for programming. Python has scripting features and in addition it uses many advanced libraries such as NumPy, SciPy, and Matplotlib which makes it extremely useful for data science.

Python is an excellent tool and a perfect fit as a python – big data combination for data analysis for the following reasons[60]:

- Open-source: Python is an open-source programming language developed using a community-based model.
- Library support: Python consists of a large number of well-tested analytics libraries which include packages like numerical computing, data analysis, machine learning, statistical analysis and visualization.
- Speed: As a high-level language, it has many benefits which accelerate the code development. It enables prototyping ideas which makes coding fast while maintaining the great transparency between the code and its execution.
- Object-oriented: Python is an object-oriented programming language which also supports advanced data structures such as lists, sets, tuples, dictionaries and many more. It supports many scientific computing operations like matrix operations, data frames, etc. These abilities within Python enhance its scope to simplify and speed up data operations.

- Data processing support: Provides advanced support for image and voice data due to its inbuilt features of supporting data processing for unstructured and unconventional data which is a common need in big data.

6.2 Step I: Radar Site and Cross-Spectra

The first phase of our method is to choose a radar site within the HF radar network and to decide which point in time and the associated cross-spectra files we are going to study. The SeaSonde HF radars we have been working with during this research operate at the following frequencies:

- 5 MHz (Long-Range): The lowest of the SeaSonde transmit frequency bands is within 4.4 and 6 MHz. This allows for greater ranges than the higher frequency bands without any increase in radiated power because the propagation loss is less than at higher frequencies[58][59]. In general, the average daytime observable range achieved by a SeaSonde operating at this long-range mode is typically 150-220 km and their range resolution $\sim$6 km.

- 13 MHz (Mid-Range): The SeaSonde medium range configuration works with a transmit frequency band around 13 MHz. These mid-range HF radars can reach typically 60 to 90 km in range from the radar location with the trade-off of having a better range resolution (between 1 and 3 km) as compared to the long-range radars.

- 24 MHz (High-Resolution): The short range HF radars work at an operating frequency around 25 MHz have the shortest range of 20 to 45 km. However, they present a great range resolution of 0.5 km and a convenient 2.4 cm/s for radial current resolution.

SeaSonde Site	Operating Frequency [MHz]	λ Radar [m]	Range [km]	Range resolution [km]	Doppler bins	Time output [minutes]	Radial current resolution [cm/s]
HATY	4.5	66	160-220	5.8	1024	30 - 60	3.2
ASBI	13.2	23	60-90	3	512	10 - 15	4.4
VCOL	24.4	12.3	20-45	0.5	1024	10 - 15	2.4

Table 6.1: Comparison of the specifications between the three operating frequencies of the SeaSonde sites that we use in this study.

Depending on the operating frequency of the HF radar of choice, we will have different specifications, summarized in Table 6.1. Generally, the most important factor in the selection of the operating frequency is the desired radar range for the specific site or application needs, as presented in Figure 6.2.

Next, we have to choose the time frame that we will be studying. The decision on which days to be analyzed can depend on the study requirements, the time of interest due to an specific event happening, e.g., tidal cycles, a storm, coastal upwelling, etc.

Once we have selected the radar site and the set of cross-spectra files that we will be processing, we need to locate the files described in the list below. The parameters in `Header.txt`, `AnalysisOptions.txt` and `FirstOrderConfigs.plist` may vary depending on the season, the tide, the weather predictions etc. and consequently we can store different versions and the operator should be able to choose which configuration to use.

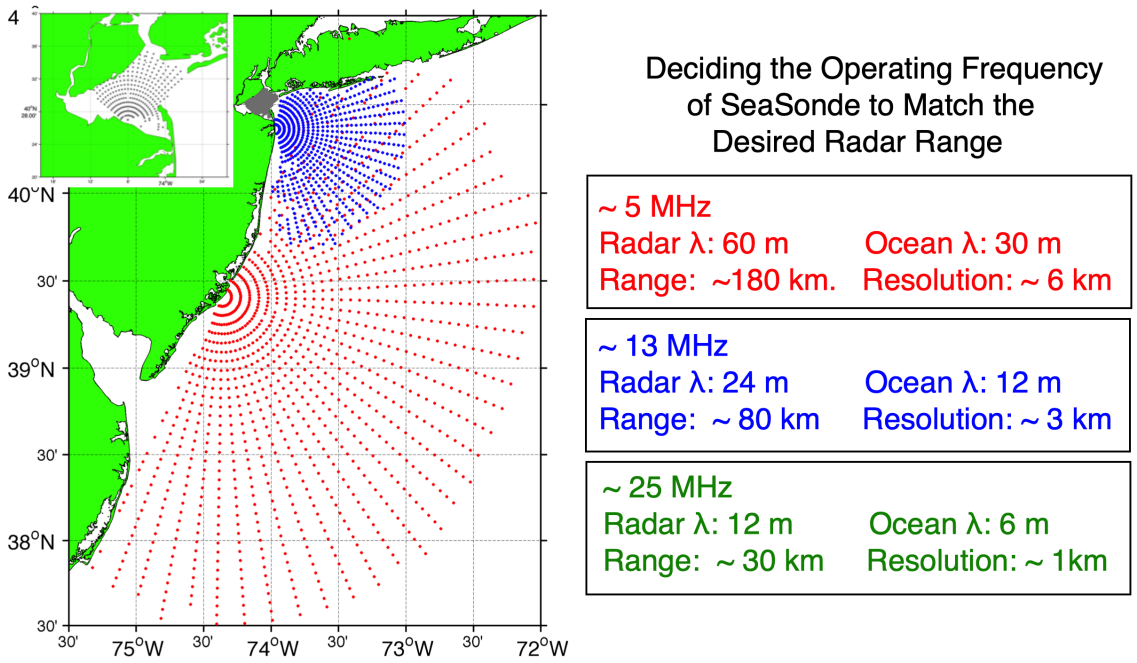

Figure 6.2: Selection of the radar's required frequency of operation depending on the desired range for our application. The possible operating frequencies for SeaSonde HF radars used in this example are ∼5 MHz, ∼13MHz and ∼25MHz. The coverage maps, credited to IOOS – the U.S. Integrated Ocean Observing System[29], correspond to an example of three SeaSonde radar sites located at the East Coast of the United States.

List of configuration parameters:

1. `Header.txt`

   In the "Header" text file we find a series of parameters specific to the corresponding radar site. These parameters control aspects that are used in the conversion from Cross-Spectra into Radial Maps and Wave Data. It is important to verify when this Header was last modified and if it needs to be updated or any changes have to be made for the present scenario. The parameters that are most relevant to our research are defined in Chapter 4 when we describe the null finding algorithm.

2. `AnalysisOptions.txt`

   In the "Analysis Options" text file we find some processing settings needed when turning the cross spectra files into radial current maps. One of the parameters is the Antenna Pattern Type: if 0 (zero) it specifies to use an ideal antenna pattern in processing cross spectra into radials, while a value 1 (one) specifies to use a measured antenna pattern in processing. Another important setting is the Doppler Interpolation parameter. If set to 2 (two), it tells the `SpectraToRadial` tool to internally double the number of Doppler points (using linear interpolation) and then calculate radial velocity and bearing. This results in a typical 10 to 15 percent increase in the number of vectors. The theory behind this is that real world signals do not typically fall exactly onto a Doppler cell and instead will split power between two Doppler cells; so this technique helps the software better discern between adjoining vectors.

3. `MeasPattern.txt` and `IdealPattern.txt`

   These files define the antenna pattern used to resolve radial bearings and therefore place them in the radial maps. `IdealPattern.txt` is a theoretical ideal pattern, while `MeasPattern.txt` is a Radial Site specific measured pattern that involves measuring the antennas and doing an Antenna Pattern Measurement. (More information can be found on the "Antenna Pattern Measure Guide" by CODAR Ocean Sensors Ltd.). The pattern bearings are CCW (counter-clockwise) degrees referenced from the antenna bearing, which is found in the `Header.txt` and is CW (clockwise) degrees from true North.

4. `FirstOrderConfigs.plist`

   The "First Order Configurations" preference file contains a series of parameters, described in Chapter 4, that configure the settings for the alternative null finding algorithm (Algorithm 16), the split Bragg recursive search (Algorithm 17) as well as the Bayesian estimation application in Algorithms 28 through 30.

As we discussed in Chapter 2 and Chapter 3, we study the backscatter spectra of the monopole antenna (Antenna 3) to extract the first-order region boundaries. The traditional method for determining the first-order Bragg boundaries is known as the "conventional" method in this study, which we detailed in Section 4.10. We are interested in comparing the aforementioned boundaries against the first-order boundaries obtained using the "alternative" multi-probabilistic null finding method that we designed – described in Section 4.11 – which first finds the local minima (*nulls*) and then applies a multi-probabilistic approach (see Figure 6.1) to obtain the final decision on the locations of the first-order boundaries.

6.3 Step II: Conventional First-Order Boundaries

In order to perform the comparisons we will have to extract the conventional first-order boundaries from the cross-spectra files; we are able to do so if the version of the cross-spectra file is 6 or newer. However, if we are working with older versions and we still want to be able to run our comparisons we will need to "generate" the conventional first-order boundaries by emulating the basic null finding described in Chapter 4 Algorithm 18.

The pseudo-code in Algorithm 20 describes how to extract the conventional first-order boundaries (labeled as *conventional_FirstOrderBoundaries*) if they are available and stored under the label `['FOLS']` in the cross-spectra file, or how to generate them using the `ConventionalFOL` algorithm mentioned above when they have not been predetermined in the file.

Algorithm 20 Retrieve Conventional First-Order Boundaries

Load cross-spectra file **as** `cs_file`
version ← `cs_file.version`
if *version* ≥ 6 **then**
  Extract the `FOLS` field from `cs_file`
  FirstOrderBoundaries ← `cs_file.FOLS`
else
  Run `ConventionalFOL` algorithm to generate the first-order boundaries
  FirstOrderBoundaries ← `ConventionalFOL(cs_file)`
end if
Return *conventional_FirstOrderBoundaries*

6.4 Step III: Alternative First-Order Boundaries

The alternative method to determine the first-order Bragg boundaries, also known as the "multi-probabilistic null-finding method" in this study, consists of many steps which are all presented in Chapter 4, specifically in Section 4.11 and illustrated in Figure 6.1. As a reminder, its two main innovative features are:

1. The null-finding algorithm allows output sets of multiple *initial* first-order boundary estimates for each Bragg region boundary and range.

2. Our newly designed algorithm determines the locations of the *final* unique first-order boundaries after applying two Bayesian tracking algorithms (both spatial and temporal) that improve the *initial* estimates using spatial and temporal models.

The pseudo-code below, Algorithm 21, summarizes our process to determine what we denominate as *initial_FirstOrderBoundaries*, i.e. the "first estimate of the first-order boundaries". As we describe next, the determination of every single boundary at a single range reported at *initial_FirstOrderBoundaries* can consist of:

- A) a single Doppler bin being the most probable location and therefore our first-order boundary determination,

- B) a set of possible locations, to which we will later apply the Bayesian estimation techniques discussed in Chapter 7, to obtain the final boundaries namely *final_FirstOrderBoundaries*, as described in Section 7.5.

Initially, as we see in Algorithm 21, we run the `AlternativeFOL` method (Section 4.11). We treat each of the four boundaries as a separate problem and thus we obtain four independent sets of boundaries following Section 5.1.4.

In addition, we also run the recursive Split Bragg Search Algorithm that we introduced in Section 4.9 and that we label as `BraggSplitAlgorithm` in Algorithm 21. The complete Bragg Split Test is described in Algorithm 17 and will try and find if any of the two Bragg regions has been divided into two or more sections, returning all the corresponding boundary locations for each of the sections (see Figure 4.9).

On top of that, a specific first-order boundary for a specific range, even when the Bragg region does not present a split, might actually output multiple possible locations as we listed under Option B above. This could happen when the region around the boundary contains a lot of noise, or if there is more than one null nearby and therefore the null-finding algorithm is not sure which Doppler bin should be considered as the most appropriate location for the first-order boundary.

Algorithm 21 focuses on obtaining the *initial_FirstOrderBoundaries* which could be multiple. Next, the tracking in range and time presented in Section 6.5 – which refer to the Range Tracking Algorithm 33 and the Temporal Tracking Algorithm 34 that will be described in Chapter 7 – will cover how we analyze the multiple *initial* boundary options and come up with the final decisions (namely *final_FirstOrderBoundaries*), which are the ultimate objective of this research.

Algorithm 21 Initial 'Alternative' Estimations of the First-Order Boundaries

Load cross-spectra file **as** `cs_file`
Run the multi-probabilistic estimation method `AlternativeFOL`
(Algorithm 19) to determine the first-order boundaries
while running `AlternativeFOL` **do**
  for each BOUNDARY in $[FOL1, FOL2, FOL3, FOL4]$ **do**
    for each RANGE in `cs_file` **do**
      1) Search for split Braggs
      if split Bragg **detected then**
        Run `BraggSplitAlgorithm`
      end if
      2) Check for multiple possible locations
      if various probable locations **detected then**
        Return set of multiple Doppler locations into
        corresponding BOUNDARY and RANGE
      end if
    end for
  end for
  $[FOL1, FOL2, FOL3, FOL4] \leftarrow$ `AlternativeFOL(cs_file)`
end while
$FirstOrderBoundaries \leftarrow [FOL1, FOL2, FOL3, FOL4]$
Return *FirstOrderBoundaries* **as** *initial_FirstOrderBoundaries*

6.5 Step IV: Tracking in Range and in Time

We saw in Algorithm 21 how *initial_FirstOrderBoundaries* might consist of some sets of multiple possible Doppler bin locations assigned to a single first-order boundary at a specific range. As we described in Section 5.4, we have developed two possible solutions that will determine the final *final_FirstOrderBoundaries* given the aforementioned multiple initial first-order boundary estimates:

1. NNF: We can choose the Nearest Neighbor Filter approach to select, between the existing possible boundary locations, which one is most probable. The NNF filter is an approximation of the optimal Bayesian filter recursive solution for multiple simultaneous measurements and to choose the most probable location NNF will search for the measurement whose statistical distance to the predicted measurement of the filter is the smallest.
   (Described in Section 5.4.2)

2. PDAF: We can choose the Probabilistic Data Association Filter approach to estimate a first-order boundary location by assigning probabilities (weights) to the existing possible boundary locations and finally determining a location that could be anywhere in between all of these possible locations.
   (Described in Section 5.4.3)

Since the Bayesian estimation algorithms and their application to tracking the first-order boundaries is such a relevant step in this research, we decided to describe them in Chapter 7, where we study their temporal and spatial tracking in detail. As a reminder, the outputs of this Bayesian tracking correction will be the sets of "final estimations of boundary locations" that are used in the comparison algorithms presented in the next sections.

6.6 Step V: Radial Vector Maps

Once we have obtained the "final estimates" of the first-order boundaries using our alternative proposed method, we start the process of comparing those solutions against the *conventional* first-order boundary estimates introduced in Chapter 4 and Algorithm 18 and extracted as described in Section 6.3. Note that we perform this comparison to verify how our multi-probabilistic method improves the radial vector current maps by removing erroneous vectors and adding useful new vectors.

The first step within those comparisons is to generate the corresponding **radial current vector maps** (Algorithm 22) where we are able to observe the status of the current vector fields in the vicinity of our radar site. Many examples of radial current velocity vector maps are presented in this document, e.g. the radial map corresponding to the Bodega Bay SeaSonde site in Northern California in Figure 2.9 and the radial maps for the VCOL site in the Straits of Georgia in Figure 8.8, Figure 8.9 and Figure 8.11.

Algorithm 22 Generating Radial Vector Maps

I. Load *conventional_FirstOrderBoundaries* as in Algorithm 20.

II. Load *proposed_FirstOrderBoundaries* after running the corresponding tracking, either NNF or PDAF, as described in Section 6.5.

III. Run MUSIC algorithm to re-process the cross-spectra files in our set. As shown in Figure 6.3 we obtain the set of corresponding radial files (`.ruv` files) that contain the list of radial current velocities for each detectable bearing and range arc progressing outwards from the SeaSonde receive antenna.

IV. Generate the corresponding radial current vector maps by placing the radial vectors given by the radial files.

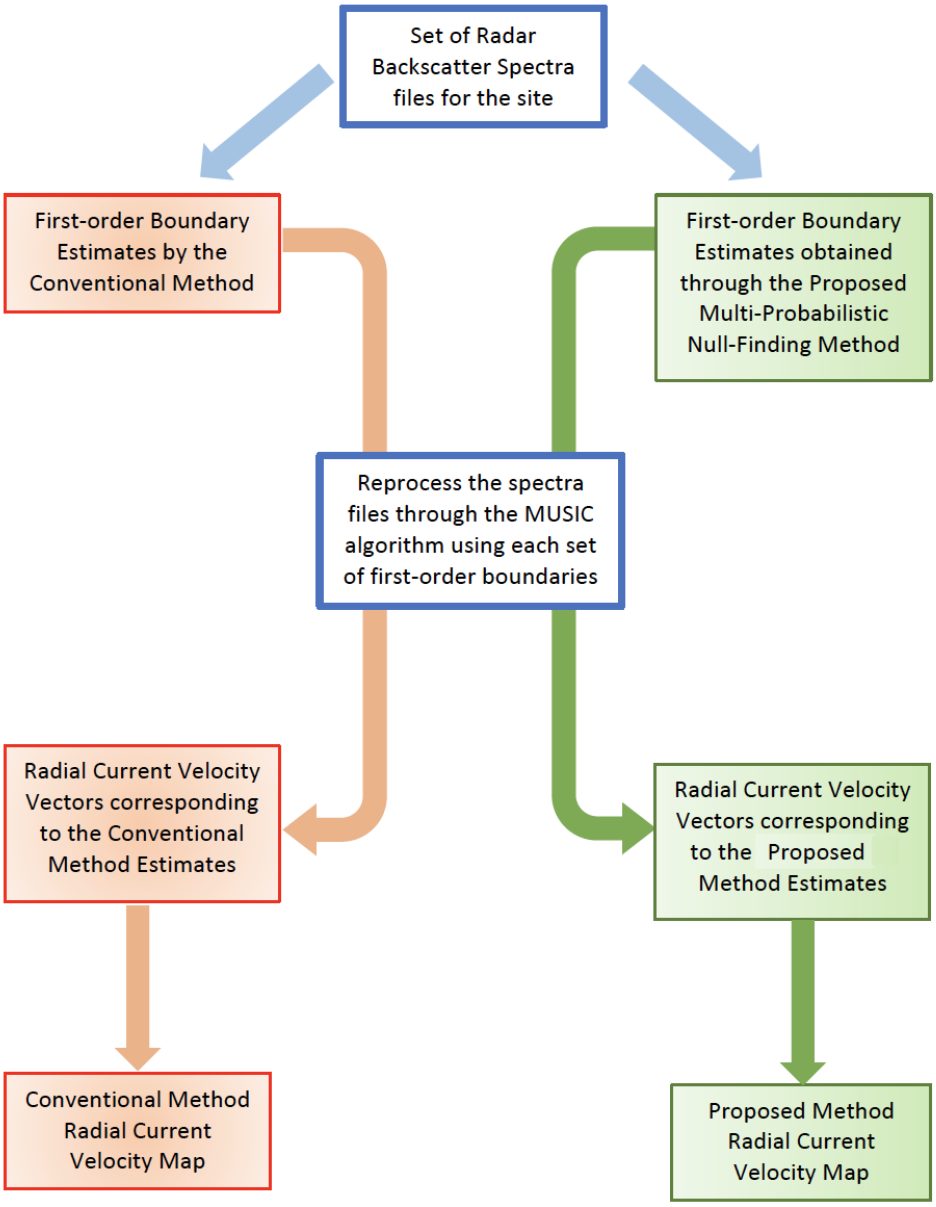

Figure 6.3: Functional block diagram showing the process between having the set of spectra files and the first-order estimates from the two methods into re-processing through the MUSIC algorithm (Section 2.3.2 and Figure 2.10) and thus obtaining the resulting sets of radial current velocity vectors and their corresponding radial current velocity maps.

The flow diagram in Figure 6.3 shows the process between the set of spectra files and the first-order estimates from the two methods into the resulting sets of radial current velocity vectors and their corresponding radial current velocity maps.

6.7 Step VI: Radial Vectors Comparisons

Once we have both sets of `ruv` radial files containing all the determined radial current vectors corresponding to 1) the *conventional* method, and 2) the *proposed* multi-probabilistic method, we are able to compare those vectors within the radial maps, following Algorithm 23. In addition, we run a series of statistical comparisons to study the validity of the proposed method radial vectors, described in detail in Chapter 8. As shown in that same chapter, we are in fact able to demonstrate how the use of Bayesian estimation techniques in the determination algorithm properly tracks and corrects the first-order boundaries, improving the overall accuracy of the radial current vectors by:

- adding radial vectors that statistically compare well with the surrounding radial vectors in the area,
- while rejecting radial vectors whose standard deviation from the surrounding ones is indeed higher than the added vectors.

Algorithm 23 Comparing Between Radial Vector Sets

I. Load *conventional_Radials*, the radial current vectors associated with the *conventional_FirstOrderBoundaries*.

II. Load *proposed_Radials*, the radial current vectors corresponding to the *proposed_FirstOrderBoundaries* after running the corresponding tracking, either NNF or PDAF, as described in Section 6.5.

III. Divide the two sets of radial vectors into:

   (a) **Agreed Radials**: Subset of radial vectors that are found by both the *conventional* first-order boundaries method as well as by the *proposed* multi-probabilistic null-finding first-order boundaries determination method.

   (b) **Added Radials**: Subset of radial vectors that are added by the *proposed* multi-probabilistic null-finding first-order boundaries determination method.

   (c) **Rejected Radials**: Subset of radial vectors that were given by the *conventional* first-order boundaries method but rejected by the *proposed* multi-probabilistic null-finding first-order boundaries determination method.

IV. Run the statistical comparisons described in Chapter 8 to demonstrate the advantages of the *proposed method*.

V. Generate comparison figures like Figure 8.10.

6.8 Step VII: Comparisons with Drifters

The last step in our series of comparisons (Algorithm 24) is the comparison against radial currents obtained by drifters in the Straits of Georgia, described in Chapter 9. Our purpose in doing a drifter comparisons is to perform a sanity check to see if indeed our method of determining the Bragg first-order boundaries tends to add radials with Root Mean Squared errors (see Section 8.1) much smaller than the RMS errors of the radials that were determined by the *conventional* method but are rejected by the *proposed* method.

Algorithm 24 Comparing Between Radial Vector Sets and Drifter Radial Vectors

---

1. Load Drifter data nearby the radar site and transform the drifter velocities into radial velocities respect to our radar site (Section 9.3), saved as *drifter_Radials*.
2. **Run comparisons** (Section 9.3) between the drifter radials and the three subsets of radial currents described in Algorithm 23.
3. Generate **Standard Deviation Comparison Figures** of radar radials versus drifter radials, as seen in Figure 9.7.

Chapter 7

Bayesian Estimation Algorithms and their Application to Tracking the First-Order Boundaries in Time and Range

In this chapter we study the application of the Bayesian estimation and tracking techniques that we first introduced in Chapter 5.

In particular, in Section 7.1 we first present the algorithm to load the cross-spectra radar files and select the necessary configuration parameters to create the desired estimation model, with the particular settings that we need for the current time, radar site and weather conditions. The relevant parameters for setting up the model can be either modified depending on our scenario, or multiple configurations can be stored. The parameters are saved at the `Header.txt` file, as well as in the `FirstOrderConfigs.plist` file as described in Section 6.2.

Next, in Section 7.2 we introduce the complete recursive Bayesian estimation cycle with all its inputs and its estimation output. More information about designing Python algorithms for Bayesian estimation is found in Labbe[38].
The recursion features two paramount steps that we introduced in Section 5.2.3:

1. The `predict` stage: Propagates the probability density function of the estimate forward, by combining the previous prediction and the current behavior of the estimation model as described in the algorithms under Section 7.3. In other words, we are translating, deforming and spreading the probability density function.

2. The `update` stage: Presented in Section 7.4, this step uses the Bayes theorem to modify the predicted probability density function based on the newly received measurement.

Recall that the four first-order boundaries we desire to determine as four separate estimation problems are:

- Left boundary of the Receding Bragg wave
- Right boundary of the Receding Bragg wave
- Left boundary of the Advancing Bragg wave
- Right boundary of the Advancing Bragg wave

As we discussed in Section 5.4 "Solutions for Object Tracking with Simultaneous Measurements", we might have tracked multiple simultaneous possible measurements for a single boundary location in range e.g. due to measurements in clutter or the presence of split Braggs.

Consequently, in Section 7.5 we address the question of how to proceed when we receive multiple boundary possible estimates for any or all of the four estimation problems above, i.e. we describe the "data association step". In our scenario, the sensor measurements for each of the four boundaries under study are received intrinsically separated, thus can skip the step of "gating" the measurements (Step 3 in the list below), which would take place as soon as the observations are received. Thus, after the prediction step, if multiple observations are received, we need to apply the data association step before updating the state of our estimation.

Finally, in Section 7.5.1 and Section 7.5.2 we focus on the temporal and spatial tracking that we apply to our multiple simultaneous measurements in the data association step and describe why they improve our final estimations of the first-order boundaries. This two-dimensional tracking in range and in time is a novel application of Bayesian tracking for this first-order boundaries determination problem.

Overview of the Estimation Steps:

1. Set up the estimation model.
2. Prediction.
3. Gating: Not necessary for our problem since the measurements for every boundary are received separately.
4. Data Association: Step that will determine which of the received observations is the best match, and therefore its value will be used on the Update step.
5. Update.

7.1 Setting up the Estimation Models

We start with Algorithm 25 below, which includes the steps to load the cross-spectra and select the necessary configuration parameters to create the desired estimation model adequate for the current time, radar site and weather conditions. Please refer to Chapter 5 for details about the formulation of the estimation models as well as their behavior and their recursive updates.
The relevant Bayesian estimation parameters mentioned in this chapter were introduced in Section 4.1 and they all need to be set and stored in the `FirstOrder Configs.plist` file.

Firstly, for the **initial estimation model** we configure the parameters below, which will be the inputs into the first recursion cycle in Algorithm 26 described next. The units for this parameters are "Doppler bin" as we saw in Figure 3.5.

- **Initial State** mean value and covariance (`state_0` and `state_cov`): the estimated *initial state* value for the current *first-order boundary* and the covariance (uncertainty) of this value.

- **Initial Model** mean value and its covariance (`model_0` and `model_cov`): the model (function describing the estimated initial behavior of the variable) and the uncertainty associated with this model.

It is important to notice that in every cycle of the recursion, including the initial recursion at time $t = 0$, the **mean value of the state variable** (first-order boundary location), its **variance** and the **model behavior and its variance** are all **updated**. For example, the model behavior variable `model_k` (which is initially set as `model_k = model_0`) changes with each time t and describes our expected change in the state variable between iterations.

If `model_0` = 1, the model predicts that the Doppler location of the first-order boundary should not vary, what we describe as a *smooth boundaries* model. But if we keep receiving measurements that are far away from the previous state value, then eventually, following the Bayesian recursion in Chapter 5, the model behavior at time $t = k$ (`model_k`) will end up updating and deviating from its initial value.

Secondly we set the parameters below for both the `Time_Tracking` algorithm and the `Range_Tracking` algorithm, since they are not necessarily the same.

- **New Measurement Covariance** (stored as `m_cov_0`) is the uncertainty initially associated with the measurements that we receive.
- **State to Measurement** (`state2measurement`) is the function describing the transition between the state of the variable and the measurement[38]. When the measurement describes the state in the same two-dimensional space (like in our problem where no transition is needed) then this parameter `state2measurement`=1.
- **Mode Selection**:

  The `Time_Tracking` correction will always be performed forward in time, in other words: we correct the present boundaries including information about previous time boundaries. Hence, the **mode selection** setting applies only to `Range_Tracking`. The value of `range_tracking_mode` can be:

  - 1 → Single range tracking inwards towards the radar,
  - 2 → Range tracking inwards and then back out,
  - 3 → Single range tracking outwards in range,
  - 4 → Range tracking outwards and then back towards the radar.

Initial Bayesian estimation cycle at *t=0*:

After the estimation model is set, we start with the first estimation cycle: the initial recursion at time $t = 0$. We find the steps in Algorithm 26 below. As we present in Figure 7.1, the inputs into the first recursion step are the initial estimates of the first-order boundary and the model behavior at $t = 0$, while the output is $\hat{\mathbf{x}}_0$, the final estimation of the first-order boundary at $t = 0$.

Algorithm 25 Loading Cross-Spectra and Creating the Estimation Model

Load the monopole antenna cross-spectra file as `cs_file`
`radar_site`: the settings of the radar site given by the `cs_file`
`n_ranges`: the number of range bins at the corresponding `radar_site`

for each `BOUNDARY` **in** [*FOL1, FOL2, FOL3, FOL4*] **define:**
`state_0`: the estimated *initial state* value for that *first-order boundary*
`state_cov`: the covariance (uncertainty) of the *initial state* value
`model_0`: the *initial model behavior* that we expect on the boundary
`model_cov`: the covariance (uncertainty) of the *initial model*
end for

Algorithm 26 Bayesian Estimation (recursion at time $t = 0$)

Inputs:
`state_0` and `model_0`: initial estimates of the first-order boundary location and the model behavior at time $t = 0$
Predict:
$\hat{\mathbf{x}}_0 \leftarrow$ `predict(state_0, model_0)`
Output:
$\hat{\mathbf{x}}_0 \rightarrow$ First-order boundary estimation at time $t = 0$

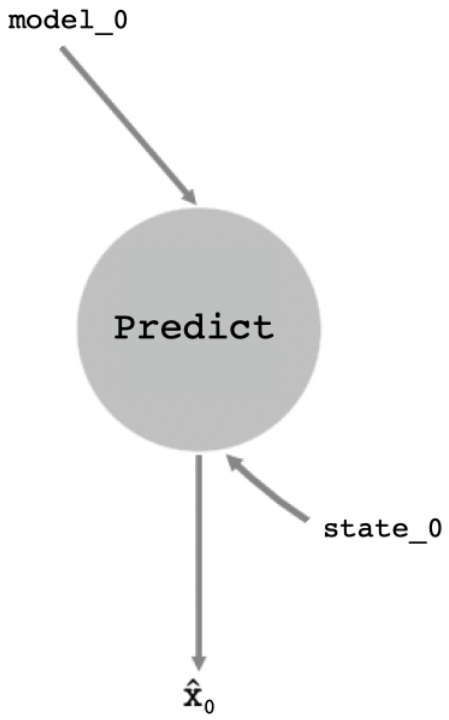

Figure 7.1: Graphic description of Algorithm 26. Inputs are the initial estimates of the first-order boundary and model behavior at $t = 0$, while the output is $\hat{\mathbf{x}}_0$, the final estimation of the first-order boundary at $t = 0$.

Next, Algorithm 27 describes how we extract each of the sets of simultaneous measurements that might be associated with each of the four boundaries. The input for Algorithm 27 is `m_k` (shown in Figure 7.2) and it corresponds to the set of *measured first-order boundaries* received at time k for the specific boundary and range that we are working with. The complete proposed multi-probabilistic null-finding algorithm that generates each of the `m_k` measurement sets can be found in Section 4.11.

Algorithm 27 First-Order Boundary Measurements at time $t = k$

for each `BOUNDARY` **in** [*FOL1, FOL2, FOL3, FOL4*]:
  for each `RANGE` **in** `n_ranges` **do**
    Load `m_k[BOUNDARY,RANGE]` from the `cs_file`

      `m_k` is the set of *measured first-order boundaries* received at time k for that specific boundary and range.

      At any given time k, the set `m_k` can be a single first-order boundary location or a number of multiple locations.

  end for
end for

7.2 The Recursive Bayesian Estimation Cycle

`m_k`, the sets of *measured first-order boundaries* received at time $t = k$ for each specific boundary and range, are the first inputs into Algorithm 28, described in Figure 7.2. As a reminder, *measured first-order boundaries* are the alternative first-order boundaries, described in Section 6.4 as Step III. This is the continuation of the recursive Bayesian estimation cycle, executed at any time $t = k$, where k is not the initial time $t = 0$.

Another input into the algorithm, besides the measurements `m_k`, is the variable `model_k` and it corresponds to the model behavior at time $t = k$. As we introduced in Section 7.1, the model behavior is initially set as `model_k = model_0` during Algorithm 25 and its behavior is updated on every recursion cycle following the Bayesian recursion updates described in Chapter 5 and Algorithm 28.

The last input for this recursive estimation algorithm is $\hat{\mathbf{x}}_{k|k-1}$ that becomes $\hat{\mathbf{x}}_{k-1|k-1}$, and it is the estimation from previous time $t = k - 1$ that includes all measurements up to $t = k - 1$. We observe those estimated states as well as the different inputs into the algorithm in Figure 7.2 .

The output of this Bayesian estimation recursive algorithm is the first-order boundary estimation at time $t = k$, $\hat{\mathbf{x}}_k$. We are able to estimate the output by combining the updated prediction $\hat{\mathbf{x}}_{k|k}$ with the updated model behavior `model_k`.

Finally, all the variables on this Bayesian estimation recursive algorithm that we describe in Figure 7.2 go through the two stages of the Bayesian estimation that we introduced in Section 5.2.3. This important steps are the `predict` stage and the `update` stage that we describe next.

7.3 The Predict Step

The `predict` step, described in Algorithm 29, focuses on predicting $\hat{\mathbf{x}}_{k|k-1}$. The `predict` step combines $\hat{\mathbf{x}}_{k-1|k-1}$ (the value of the first-order boundaries from the previous state) with the updated model behavior at the current time `model_k`. Running `predict` leads to $\hat{\mathbf{x}}_{k|k-1}$, which will be then passed into the `update` step as shown in Algorithm 28. This prediction at time $t = k$ is known in Chapter 5 as the *prior* or *predicted probability density function* of the object state at time k, but before receiving and processing the measurement `m_k`.

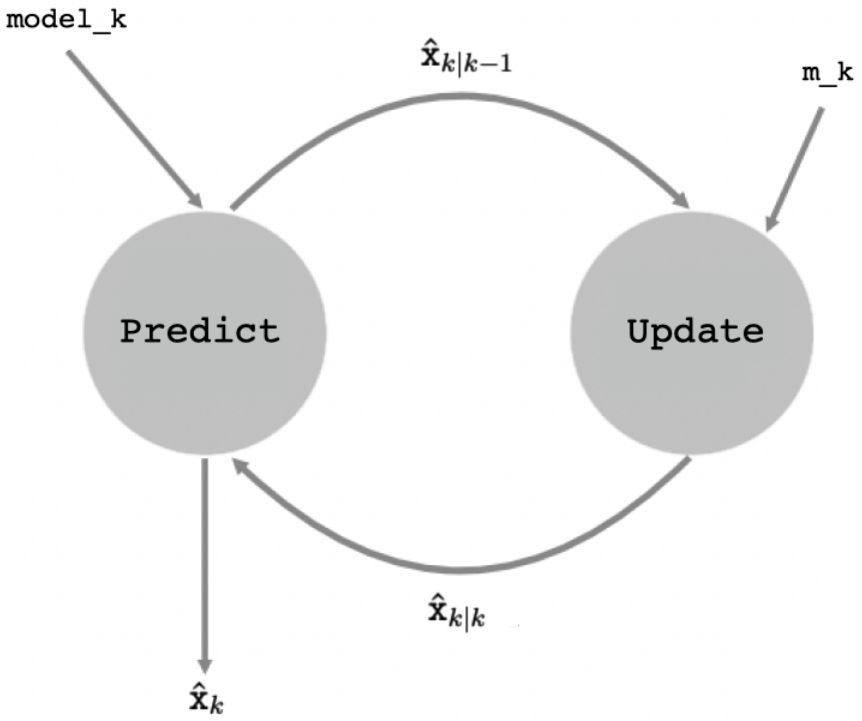

Figure 7.2: Graphic description of Algorithm 28, the recursive Bayesian estimation algorithm. Inputs are: 1) `m_k`, the sets of *measured first-order boundaries* received at time $t = k$ for each specific boundary and range; 2) `model_k`, the model behavior at time $t = k$; 3) $\hat{\mathtt{x}}_{k|k-1}$ that becomes $\hat{\mathtt{x}}_{k-1|k-1}$, and corresponds to the estimation from previous time $t = k - 1$ that includes all measurements up to $t = k - 1$. The output of this recursive Bayesian estimation algorithm is $\hat{\mathtt{x}}_k$, the final estimation of the first-order boundary at the current estimation time $t = k$.

Algorithm 28 Bayesian Estimation (Recursion at time $t = k$)

Inputs:

`m_k`: set of *measured first-order boundaries* received at time $t = k$ for that specific boundary and range.

`model_k`: model behavior at $t = k$.

$\hat{\mathtt{x}}_{k-1|k-1} \leftarrow \hat{\mathtt{x}}_{k|k-1}$: estimation from previous time $t = k - 1$ that includes all measurements up to $t = k - 1$.

Predict:

$\hat{\mathtt{x}}_{k|k-1} \leftarrow$ `predict(`$\hat{\mathtt{x}}_{k-1|k-1}$`, model_k)`: uses the previous state at time $t = k - 1$ and the updated model behavior to $t = k$ to predict the current value of the first-order boundaries.

Update:

$\hat{\mathtt{x}}_{k|k} \leftarrow$ `update(`$\hat{\mathtt{x}}_{k|k-1}$`, m_k)` : update the $\hat{\mathtt{x}}_{k|k-1}$ prediction by combining it with `m_k`, the measured state of the variable at $t = k$, which leads to the updated prediction $\hat{\mathtt{x}}_{k|k}$.

Output:

$\hat{\mathtt{x}_k} \leftarrow$ `output(`$\hat{\mathtt{x}}_{k|k}$`, model_k)`: estimate the output $\hat{\mathtt{x}_k}$ by combining the updated prediction $\hat{\mathtt{x}}_{k|k}$ with the updated model behavior `model_k`.

$\hat{\mathtt{x}}_k \rightarrow$ First-order boundary estimation at time $t = k$

Algorithm 29 Bayesian Estimation (`predict` step)

The `predict` algorithm steps are based on Equations (5.13) and (5.15), describing the *prior probability density function*.

$$P(X_k \mid Y^{k-1}) = \int_{X_{k-1}} P(X_{k-1} \mid Y^{k-1}) P(X_k \mid X_{k-1}) dX_{k-1}$$
$$= \int \mathcal{N}(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \cdot \mathcal{N}(\mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_k) = \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$$

where $P(X_k \mid Y^{k-1})$ is the output $\hat{\mathbf{x}}_{k|k-1}$, i.e. the prediction that will become the input of the next iteration. From Y^{k-1} we see that the measurements considered are only those up to time $t = k - 1$.

Therefore, the `predict` algorithm steps correspond to the following Gaussian multiplications:

$$\hat{\mathbf{x}}_{k|k-1} = \texttt{predict}(\hat{\mathbf{x}}_{k-1|k-1}\texttt{, model\_k}) =$$
$$\texttt{gaussian\_multiply(previous\_posterior, model\_transition)}$$

where the `previous_posterior` is $P(X_{k-1} \mid Y^{k-1})$ and the `model_transition` $P(X_k \mid X_{k-1})$ has been updated from the previous `model_k` through its uncertainty `model_cov`.

7.4 The Update Step

The `update` step, described in Algorithm 30, focuses on updating the prediction $\hat{\mathbf{x}}_{k|k-1}$ that we receive from the `predict` step. The `update` step combines it with the first-order boundary measurement `m_k` for that specific boundary and range that we receive at the current time $t = k$. Running `update` leads to $\hat{\mathbf{x}}_{k|k}$: the updated prediction at time $t = k$, known throughout Chapter 5 as the *posterior probability density function* given all measurements up to $t = k$.

Algorithm 30 Bayesian Estimation (`update` step)

The `update` algorithm steps follow Eq.(5.16),

$$P(X_k \mid Y^k) = \frac{\mathcal{N}(\mathbf{H}\mathbf{x}_k, \mathbf{R}_k) \cdot \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})}{\mathcal{N}(\mathbf{H}\hat{\mathbf{x}}_{k|k-1}, \mathbf{S}_k)} = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}),$$

where X_k is the prediction $\hat{\mathbf{x}}_{k|k-1}$ received from the `predict` step and the measurements up to measurement `m_k` are listed as Y^k.

Eq.(5.16) follows the Bayes theorem description of the *posterior probability density function* in Eq.(5.6),

$$\texttt{posterior} = \frac{\texttt{prior} \times \texttt{measurement}}{\texttt{normalization factor}}$$

where the actual term `likelihood` from Eq.(5.6) has been replaced by `measurement` for clarity.

Hence, the actual steps to execute in the `update` step are the following Gaussian multiplications:

$$\hat{\mathbf{x}}_{k|k} = \texttt{update}(\hat{\mathbf{x}}_{k|k-1}, \texttt{ m\_k}) =$$

```
gaussian_multiply(prior, measurement, normalization_factor)
```

7.5 The Data Association Step

In addition to the calculations presented in Chapter 4 and Chapter 6 to find the nulls and the locations of the first-order boundaries, our approach also keeps track of the spatial and temporal history of data and applies Bayesian estimation techniques to the observed radar backscatter spectra to correct and filter out erroneous observations.

Several Bayesian sequential estimation techniques were tested to improve the estimation of the first-order Bragg echo boundaries. As described in Chapter 5 we studied simple filters like g-h and Kalman as well as more advanced ones like the Nearest Neighbor and the Probabilistic Data Association filters.

In this section we describe this step known as "data association" that will be applied in two instances:

1) On the first hand, when we receive multiple detection measurements for any or each of the four first-order boundary estimations and we want to decide the most probable location for that boundary. In our scenario, the sensor measurements for each of the four boundaries under study are received intrinsically separated but we might have more than one for a single boundary location and range. Therefore in this case, after the prediction step we need to apply the data association step before updating the state of our estimation.

2) On the other hand, we also apply "data association" to the tracking in range and time that we apply to the boundaries. This Bayesian tracking is indeed the last correction applied to the boundaries which will take into consideration the values that the boundaries had in previous timestamps and in the surrounding ranges, as described in detail in Sections 7.5.1 and 7.5.2.

The data association step for a case where a set of multiple measurements are received simultaneously for the estimation of a single state variable (described in Section 5.4) is presented in Figure 7.3:

The figure shows an example when two simultaneous measurements are received. The horizontal axis are the time steps, while the vertical axis corresponds to the value of the state variable we are estimating. At the previous iteration $t = 1$ we placed the *prior estimate* at $y = 2.4$, and we update that estimate by using the `model` which gives us the *new estimate* for the present $t = 1$ around $y = 2.6$.

In addition, at the current iteration $t = 1$ we receive two *new measurements* (blue dots) and assign a particular weight to each (β in (Algorithm 32)). At the update step we combine the *new estimate* with the *new measurements* and we obtain the final *prediction* (green dot) for the current time $t = 1$ at $y = 2.8$.

As we see next, Figure 7.3 is an example of a PDAF result (Algorithm 32) because the final decision on the *prediction* is placed at a value between measurements, and not forced to be one of the measurements like an NNF approach would do.

The Nearest Neighbor Filter recursion (Algorithm 31) is a simplified filter that chooses which measurement is nearest to the predicted measurement to update the state. As opposed to "all-neighbors filters", the nearest neighbor filter can lead to very poor performance when the density of clutter or spurious measurements is high because the measurement chosen for the update could be erroneously selected in the clutter and thus unrelated to the object state.

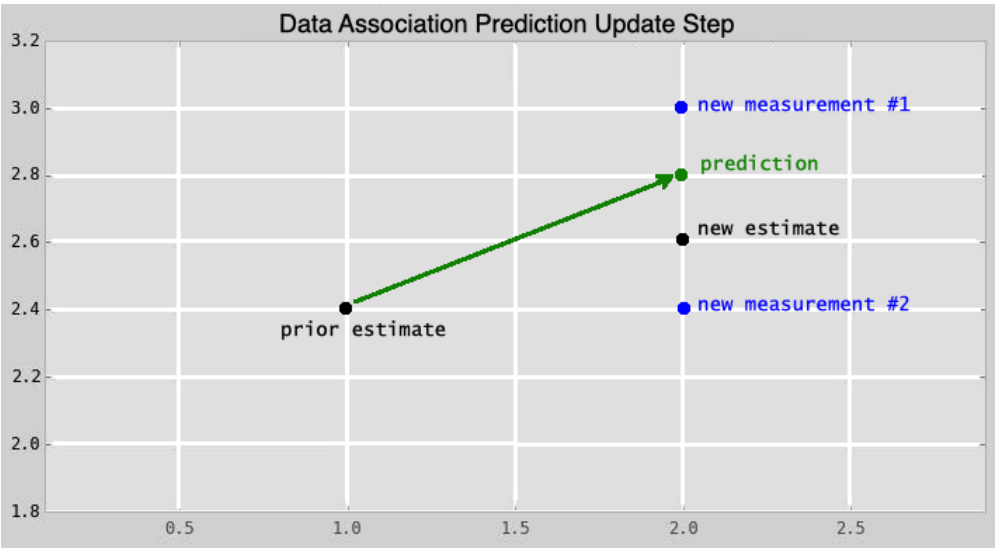

Figure 7.3: Data association prediction update step: example with two received measurements. The horizontal axis are the time steps, while the vertical axis corresponds to the value of the state variable. At the previous iteration $t = 1$ we placed the *prior estimate* at $y = 2.4$, and we update that estimate by using the `model` which gives us the *new estimate* for the present $t = 1$. In addition, at the current iteration $t = 1$ we receive two *new measurements* (blue dots) and assign a particular weight to each. At the update step we combine the *new estimate* with the *new measurements* and we obtain the final *prediction* (green dot) for the current time $t = 1$ at $y = 2.8$.

Hence, for our estimation problem we decide to use the Probabilistic Data Association Filter (Section 5.4.3) instead, described in Algorithm 32, because it is a robust approximation of the **optimal** Bayesian filter recursive solution for scenarios where multiple simultaneous measurements are available and one of the most successful algorithms in the class of Bayesian all-neighbors filters[17].

Algorithm 31 Nearest Neighbor Filter Recursion

For both SPATIAL and TEMPORAL tracking:
Select the object dynamics, measurement and noise models.
Load the parameters into our `ModelNNF`.
At time $t = 0$:

1. **Input** the initial *state estimate* and the initial *model prediction*
2. **Update estimation** with the two initial values above
3. **Update prediction** with the *updated estimation*
4. **Output** *first-order boundary* prediction at $t = 0$

At time $t = k$:

1. **Input** first-order boundary *measurements* at $t = k$ and the first-order boundary *prediction* at $t = k - 1$ is applied as the *current model prediction*.
2. **Update estimation** using the *model prediction* at $t = k$ and the first-order boundary *estimation* from measurements at $t = k$.
3. Among all the measurements received at $t = k$, only one of the measurements is selected to update the prediction. The decision is based on the following criterion (Equation 5.21):

$$\mathbf{y}_k(i) = \underset{y_k(j),\ \forall j \in \{1,\ldots,m_k\}}{\operatorname{argmin}} \left[\mathbf{y}_k(j) - \mathbf{H}\hat{x}_{k|k-1}\right]^T \mathbf{S}_{k|k-1}^{-1} \left[\mathbf{y}_k(j) - \mathbf{H}\hat{x}_{k|k-1}\right]$$

   where $\mathbf{S}_{\mathrm{k|k-1}} = \mathbf{H}\mathbf{P}_{\mathrm{k|k-1}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{\mathrm{k}}$.
4. **Update prediction** at $t = k$ with the *updated estimation*.
5. **Output** *first-order boundary* prediction at $t = k$.

The Probabilistic Data Association Filter, presented in Algorithm 32 below, performs by updating the object state estimate using all of the validated measurements and their respective posterior probability weightings and then estimates a boundary location somewhere in between the multiple single boundary location possibilities we have received as validated measurements from Algorithm 21.

7.5.1 Bayesian Tracking in Range

The purpose of applying a spatial Bayesian Tracking in range is to smooth and correct the boundaries from the understanding that

1. the boundary location in a specified Doppler bin corresponds to the maximum radial velocity current present in that range cell,

2. oceanographic currents would generally not increase nor decrease excessively in the relatively reduced distance of 5-10 km corresponding to a range width,

3. thus we study the current variability in the corresponding radar site and establish a "maximum realistic change" in the ocean current velocity which we then set as our initial estimated model behavior `model_0` on the `First OrderConfigs` file corresponding to the site.

4. Generally, we initially set `model_0`=1 which is a model that we describe as “smooth boundaries model” which just means that the boundary should stay constant.

Algorithm 33 presents the Range Tracking steps that we perform to the "initial first-order boundaries" that resulted from Algorithm 21.

Algorithm 32 Probabilistic Data Association Recursion

For both SPATIAL and TEMPORAL tracking:
Select the object dynamics, measurement and noise models.
Load the parameters into our `ModelPDAF`.
At time t=0:

1. **Input** *initial state estimate* and *initial model prediction*
2. **Update estimation** with the two initial values above
3. **Update prediction** with the *updated estimation*
4. **Output** *first-order boundary* prediction at t=0

At time t=k:

1. **Input** all the *first-order boundary measurements* at $t = k$ and the *first-order boundary prediction* at $t = k - 1$ as the present *model prediction* for time $t = k$.
2. **Update estimation** using *model prediction* at $t = k$ and the *first-order boundary estimation* from measurements at $t = k$.
3. The **data association probabilities** are denoted by $\beta_k(i)$ (known as "weights") and are calculated for the ith validated measurements.
4. The *posterior* probability density function is a first-order boundary can be a location anywhere within the possible measurements, as in Eq.(5.31), and will become the **updated estimation** at this time.

$$P(X_k \mid Y^k,\ m^k) = \beta_k(0)\, \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) + \sum_{i=1}^{m_k} \beta_k(i)\, \mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}^i_{k|k-1}, \mathbf{P}^i_{k|k-1})$$

5. **Update prediction** at this time $t{=}k$ with the *updated estimation*.
6. **Output** *first-order boundary* prediction at $t{=}k$.

Algorithm 33 Range Tracking to the Initial First-Order Boundaries

1) **Load** the initial first-order boundary estimates from Algorithm 21
$[FOL1, FOL2, FOL3, FOL4] \leftarrow$ `AlternativeFOL(cs_file)`

for each BOUNDARY in $[FOL1, FOL2, FOL3, FOL4]$ **do**
for each RANGE **do**
2) PDAF steps described in Algorithm 32 to:

i. Load the estimates of the possible Doppler bin locations depending on the *tracking mode* selected under parameter `range_tracking_mode`. (Section 7.1)

ii. Determine the data association probabilities (weights) of each of those possible Doppler bin locations.

iii. Calculate the posterior probability density function that is translated into the updated estimation.

iv. Use this prediction at the current time t=k into the new recursion at time t=k+1.

end for
end for

3) **Return** *FirstOrderBoundaries_Range*

7.5.2 Bayesian Tracking in Time

Similarly to what we described in Section 7.5.1, the purpose of applying a temporal Bayesian tracking in time is to smooth and correct the boundaries from the understanding that, since the boundary location in Doppler bin corresponds to the maximum radial velocity current present in that range cell, this location should not increase or decrease excessively in the period of time between spectra files (normally around 10 to 30 minutes). Hence, we usually also consider a "smooth boundaries model" in time and set `model_0`=1.

Algorithm 34 presents the temporal tracking steps that recursively update the boundaries with their immediately preceding estimates from the anterior timestamp. The inputs of this time tracking algorithm are the first-order boundary estimates after they have gone through the spatial correction in Algorithm 33 from Section 7.5.1.

Summarizing this chapter: we developed a set of spatial Bayesian filters, which track the measurements in range (both outwards from the radar and inwards from the furthest range) and are based on the assumption that the maximum radial ocean current found on two neighboring ranges from the radar should not present a strong variation between them. On the other hand, we also developed a temporal Bayesian filter, which tracks the estimates in time. In order words, every filtered set of first-order boundary estimates will be based on all available information from previously processed datasets.

Therefore, at the end of the complete Bayesian tracking process, the estimates have been decided based on a weighted combination of

1. their present measurement,
2. their immediately previous measurement from the neighbor ranges,
3. and their immediately preceding measurement from the anterior timestamp.

Finally, the next two chapters present a series of comparisons that demonstrate the improved results obtained from applying these Bayesian filters to the proposed boundary estimates resulting from Algorithm 21.

Algorithm 34 Temporal Tracking to the Initial First-Order Boundaries

1) **Load** the first-order boundary estimates from Algorithm 33 that are already corrected in range.
$[FOL1, FOL2, FOL3, FOL4] \leftarrow FirstOrderBoundaries\_Range$

for each BOUNDARY in $[FOL1, FOL2, FOL3, FOL4]$ **do**
for each RANGE **do**
2) PDAF steps described in Algorithm 32 to:

i. Load the estimates of the possible Doppler bin locations from the previous spectra file.

ii. Determine the data association probabilities (weights) of each of those possible Doppler bin locations.

iii. Calculate the posterior probability density function that is translated into the updated estimation.

iv. Use this prediction at the current time $t{=}k$ into the new recursion at time $t{=}k{+}1$.

end for
end for

3) **Return** *FirstOrderBoundaries_Time*
as the **final** proposed first-order boundary estimates.

Chapter 8

Demonstration of the Validity of the Method

In this chapter - as well as in Chapter 9 - we study the performance and the validity of our proposed, multi-probabilistic method for the determination of the first-order Bragg line boundaries. We demonstrate how using Bayesian estimation in the determination algorithm properly tracks and corrects first-order boundaries, improving the overall accuracy of the radial current vectors by:

- adding radial vectors that statistically compare well with the surrounding radial vectors in the area,
- while rejecting radial vectors whose standard deviation from the surrounding ones is indeed higher than the added vectors.

In Section 8.1 we review the standard statistical methods that are used to demonstrate the efficacy of our proposed multi-probabilistic method, namely: the **standard deviation** and the **root mean square error** as measures of the difference between the conventional and the new proposed method.

In Section 8.2 we present comparisons of the first-order boundary solutions from the currently-used null finding method (namely the "Conventional" method) and the proposed Bayesian method we developed that we call 'multi-probabilistic first-order boundary determination' method. We will be showing scenarios where the Bragg region presents splits, like introduced in Section 3.3.2, and where the proposed multi-probabilistic method is indeed very successful. These comparisons between methods will be studied both from the radar backscattered cross-spectra point of view as well as on the corresponding radial current maps that are generated from these cross-spectra.

Lastly, in Section 8.3, we study a series of statistical comparisons between the methods. The objective of this statistics is to demonstrate how the use of multi-probabilistic estimation techniques helps to properly track and correct the boundaries and improves the overall correlation of the radial current vectors that are obtained with the proposed method.

For perspective, we note that in Chapter 9 we demonstrate the validity of the proposed multi-probabilistic method, but –contrary to Section 8.2– focusing in scenarios where there are no splits in the Bragg regions. We compare the resulting radial current vectors from the cross-spectra with a series of drifters deployed along the Straits of Georgia, and we conclude that the vectors added by the proposed multi-probabilistic method match better with the radial current velocity vectors extracted from the drifters, as opposed to the rejected vectors that present a higher average root mean square error when compared to the drifter velocities.

8.1 Statistical Comparisons and Error Analysis

In this section we define the statistical methods that are used in this chapter to demonstrate the efficacy of our multi-probabilistic method and its statistical comparison against the conventional method. These standard statistical methods are, namely, the **standard deviation** and the **root mean square error** as measures of difference between the radial velocity vectors obtained by the two different first-order Bragg boundary determination methods.

The most probable estimate of the **mean** μ of a random set of observations is the average $\left(\overline{x} = \frac{1}{N}\sum x_i\right)$ of the observations. Assume we have a set of N data points, distributed according to the Gaussian distribution, with mean μ and standard deviation σ. Therefore, the probability dP_i for making any single observation x_i within the interval dx is given by $dP_i = p_i dx$, where p_i is the probability density function.

The **standard deviation** σ is used to measure the spread of data around the mean. Its form, in Formula 8.1, is the square root of the **variance** σ^2. The variance is defined as the average of the squared differences of each data point value from the **mean** μ.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{8.1}$$

The probability P_i of making an observation x_i is:

$$P_i = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right]. \tag{8.2}$$

We define μ' as the estimated mean, since generally we do not know the actual value of the mean μ of the distribution. We apply the method of **maximum likelihood**, and it states that the most probable value for μ' is the one that gives the maximum value for the probability $P(\mu')$.

$$P(\mu') = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum\left(\frac{x_i - \mu'}{\sigma}\right)^2\right]. \tag{8.3}$$

Maximizing the probability $P(\mu')$ is equivalent to minimizing the argument X of the exponential,

$$X = \frac{-1}{2}\sum\left(\frac{x_i - \mu'}{\sigma}\right)^2, \tag{8.4}$$

by setting its derivative to 0.

In other words,

$$\frac{dX}{d\mu'} = \frac{-1}{2}\sum\frac{d}{d\mu'}\left(\frac{x_i - \mu'}{\sigma}\right)^2 = \sum\left(\frac{x_i - \mu'}{\sigma^2}\right) = 0, \tag{8.5}$$

which, since σ is a constant, gives

$$\mu' = \overline{x} = \frac{1}{N}\sum x_i. \tag{8.6}$$

Thus, as we discussed at the beginning of this section, the maximum likelihood method for estimating the mean by maximizing the probability $P(\mu')$ of Equation (8.3) shows that the most probable value of the mean is just the average $\overline{x}$.

Estimating the error in the mean μ': In order to calculate the uncertainty of the mean, we apply Equation (8.7) to find the variance σ_μ^2 of the mean μ'.

$$\sigma_\mu^2 = \sum \left[\sigma_i^2 \left(\frac{\partial \mu'}{\partial x_i} \right)^2 \right]. \tag{8.7}$$

If the uncertainties are equal for each data point[14], $\sigma_i = \sigma$, the partial derivatives in Equation (8.7) are

$$\frac{\partial \mu'}{\partial x_i} = \frac{\partial}{\partial x_i} \left(\frac{1}{N} \sum x_i \right) = \frac{1}{N}, \tag{8.8}$$

and combining Equations (8.7) and (8.8), we finally obtain the **variance of the mean** (when all uncertainties are equal) as

$$\sigma_\mu^2 = \sum \left[\sigma_i^2 \left(\frac{1}{N} \right)^2 \right] = \frac{\sigma^2}{N}. \tag{8.9}$$

Formula (8.9) corresponds to the **estimated error in the mean when all uncertainties are equal**, and it demonstrates how the precision of our estimate of the mean μ improves as the square root of the number of measurements N.

If the uncertainties of the measurements are not equal, we evaluate the differential in Equation (8.8) differently, as follows:

$$\frac{\partial \mu'}{\partial x_i} = \frac{\partial}{\partial x_i} \frac{\sum (x_i/\sigma_i^2)}{\sum (1/\sigma_i^2)} = \frac{1/\sigma_i^2}{\sum (1/\sigma_i^2)}, \tag{8.10}$$

and therefore, the **general formula for the uncertainty of the mean** σ,

$$\sigma_\mu^2 = \sum \left[\sigma_i^2 \left(\frac{\partial \mu'}{\partial x_i} \right)^2 \right] = \sum \frac{(1/\sigma_i^2)}{\left[\sum (1/\sigma_i^2) \right]^2} = \frac{1}{\sum (1/\sigma_i^2)}. \tag{8.11}$$

Another measure we use is the **Root Mean Squared Error**, which is the standard deviation of the residuals. While the standard deviation σ is used to measure the spread of data around the mean, the RMSE is used to measure distance between predicted and actual values. To compute the RMSE we have to calculate the residual (difference between the prediction and the truth value) for each data point, then compute the norm of the residual for each data point, compute the mean of residues and finally take the square root of that mean. The RMSE Formula (8.12) is below:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{8.12}$$

RMSE is a measure of how spread out these residuals are. In other words, it tells us how concentrated the data are around the line of best fit. The RMSE provides an average model prediction error in units of the variable of interest. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors.

8.2 Analogy of the Two Methods Radial Current Outputs

In this section we present a series of comparisons of the first-order boundary solutions from the currently-used null finding method (the "Conventional" method) and the proposed method we developed i.e. the "multi-probabilistic first-order boundary determination" method.

The differences in the determinations of the first-order boundaries will be shown

1. on the radar backscattered cross-spectra, where we will observe the two first-order Bragg regions, one on each side of the zero-Doppler line, and the determined boundaries drawn on top – separating these first-order regions from second order echoes and noise or interference that might be present.

2. on their corresponding radial current maps, where a different amount of vectors will be present depending on the boundaries corrections.

The proposed multi-probabilistic method is particularly successful for instances where there was splitting of the first order Bragg regions (Chapter 3). We will observe, indeed, many "added" vectors appearing on the radial current map corresponding to the proposed multi-probabilistic method, mainly in the ranges (distance from the radar site) where the Bragg regions had split into multiple regions as we discussed in Section 3.3.2 and in Chapter 4 as well as observed in Figure 4.9, Figure 8.3 and Figure 8.4.

On the other hand, the radial current map for that same cross-spectra, but the one corresponding to the first-order boundaries determined by the *conventional* method, will miss some of those vectors, due to the fact that the method is not able to track the first-order boundaries properly under the presence of strongly split Bragg regions.

In addition to displaying results of both methods in terms of the radar back-scattered cross-spectra, and the corresponding radial current maps for each of the two methods, we will also present a study of the validity of the proposed first-order boundary determination algorithm. To do so, we introduce a series of statistical comparisons with the objective of demonstrating how the implementation of our multi-probabilistic estimation techniques to properly track and correct the boundaries improves the overall correlation of the retrieved radial vectors.

This validation rests the fact that, as we will see below, the proposed multi-probabilistic method is adding radial vectors that statistically compare well with the surrounding vectors in the area, while it rejects radial vectors whose deviation with respect to the vectors around them is indeed higher.

Study of the First-Order Boundaries and the corresponding Radial Current Velocity Maps for the VCOL site at the Straits of Georgia: The SeaSonde radar sites at the Straits of Georgia, like the ones shown in Figures 8.1 and 9.3, often present splitting of their Bragg regions. In this case, it happens due to the behavior of the ocean currents in the strait and the tide cycles associated with this type of oceanographic feature.

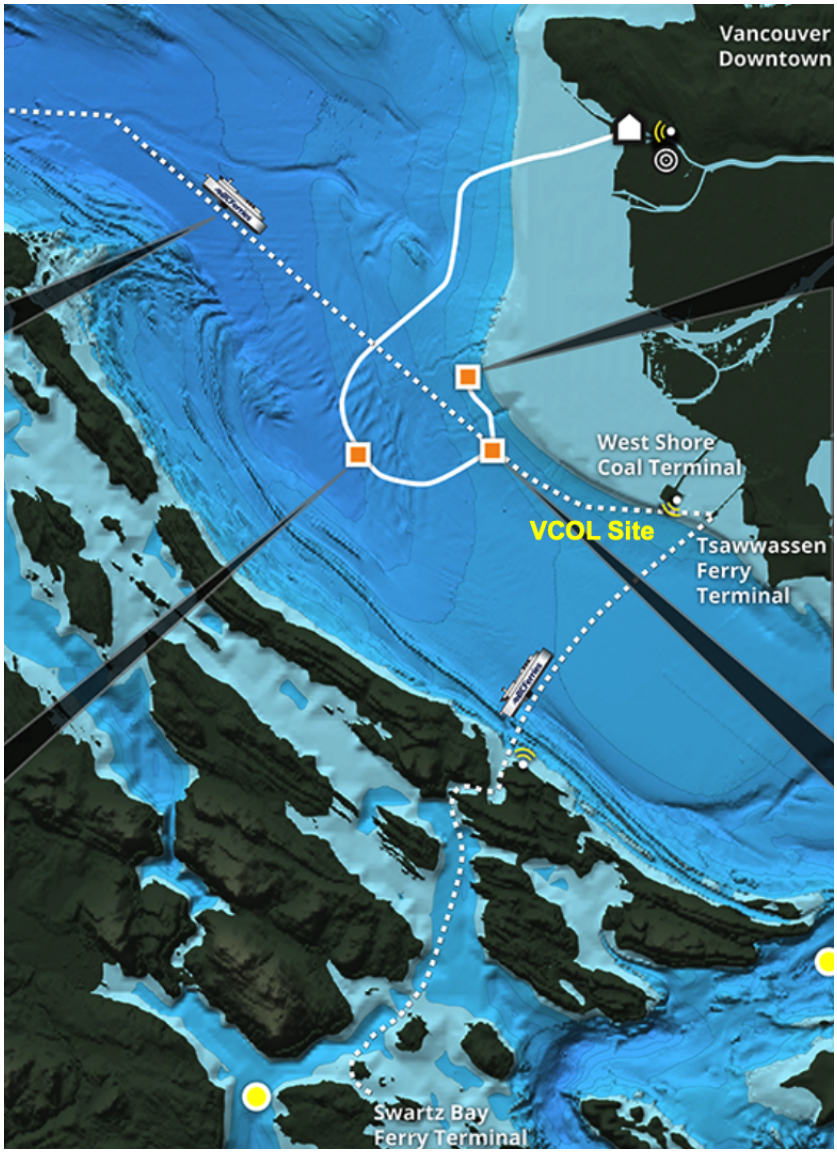

Figure 8.1: Location of the VCOL SeaSonde radar site at the Strait of Georgia. Map credits to Ocean Networks Canada[16]. The VCOL site is found at the end of the West Shore Coal Terminal, near the Tsawwassen Ferry Terminal in Delta, British Columbia. The map also shows some ferry routes in dashed lines, and a fiber-optics cable in a solid white line, along with the square orange and round yellow symbols that represent other instruments that ONC has on site. The map also shows other radars (symbols in white with small yellow waves) that correspond to other SeaSonde sites built by CODAR along the Strait of Georgia.

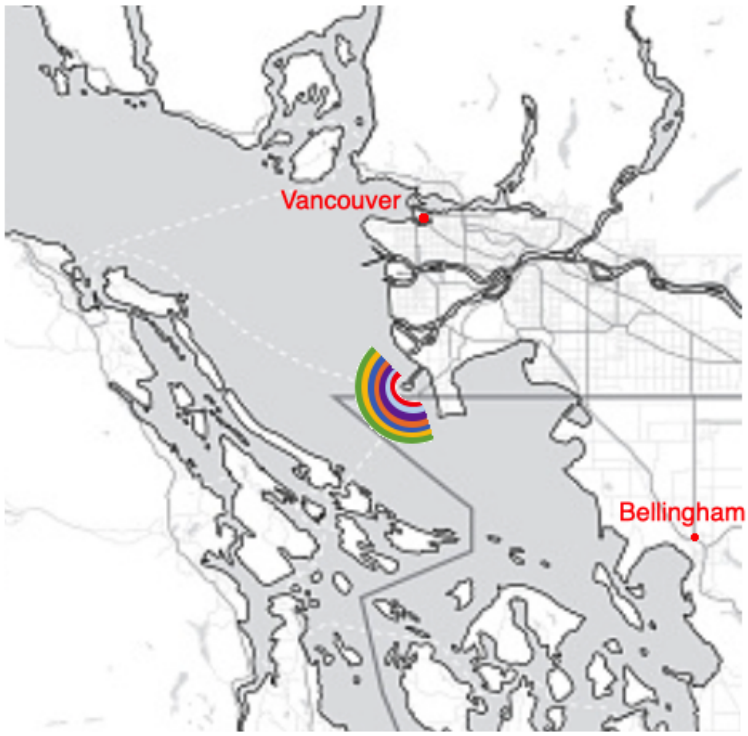

Figure 8.2: Map showing the constant range arcs (range cells) at the VCOL SeaSonde radar site near the ferry terminal in Vancouver, Canada. Each range arc, presented in different colors, provides a separate 'Range Spectra' from which we should extract the corresponding first-order region boundaries.

For this chapter we focus on a set of cross-spectra obtained during the first day of July 2020 from the VCOL SeaSonde radar site located at the West Shore Coal Terminal near Delta, British Columbia. The exact location of the radar site (49°01.083'N,123°10.313'W) and the surrounding area are presented in Figure 8.1.

VCOL is a radar site working at 24.4 MHz center frequency and its maximum range is about 45 km, divided into range rings of about 500 meters width. The sweep rate for this radar is 4 Hz and every Doppler spectrum is divided into 1024 Doppler bins. We obtain a cross-spectra file every ten minutes at this site, and it corresponds to a temporal coverage of 15 minutes averaged.

Before we start comparing the first-order boundaries determinations by each of the methods, it is worth having a first look at an example radar backscattered signal corresponding to the 7:50 a.m. UTC on July 1st 2020. Figure 8.3 shows the backscattered signal for the three antennas. Namely, cross-loop Antenna 1 and Antenna 2, and the monopole Antenna 3 as we introduced in Chapters 2 through 4 (particularly Section 2.3, Section 3.2.1 and Section 4.1).

Overall we can see how the energy backscattered on the three antennas presents split regions, especially on the right Braggs on Figure 8.3. As introduced in Section 4 the *conventional* method is not prepared to adapt to split Braggs and when it is split the method would erroneously place one of the first-order boundaries by the lowest peak on the split, as we will see later in this section.

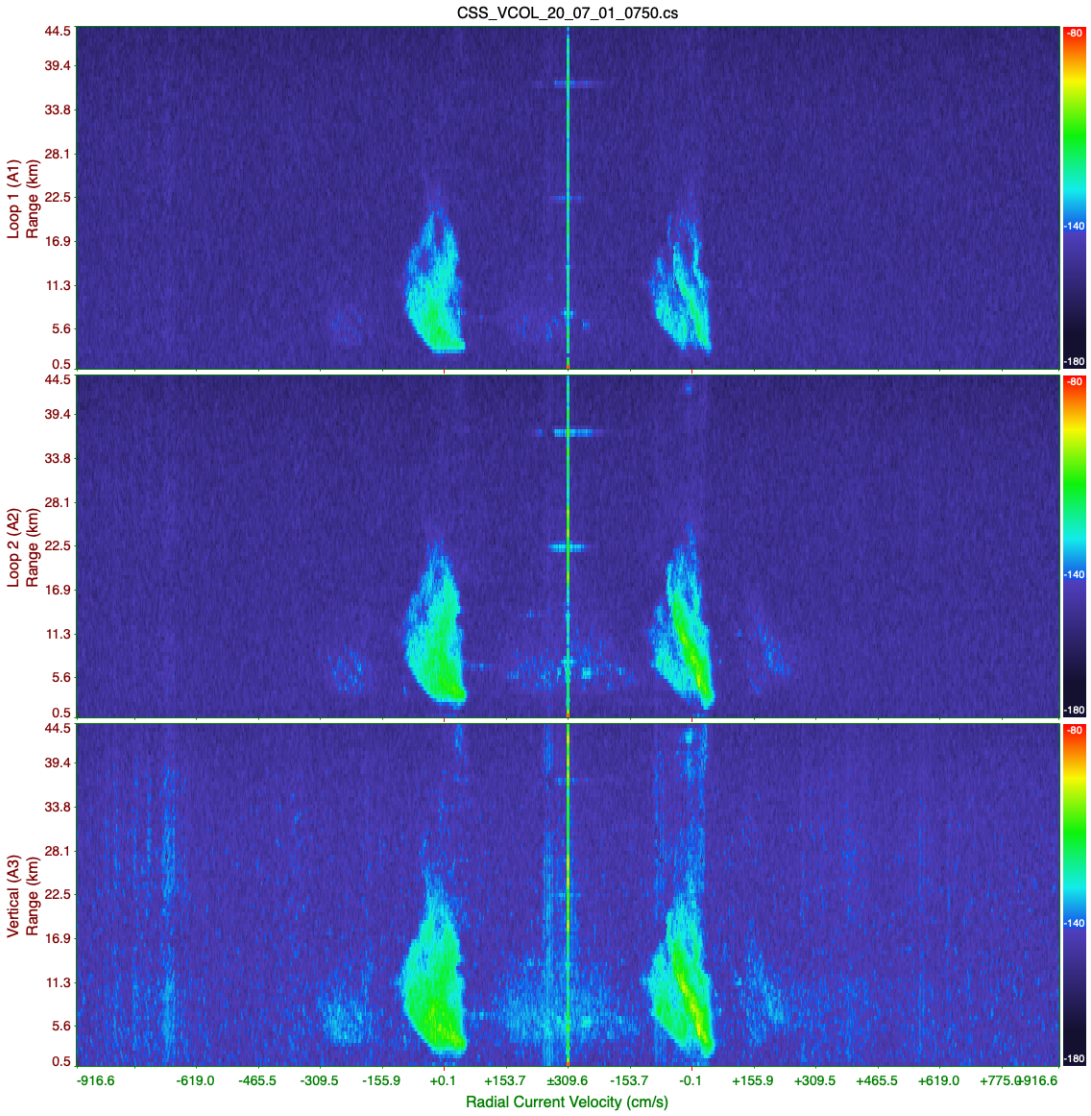

Figure 8.3: VCOL radar site backscattered signal from the 1st day of July 2020 at 7:50 a.m. (UTC). The top two figures correspond to the backscattered spectra received at the crossed-loop Antennas 1 and 2, respectively. On the other hand, the bottom section describes the backscattered signal from the monopole Antenna 3. We observe some instances of the Bragg region energy splitting, mainly on the right Bragg portions on the three antennas.

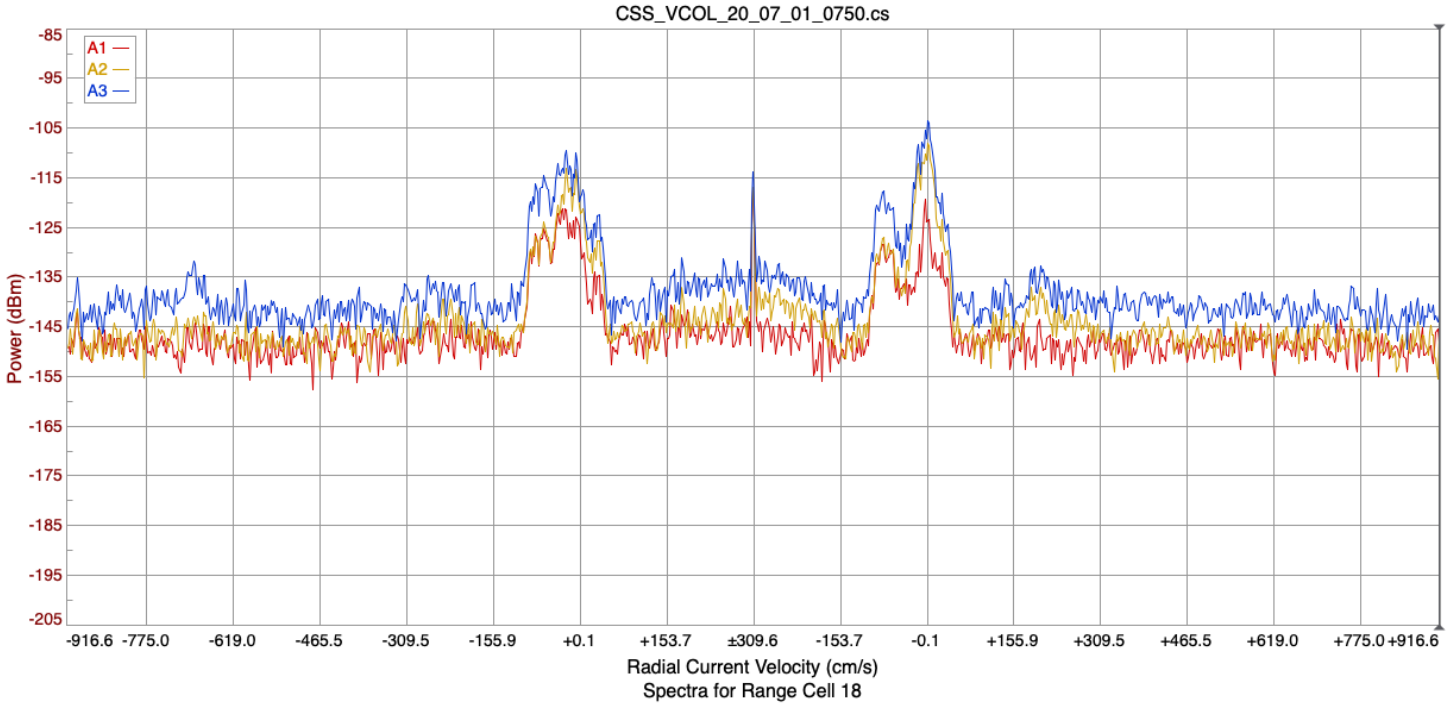

Figure 8.4: Detail of the 18th range cell backscattered signal from the VCOL radar site on the 1st of July of 2020 at 7:50 a.m. (UTC). The red and yellow lines correspond to the crossed-loop Antennas 1 and 2, respectively. On the other hand, the blue curve represents the backscattered signal from the monopole Antenna 3. We can clearly observe the split of the right Bragg region, where the peak power for Antenna 3 is slightly above -105 dBm, then the split dips down to below -130 dB. The left side of the split, weaker than the peak power, has a maximum around -117 dBm.

An instance of the Bragg region splitting, described in Section 3.3.2, is well illustrated in Figure 8.4. We observe that the first-order Bragg region for positive Doppler shift (right hand side) is not a single peak, but has split into two peaks. 8.4 corresponds to the 18th range cell, in other words about 9 km distance from the radar site, for the 1st of July of 2020 around 7:50 a.m. (UTC).

At the 9 km range presented on Figure 8.4, we easily observe the split of the right Bragg region. Focusing on the monopole Antenna 3, we see its peak power slightly above -105 dBm, then the split drops down to below -130 dB. The left side of the split, weaker than the peak power on the right portion, has a maximum around -117 dBm at this range.

Next, we start comparing how well the two methods respond to this phenomenon when applying their respective algorithms to determine the first-order boundaries.

Boundaries from the conventional method:

Our first step is to extract, from the cross-spectra, the original first-order boundaries as determined by the *conventional* method. Figure 8.5 presents the first-order boundaries drawn in red, corresponding to the first-order boundaries obtained from the *conventional* method described in Chapter 4 Algorithm 18.

We observe how overall the *conventional* boundaries are slightly wider than the strong signal corresponding to the first-order Bragg region. This is especially evident in the left side of the left Bragg region, that corresponds to the receding Bragg, around the first 10 km in range. This leads to believe that the power threshold for the first-order boundaries that was set for this radar at the time of acquisition was not too restrictive.

More importantly is what we observe on the right Bragg region. As we can see, the strongest backscattered signal received is around the center of the right Bragg region for ranges between 3 and 12 km. In addition, for those ranges the right Bragg presents a strong split where the backscattered signal is strong on both sides of the split of the right Bragg, even though it is predominant on the right split portion.

Each one of the range cells with split Bragg regions in this area would look similar to what we saw on Figure 8.4 for the 18th range bin, since from Figure 8.5 we observe they all present a similar split.

The fact that the first-order boundaries on some of the ranges, as observed on the right Bragg on Figure 8.5, miss the split and therefore a relevant portion of the backscattered energy is neglected, will impact the number of radial current velocity vectors obtained using the conventional method. As expected, all the radial current velocity vectors corresponding to those areas of the cross-spectra in Figure 8.5, will be lost and the MUSIC algorithm will not consider them for placement in the radial map.

Examples of the radial map solutions will be discussed later in this section, once we evaluate the results from the Bayesian method for comparison.

Boundaries from the proposed multi-probabilistic null-finding method:

In order to obtain the first-order boundaries as determined by the proposed multi-probabilistic null-finding method we have to execute two algorithms, which are detailed in Section 7.5.

The first stage is to run the proposed null-finding algorithm described on Chapter 4. This algorithm, contrary to the conventional null-finding method, allows multiple choices for the location of the boundary separating first and second order in each of the range bins. This first step provides multiple possible locations of the First-order boundaries shown in Figure 8.6.

Once we have completed the null finding, the second stage is to apply the two-steps Bayesian estimation algorithms described on Chapter 5 and Chapter 7. The first Bayesian step is to perform a tracking in Range (Section 7.5.1 Algorithm 33), which will place each first-order boundary location based on the location of the previous-range first-order boundary, but nudged by a correction determined by the observational data at the range under consideration.

Once the boundaries have been corrected in range, we will apply the second Bayesian step which is the temporal tracking (Section 7.5.2 Algorithm 34). We filter the boundaries using the Bayesian estimation algorithm in time, in other words: the final first-order boundaries will be decided based on where the first-order boundaries were placed at the previous timestamp, but again nudged by the observational data.

The final decision for the first-order boundary locations - as provided by the proposed multi-probabilistic null-finding method (first calculated by the proposed null-finding algorithm and finally corrected through the two steps of Bayesian tracking) - are presented in Figure 8.7.

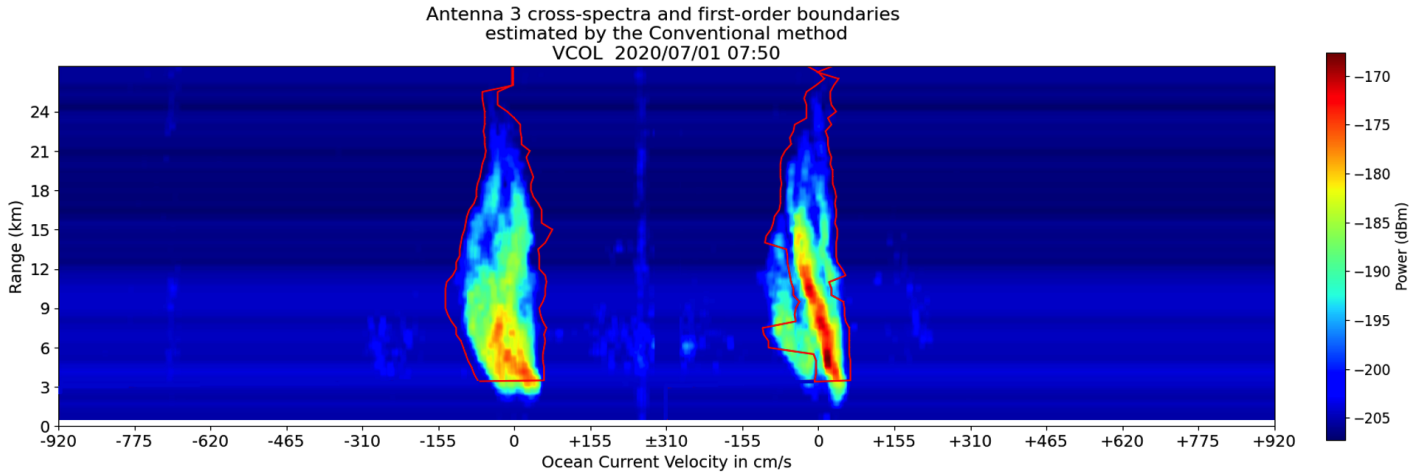

Figure 8.5: First-order boundaries determined by the Conventional method. backscattered cross-spectra for Antenna 3 from the VCOL radar site on the 1st of July of 2020 at 7:50 a.m. (UTC). The first-order boundaries drawn in red in the Figure correspond to the Conventional method results. Observe how in the ranges between 3 and 15 km the red boundaries do not properly track the right first-order Bragg region, due to the fact that the Conventional method is not able to detect the splitting of the Bragg region.

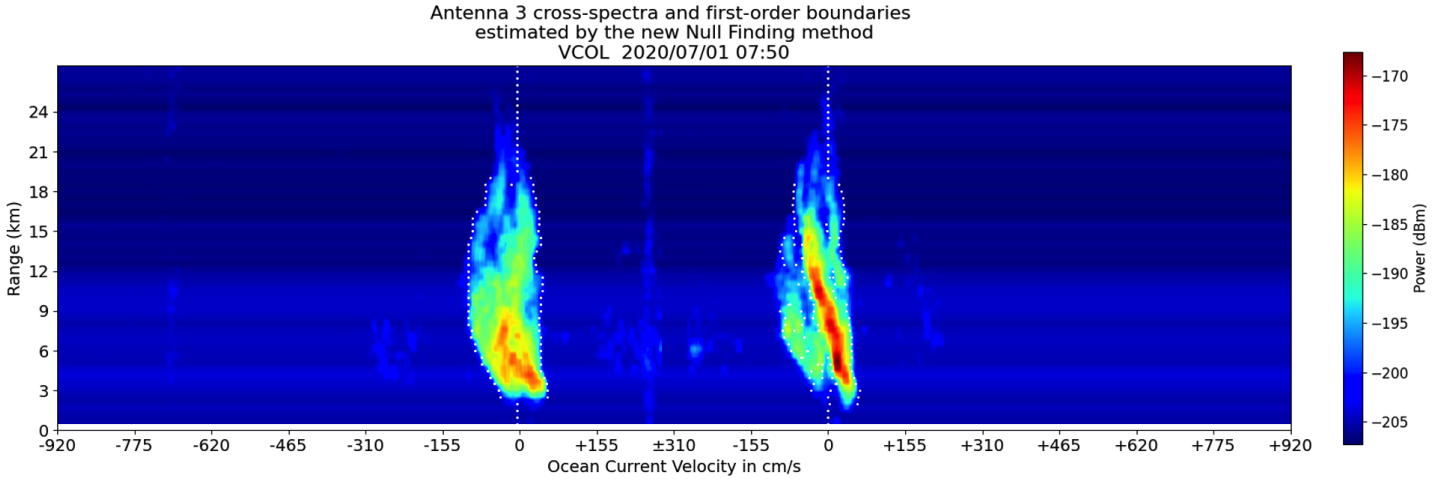

Figure 8.6: Initial possible locations of the First-order boundaries provided by the proposed multi-probabilistic null-finding method. Backscattered cross-spectra for Antenna 3 from the VCOL radar site on the 1st of July of 2020 at 7:50 a.m. (UTC). The first-order boundaries marked as white dots, correspond to the initial first-order boundary options as provided by the proposed multi-probabilistic null-finding method. We observe how in many ranges there are multiple possible options for the location of the first-order boundary, since the method is able to detect that the Bragg region has split in two, or even more, branches. After applying the corresponding Bayesian tracking in range and time to this initial first-order boundary options we obtain the final first-order boundaries that the proposed multi-probabilistic null-finding method outputs. These final boundaries are shown next, in Figure 8.7

.

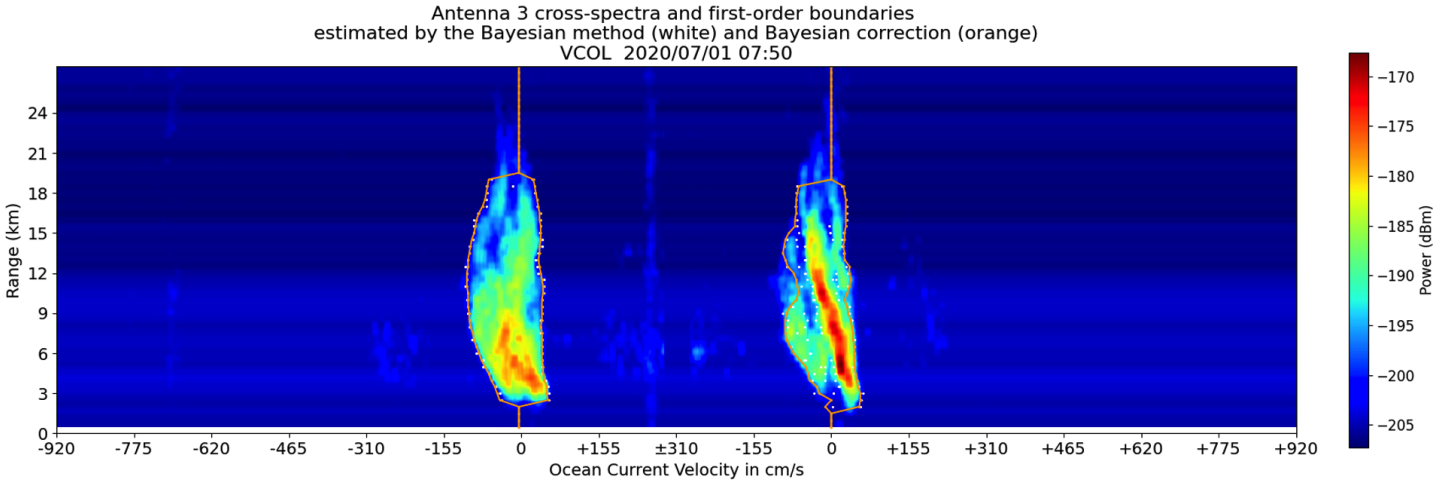

Figure 8.7: Final decision for the locations of the First-order boundaries provided by the proposed multi-probabilistic null-finding method after Bayesian tracking in Range and Time. Backscattered cross-spectra for Antenna 3 from the VCOL radar site on the 1st of July of 2020 at 7:50 a.m. (UTC). The first-order boundaries drawn in orange, correspond to the first-order boundary final decision as provided by the proposed multi-probabilistic null-finding method. These orange boundaries are determined after applying Bayesian filtering in Range and Time to the initial locations shown in Figure 8.6, as detailed in Chapter 5. We observe how these solutions, opposed to the Conventional method solutions in Figure 8.5, do a better job when tracking the first-order boundaries and include the lower energy portion of the split Bragg.

Evaluation of the radial current velocity maps for the VCOL site:

Up until this point, we have evaluated the effects of the proposed multi-probabilistic method first-order boundaries from the point of view of the back-scattered cross-spectra. We have determined the first-order boundaries by both conventional and our new multi-probabilistic methods on a dataset from July 1, 2020 in the VCOL radar site at the Straits of Georgia.

Next, we present the differences that the proposed multi-probabilistic method brings to the radial ocean surface current velocity maps. As one can imagine, working with more accurate first-order boundaries like the ones shown in Figure 8.7 as opposed to the erroneous Conventional method first-order boundaries in Figure 8.5 means that a larger number of radial current vectors are brought into the radial maps, since a larger region of first-order backscattered energy is accepted and thus, for a given range bin the MUSIC algorithm will place more vectors at their estimated bearings from the radar location. Further, some vectors that were accepted as valid by the conventional method will be rejected, using the new boundaries.

We assign the following names to the radial current vectors:

- "Agreed Vectors": set of vectors that both the Conventional and the proposed multi-probabilistic method place in the map.
- "Added Vectors": set of vectors provided by the proposed multi-probabilistic method that the Conventional method missed.
- "Rejected Vectors": set of vectors from the Conventional method that the proposed multi-probabilistic method rejects due to the revised boundaries.

The objective is, of course, to add as many good vectors as possible, while rejecting only vectors that do not correspond to actual ocean surface currents present in the field of view. Next, in Section 8.3, we will evaluate the statistical features of each of the vector sets.

But first, we are going to evaluate how the aforementioned vector sets look placed in their corresponding radial maps. To do so, we are presenting two sets of cross-spectra from the VCOL site in the morning of July 1, 2020. One corresponds to 3:20 a.m. UTC while the other one is from 7:40 a.m. UTC.

Figure 8.8 presents three panels corresponding to July 1, 2020 at 7:40 a.m. UTC. The left one corresponds to the set of "Agreed vectors", while the one in the middle is the smaller set of vectors that the proposed multi-probabilistic null-finding method adds. We can see how the vectors are placed all around a smaller region in range. This is because those vectors correspond to the ranges where there was a split Bragg region (ranges from 3 to 12 km from the radar). As seen in Figure 8.5 and Figure 8.7, the Conventional method missed some of the first-order backscattered energy when placing the boundaries, while the proposed multi-probabilistic method was able to capture it more accurately, as described in Section 8.3. This translates into a number of vectors being added by the proposed multi-probabilistic null-finding method mainly in the ranges where the split Bragg occurs. Finally, the third portion of Figure 8.8 shows the complete set of vectors as provided by the proposed multi-probabilistic null-finding method.

Similarly, Figure 8.9 shows similar results as Figure 8.8 but corresponds to the cross-spectra obtained at 3:20 a.m. UTC.

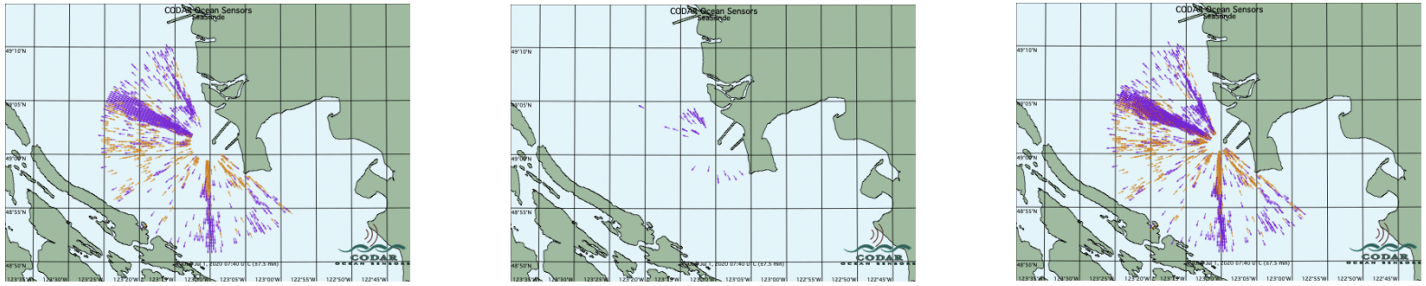

Figure 8.8: First set of radial current velocities map at the VCOL radar site in the Straits of Georgia from July 1, 2020 at 7:40 a.m. UTC. The left map includes the radial vectors provided by both the conventional and the proposed multi-probabilistic null-finding method used to determine the first-order boundaries. The map in the middle portion shows the radial current velocities that the proposed method determines and the conventional method misses. Finally, the map on the right presents all the vectors provided by the proposed method combined.

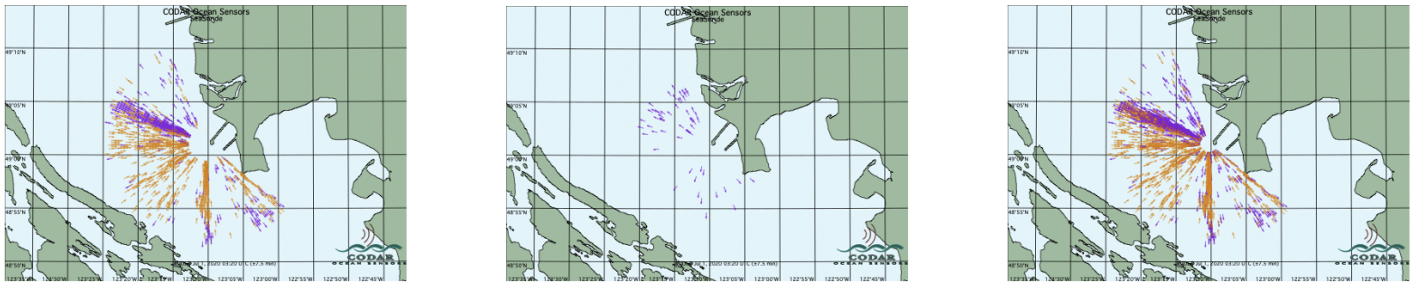

Figure 8.9: Second set of radial current velocities map at the VCOL radar site in the Straits of Georgia from July 1, 2020 at 3:20 a.m. UTC.. The left map includes the radial vectors provided by both the conventional and the proposed multi-probabilistic null-finding method used to determine the first-order boundaries. The map in the middle portion shows the radial current velocities that the proposed method determines and the conventional method misses. Finally, the map on the right presents all the vectors provided by the proposed method combined.

8.3 Study of Statistical Differences of the Radial Current Outputs Between Methods

In Section 8.2 we presented comparisons of the first order boundary solutions from the currently-used null finding method (the "Conventional" method) and the proposed method we developed (the "multi-probabilistic" method).

We saw how the proposed Bayesian method helps to properly track and correct the boundaries, and therefore improves the accuracy of the resulting radial current vectors maps that are obtained with the proposed method, as we presented in Section 8.2. We evaluated the results from the radar backscattered cross-spectra point of view as well as on the corresponding radial current maps.

The radial vectors that we place on the maps are obtained from the cross-spectra computed between voltage time series from the elements of the receive antenna as described by Barrick and Lipa (1997) based on the MUSIC algorithm of Schmidt (1986). In order to estimate cross-spectra from signal voltage time series between antennas, time averaging is required, typically ranging from several minutes to about an hour, depending on radar operating parameters.

Finally, in this Section, we study a series of statistical comparisons between the methods. The objective of this statistics is to demonstrate how the vectors added by the method using Bayesian estimation techniques improves the overall correlation of the radial current vectors with the neighborhood vectors in the map.

8.4 Evaluation of the Error on the First-Order Boundary Determination Methods

As previously mentioned, we assign the following names to the radial current vectors:

- "Agreed Vectors": set of vectors that both the Conventional and the proposed multi-probabilistic method place in the map.
- "Added Vectors": set of vectors provided by the proposed Bayesian method that the Conventional method missed.
- "Rejected Vectors": set of vectors that the Conventional method had provided but the proposed Bayesian method decides to reject.

The objective of the proposed multi-probabilistic method is, of course, to add as many valid vectors as possible, while rejecting only vectors that do not correspond to actual ocean surface currents present in the field of view.

Study of the First-Order Boundaries and the corresponding Radial Current Velocity Maps for the VCOL site at the Straits of Georgia: We make a series of statistical comparisons in order to evaluate how well the radial vectors from the two methods match with their surrounding vectors. In other words, we check:

- "Agreed Vectors": Value of the average Standard Deviation of the set of vectors that both the Conventional and the proposed Bayesian method place in the map when comparing them with their surrounding Agreed vectors.

- "Added Vectors": Value of the average Standard Deviation of the set of vectors provided by the proposed Bayesian method (that the Conventional method missed) when comparing them with their surrounding vectors.

- "Rejected Vectors": Value of the average Standard Deviation of the set of vectors that the Conventional method had provided but the proposed Bayesian method rejects.

As an example, we are going to compare the three sets of vectors presented in Figure 8.8. We are, therefore, comparing how well the added vectors on the middle panel of Figure 8.8 correlated with their neighbor vectors.

Figure 8.10 presents two panels:

1) The top panel describes the number of vectors that we are comparing in average, for each of the timestamps of our VCOL cross-spectra files. The blue and red lines correspond to the total number of vectors by each method (and the agreed ones – in green – are included in that average sample size). In this case we present the comparison of cross-spectra files from the VCOL site starting on July 1st 2020 at 00:00 until July 2nd 2020 at 7:40 am.

2) In the bottom panel, we can see those same timestamps over the horizontal axis, while the vertical axis corresponds to the average standard deviation of the vectors being compared. The average standard deviation is in the same units as the current velocity of the vectors, which is cm/s.

Overall, we observe in Figure 8.10 how the added vectors (blue dots) correlate well with the agreed vectors surrounding them. This means that the radial current velocity vectors "added" by the proposed Bayesian method are statistically as good of a match to their surroundings as the "agreed" vectors i.e. the vectors that both the Conventional and the proposed Bayesian method had placed in the radial maps. On the other hand, the vectors that the Conventional method had generated but the proposed Bayesian method rejects - the "Rejected" vectors marked as red dots - have a higher average standard deviation and don't match that well the "Agreed" vectors presented as a green line in Figure 8.10.

Particularly, Figure 8.11 includes a red circle that points to the region where more vectors have been added by the proposed method. Indeed the range bins – where those vectors within the red circle are located – correspond with the ranges in the monopole spectra where there were split Braggs (see Figures 8.5 - 8.7).

We thus conclude, that it was a good decision to remove those "Rejected" vectors and to add the "Added" vectors, since the resulting surface current radial maps present higher statistical relevance.

Finally, as we will see in Chapter 9, there will be situations where we do not see clearly that the added vectors are better correlated to the agreed as compared to the rejected. But when comparing the vector sets with the velocities determined by drifters in Chapter 9, we see that indeed the added vectors present radial velocities that match with the drifter velocities better than the rejected vectors.

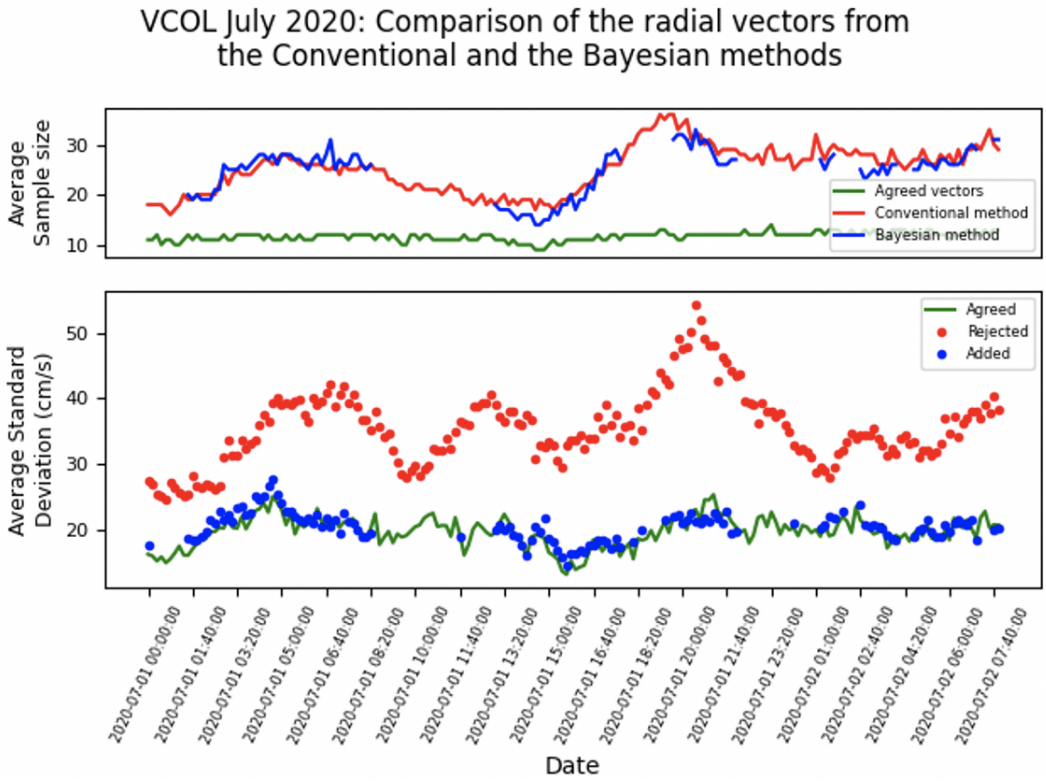

Figure 8.10: Average standard deviation comparison between vector sets. The top panel describes the amount of vectors that we are comparing in average, for each of the timestamps of the selected set of VCOL cross-spectra files (starting on July 1st 2020 at 00:00 until July 2nd 2020 at 7:40 am). In the bottom panel, we can see those same timestamps over the horizontal axis, while the vertical axis corresponds to the average standard deviation of the vectors being compared. The average standard deviation is in the same units as the current velocity of the vectors, cm/s. The line in green corresponds to the vectors placed by both the Conventional method and the proposed Bayesian method. On the other hand, the blue dots represent the vectors added by the proposed Bayesian method. Finally, the red dots correspond to the vectors determined by the conventional method but rejected by the proposed Bayesian method. We can see how the added vectors (blue dots) correlate better with the agreed vectors surrounding them, and present a lower average standard deviation. This means that the radial current velocity vectors "added" by the proposed Bayesian method are statistically significant and matching the "agreed" vectors.

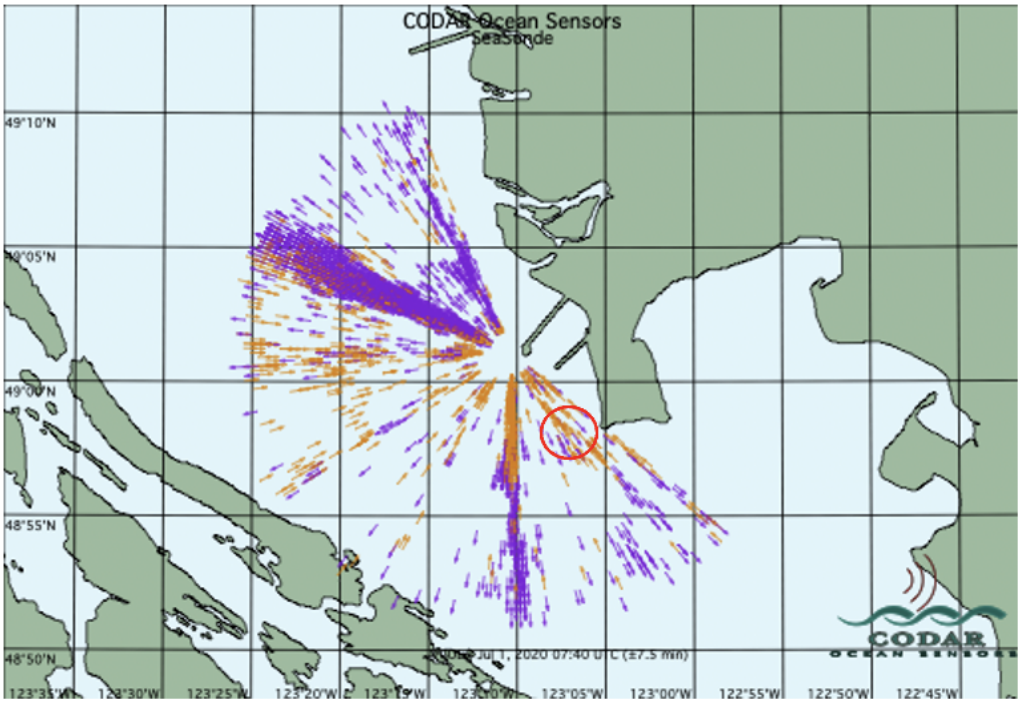

Figure 8.11: Radial current velocities map at the VCOL radar site in the Straits of Georgia from July 1, 2020 at 7:40 a.m. UTC presenting all the vectors provided by the proposed method to determine the first-order boundaries. The red circle in the figure points to the region where a higher amount of vectors have been added by the proposed method. Indeed the range bins – where those vectors within the red circle are located – correspond with the ranges in the monopole spectra where there were split Braggs (see Figures 8.5 through 8.7).

Chapter 9

Assessing the Validity of the Multi-Probabilistic Method through Comparisons with Drifters

9.1 Introduction to the UBC Drifters

A remarkable amount of fresh water enters the Strait of Georgia from the Fraser River, which then mixes with the salty ocean water in the Strait and eventually flows out into the Pacific Ocean.

Professor Rich Pawlowicz and his team at the University of British Columbia have worked for years on their "ODL Drifter Project" deploying drifters along the Straits of Georgia[54][55][56][64] in order to track the movement of the water parcels and thus determine how fast and how far they go, and to learn about turbulence of the circulation in the Strait as well.

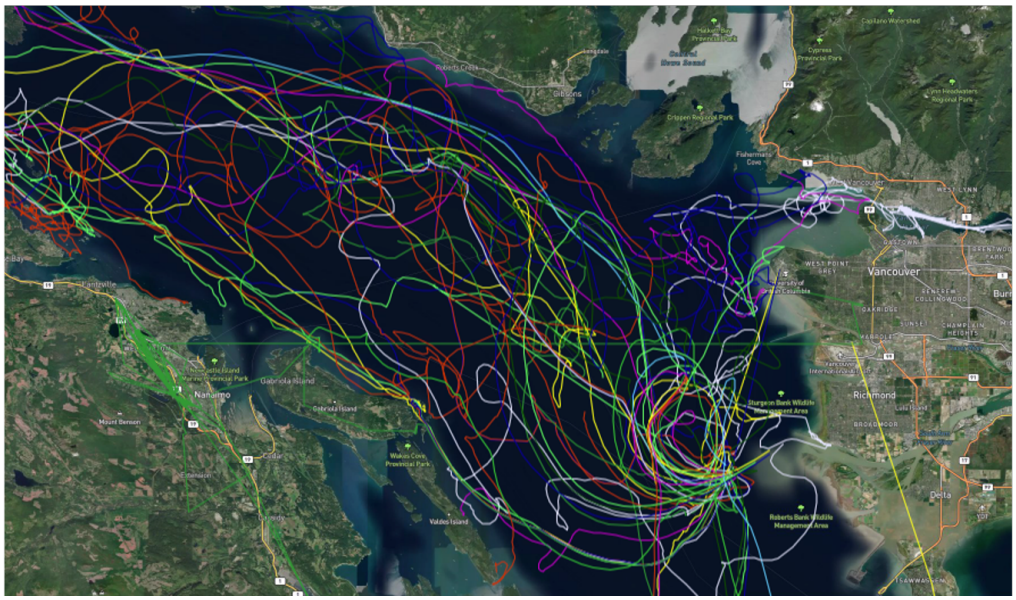

Figure 9.1: This map presents a series of drifter tracks, from several deployments along the Straits of Georgia. The buoys are mainly deployed from the south of Richmond, BC and then they will drift freely with ocean currents. As we will see in Figure 9.5, some drifters travel all the way up the Strait towards Quadra Island.

Professor Pawlowicz allowed us to study more than 400 drifter tracks dated from September 2014 until the end of March 2018, from several of their drifter deployments shown in Figure 9.1. The high-resolution expendable surface UBC drifters used were designed at the University of British Columbia and their performance was evaluated by Page et al.[53].

As shown in Figure 9.2, the UBC drifters consist of flat wooden floats above a rigid cross-shaped drogue of sheet metal about 50 cm across and 40 cm tall, their cross-sectional area is about 1800 cm^2. These drifters are optimized for measuring average ocean currents in the upper 50 cm of the water column.

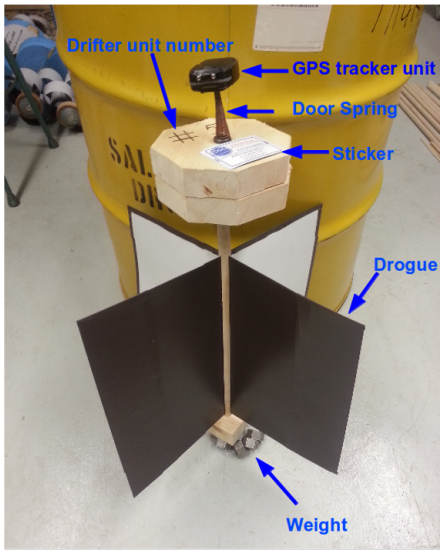

Figure 9.2: Design of a UBC Drifter [54]

9.2 Differences in Surface Current Measurements Performed by Drifters and by HF Radar

Drifters are small buoy-like instruments with GPS, which send their locations to satellites as they drift on ocean currents. Their location gets tracked for weeks or months at a time and they provide an exceptional amount of location-based data regarding surface currents. As mentioned above, the UBC drifters used in this study are optimized for measuring average currents in the upper 50 cm of the water column.

Stewart and Joy [65] used theoretical arguments to estimate that HF radars respond to velocity measurements that are an exponentially-weighted average over the uppermost $\lambda_{Bragg}/4\pi$ of the water column, approximately 0.5 m at 25 MHz. This estimate of the effective depth of HF current measurements corresponds reasonably well with the drifter effective depth of 0.5 m.

However, there are a variety of effects that could cause differences between HF radar and drifter measured currents. Carter Ohlmann[50] summarized them in several types, as follows:

1. Resolution of High-Frequency radar Doppler spectra i.e. ∼2.5 to 5 cm/s difference.

2. Vertical scale of measurements – we have a good match in this respect.

3. Horizontal scale of measurements – HF radar data is averaged over 0.25 km^2 while drifter drifts with currents over a path length of ∼0.8 km in an hour.

4. Measurements by HF radar and drifter not necessarily coincident in time.

5. Stokes drift may not be reconciled consistently among platforms – and it corresponds to about a 0.8% of wind speed. [25][48].

6. Use of "ideal" versus "measured" antenna patterns could add ∼10 cm/s error.

We have tried to limit these sources of error to the extent possible. We suspect that the largest source of error would be in the accuracy of the antenna pattern measurements. Based on previous studies [50], as we will discuss in Section 9.3, we expect that the hourly ocean current fields are likely to have random errors of about 15 cm/s, and biases around ±5 cm/s. However, there can be significant variation – in one instance, the difference increases from near zero to more than 20 cm/s within 2 hours.

Our purpose in doing a drifter comparison is to perform a sanity check to see if indeed our method of determining the Bragg first-order boundaries tends to add radials with Root Mean Squared errors (see Section 8.1) much smaller than the

RMS errors of the radials that were determined by the conventional method but are rejected by the proposed method. Here, we simply remind the reader of the typical differences between the high-frequency radar and the drifters approach and quote results of previous investigations that focused on those differences between measurements.

9.3 Drifters Dataset and Calculation of the Radial Velocities from their Trajectories

For the present research, we processed this drifter data in order to find correspondence between the locations where the drifters crossed the fields of view of the three SeaSonde sites in the area of study in the Strait of Georgia.

These sites are, from North to South: VATK at the Point Atkinson Lighthouse, VION at the Iona island and finally VCOL at the Westshore Coal Terminal as shown on the map in Figure 9.3.

Thus, the first step was to find which of the drifter tracks traveled in front of our three radar sites of interest. Unfortunately, we have very limited SeaSonde radar radial currents with correspondence in time and space with the drifter tracks. Contrary to other studies, like the ones performed by Halverson[25] or Ohlmann[50], the purpose was to validate HF radar ocean current measurements by deploying drifters on the field of view of those radar sites. In our case, conversely, the drifters were circulating freely and most of the times we do not find them crossing the coverage area of the radars, within 40 km from the shore.

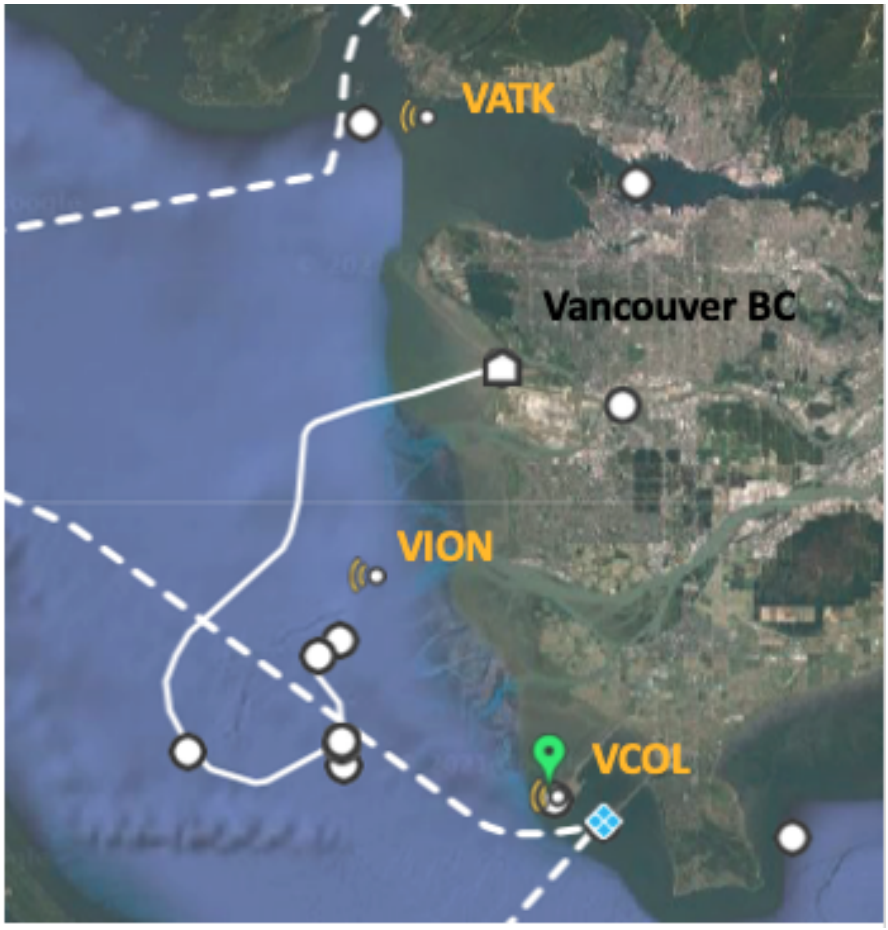

Figure 9.3: Map of the three SeaSonde locations along the Straits of Georgia and near Vancouver, British Columbia (Credit to ONC - Ocean Networks Canada[16]). Northernmost we show the VATK SeaSonde site at Point Atkinson, in the municipality of West Vancouver, BC. Next, the VION radar site is located on Iona Island near Richmond, BC. Finally, the southernmost SeaSonde radar location in the map is VCOL, at the Westshore Terminals in Delta, BC. The solid line that ONC includes in this map corresponds to a fiber-optics cable present in that location, while the dashed lines correspond to instrumented Ferry routes. On the other hand, the smaller white symbols mark the locations of stations and platforms for other instruments.

When we have a set of drifter tracks that pass nearby the radar location, we extract the velocities of each drifter from their locations at each time of measurement, i.e. their velocity on the trajectory direction. On the other hand, a single HF radar site measures line-of-sight velocities which are commonly referred to as radials. We choose to compare drifter velocities to radial velocities because radial velocities are more fundamental[25] whereas total velocities (U , V) are derived by combining radial velocities from multiple HF radar sites, where U is the eastward component and V the northward component.

First, to calculate a drifter radial velocity we need to compute the U and V components with a centered finite difference of consecutive GPS positions and times. Next, to compute radial velocities, the drifter velocities are projected on to the vector connecting the radar radial cell closest to the drifter and the radar station location. The angle of this vector is the bearing of the corresponding radial grid cell. The sign convention is such that a positive radial velocity indicates flow toward the radar site, while a negative velocity indicates flow away from the radar site.

Once we have a set of radial velocities from a given location and time, we compare them with the radial velocities obtained from the corresponding high-frequency radar site that fall into the "comparison grid" for that given time. The area of this grid has been determined by evaluating the resolution of the bins: the maximum range bin size is 0.5 km and the bearing bins resolution of 5°, as mentioned in 9.1, which corresponds to a maximum arc length of 1.2 km at the further range.

Finally, in the next Section we compare how well the drifter radial velocities within the comparison grid match with the HF radar radial velocities on that grid obtained from:

- the "conventional" method
- the proposed "Bayesian" method
- the radial vectors that both methods agree on.

9.4 Statistical Comparisons of the Radial Vectors from the HF Radars against the Corresponding Drifters

As we introduced at the beginning of this chapter, and opposed to the different comparisons evaluated in Section 8.2, there are times where the number of "added" vectors obtained from the proposed Bayesian method is minimal, when comparing to the set of vectors provided by the Conventional method.

Most of the time, this is due to the fact that the Bragg regions do not present splits and therefore when evaluating the Bayesian method results we will observe a smaller number of added radial current vectors, as compared to instances where there were split Braggs and our proposed method was able to output many more vectors. In addition, when there are no split Braggs, the difference in vectors output is mainly from the smoothing of the boundaries and the removal of outliers performed by the Bayesian method.

As presented in Figure 9.4, and as opposed to what was shown in Figure 8.10 in Chapter 8 Section 8.3, if we apply the same comparison statistics between first-order boundary determination methods for instances where we do not see relevant split Braggs, we do not observe remarkable differences on the average standard deviation of the output radial current vectors. Both sets seem to have similar characteristics. Thus, in order to evaluate the validity of the proposed method for the cases where the Bragg lines are not split and the number of added vectors is small, we compare with the series of drifters deployed along the Straits of Georgia, as described previously in this Chapter. Also the drifter comparison is a different method and provides a separate validity check. More specifically, to measure how close every set of radial vectors is to the drifters data points we calculate the Root Mean Square Error (RMSE), as described in Section 8.1.

As we will see next, the proposed method added vectors that match better with the radial vectors extracted from the drifters, as opposed to the rejected vectors that present a higher average RMSE. Even with no split Braggs, we observe that the added vectors present better correlation with their neighbors, while the vectors that the proposed Bayesian method rejects have indeed a lower correlation.

Results when comparing HF radial currents with Drifters on Deployment 21:

Figure 9.6 presents a subset of drifter tracks from Deployment 21 (full set shown in 9.5) that we compare against the radial current vectors obtained from the VATK radar site. This set of drifters corresponds with available radar data from the Point Atkinson (VATK) SeaSonde site, shown in Figure 9.3.

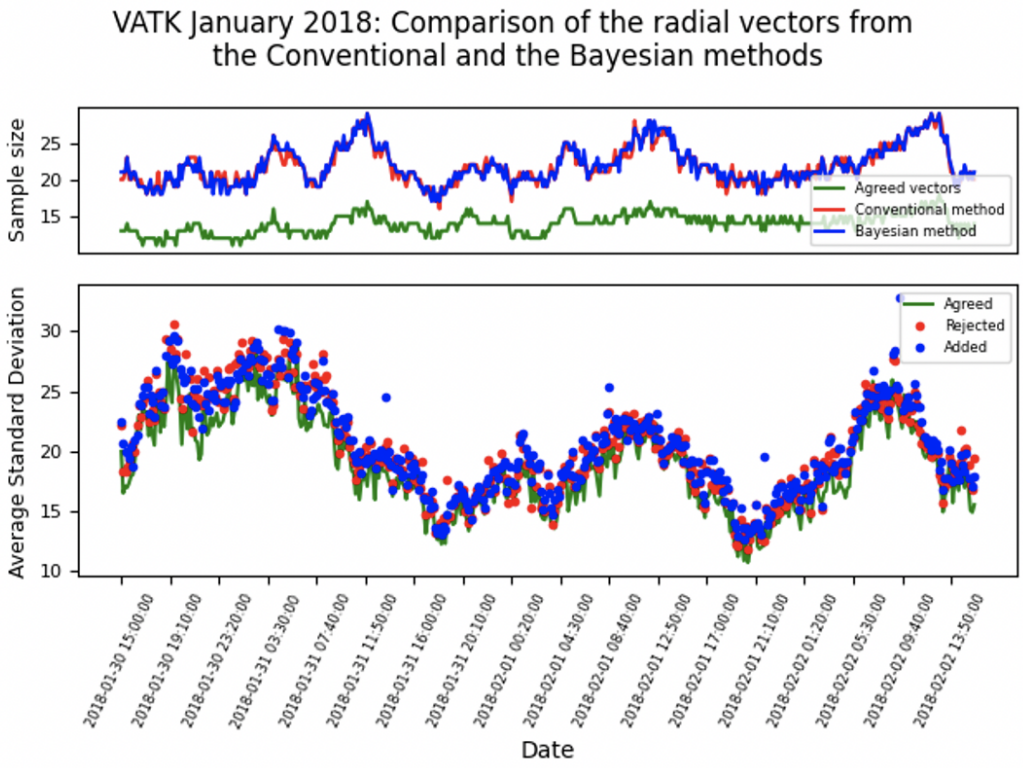

Figure 9.4: Comparison statistics for a dataset at the VATK site in Point Atkinson, West Vancouver, British Columbia. As seen, for a situation like this where there are no split Braggs in the radar backscattered cross-spectra, both of the sets of output radial current vectors present similar average standard deviation with their corresponding neighbors and therefore we can not assume that one first-order determination method would be better than the other. For this type of scenario, we apply the validation through comparing with drifter radial velocities to prove that the proposed Bayesian method provides improved results that are statistically significant.

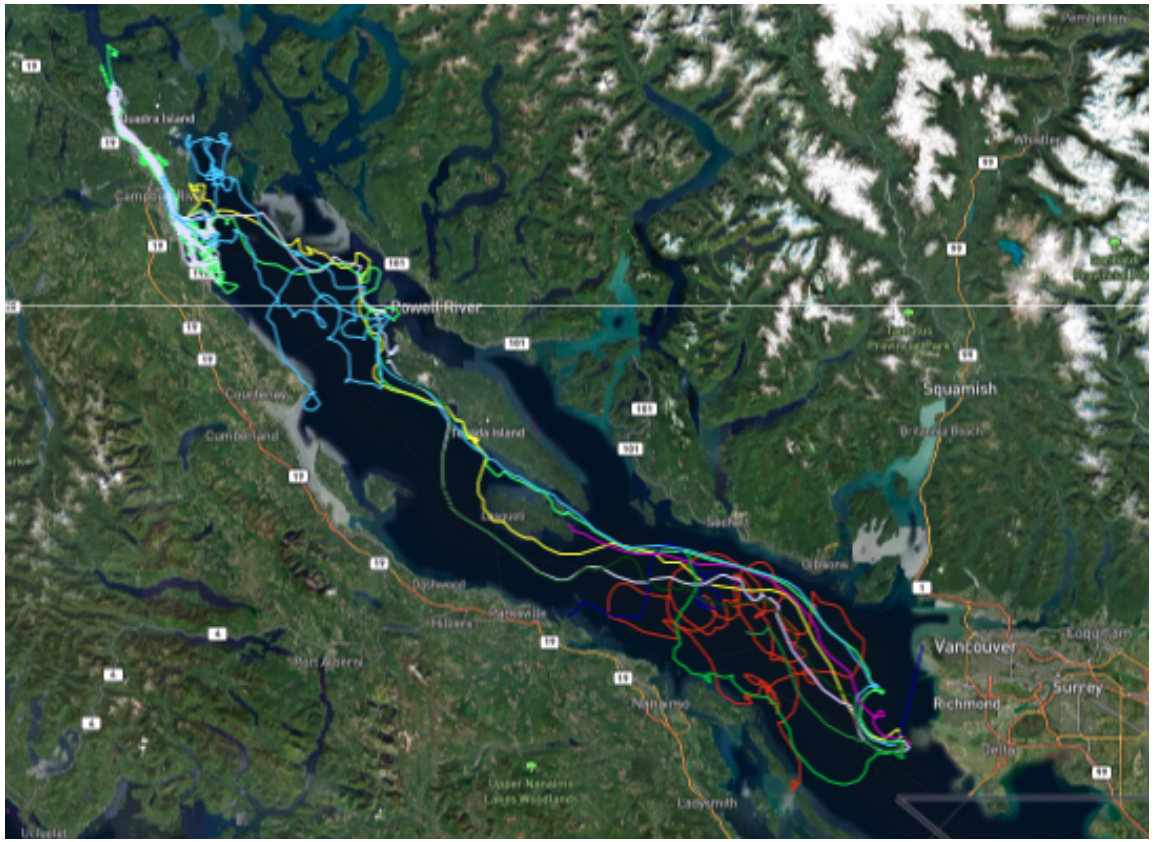

Figure 9.5: "Deployment 21" Drifter tracks. These UBC drifters were deployed from January 22 to February 20, 2018. The comparison region is near Vancouver and Richmond.

Using this match-up data set we can study the success of the proposed Bayesian method when retrieving radial ocean surface currents from the VATK site by comparing the results of the radar's radial current vectors against the radial velocities from the set of drifter tracks on Deployment 21.

The corresponding comparison results are presented in Figure 9.7, which is divided in three sections.
All three of them have a curve in common, the one with a color gradient. This multi-colored curve represents the change in radial velocity over time as detected by the UBC drifters. The color changes depending on how far the drifter was from the VATK radar site, as noted in the legend of each of the Figure sections, namely (A), (B) and (C).
We added this color change along the curve to evaluate a potential range dependency on the correlation between the radial velocity measurement from the drifters and from the radar, but we concluded that there is no apparent direct correlation.

The first section, (A), presents the comparison between the drifters radial velocities and the set of radar radial vectors that both the 'conventional' method and the proposed 'Bayesian' method agree on. As we can see, for this comparisons during January 30th 2018, the correlation of the drifters with the agreed vectors is about 0.6 and their RMSE is 16 cm/s.

The middle section, (B), of Figure 9.7 presents the same comparisons but this time the buoy radial velocities are compared against the vectors added by the proposed 'Bayesian' method. The correlation is somewhat worse, around 0.3 while the RMSE between the vectors has increased to almost 30 cm/s.

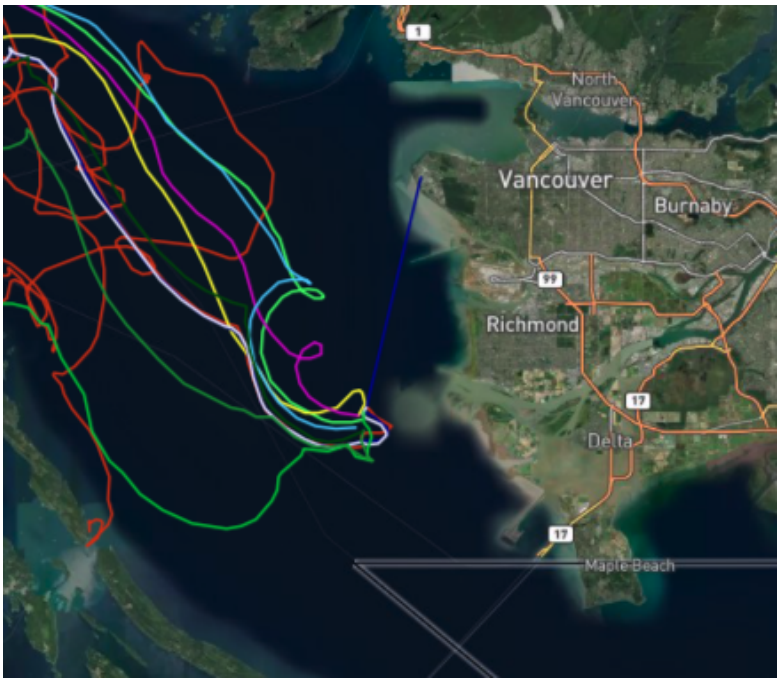

Figure 9.6: Detail of the portion of drifter tracks that travel around the VATK site and that are being compared against the radial currents provided by this SeaSonde radar.

Finally, in the third portion, (C), of 9.7 we compare the drifter radial velocities against the vectors that the 'conventional' method accepted but the proposed 'Bayesian' method rejected. We expect to see that rejecting these vectors was a good decision due to their lack of consistency with their neighbor vectors. We see, indeed, that those vectors have a very low correlation, of only 0.021, and a high Root Mean Square Error of 64.4 cm/s.

Using data from the Straits of Georgia and also when comparing with drifters, we conclude that the vectors added by the proposed Bayesian method had higher correlation with their neighbor vectors and lower deviation σ from the mean.

9.5 Discussion of the results

Previous studies comparing CODAR radial velocities to drifter-derived radial velocities have found Root Mean Square (RMS) differences ranging from 1.4 to 13.3 cm/s, and bias values ranging from -9.3 to 10.7 cm/s as described by Ohlmann[50].

Paduan et al.[51] and Lipa et al.[45], using 25 MHz radar systems like in our case, calculate RMS differences of 8.5 cm/s to 12.7 cm/s and bias values of -4 to -1 cm/s. Further results from prior work comparing HF radar and current meter measurements report similar differences as well, e.g. Janopaul et al.[30], Schott et al.[62], Graber et al.[24], Chapman et al.[18] and Emery et al.[21].

In conclusion, the studies above suggest that we should expect comparison errors as as high as 15 cm/s in RMS differences, and biases as high as ± 5 cm/s. As shown in the previous Section, even with the small number of drifter tracks that we can compare to, the RMSE of 16 cm/s seems acceptable for this type of drifter comparison and the results are within tolerance of what prior work has published. We also conclude that the vectors added by the proposed Bayesian method are statistically more significant when compared to the 'rejected' vectors which present a lower correlation. Note that the rejected radial vectors are the ones that the proposed Bayesian method rejects, but the 'conventional' method had mapped as reliable.

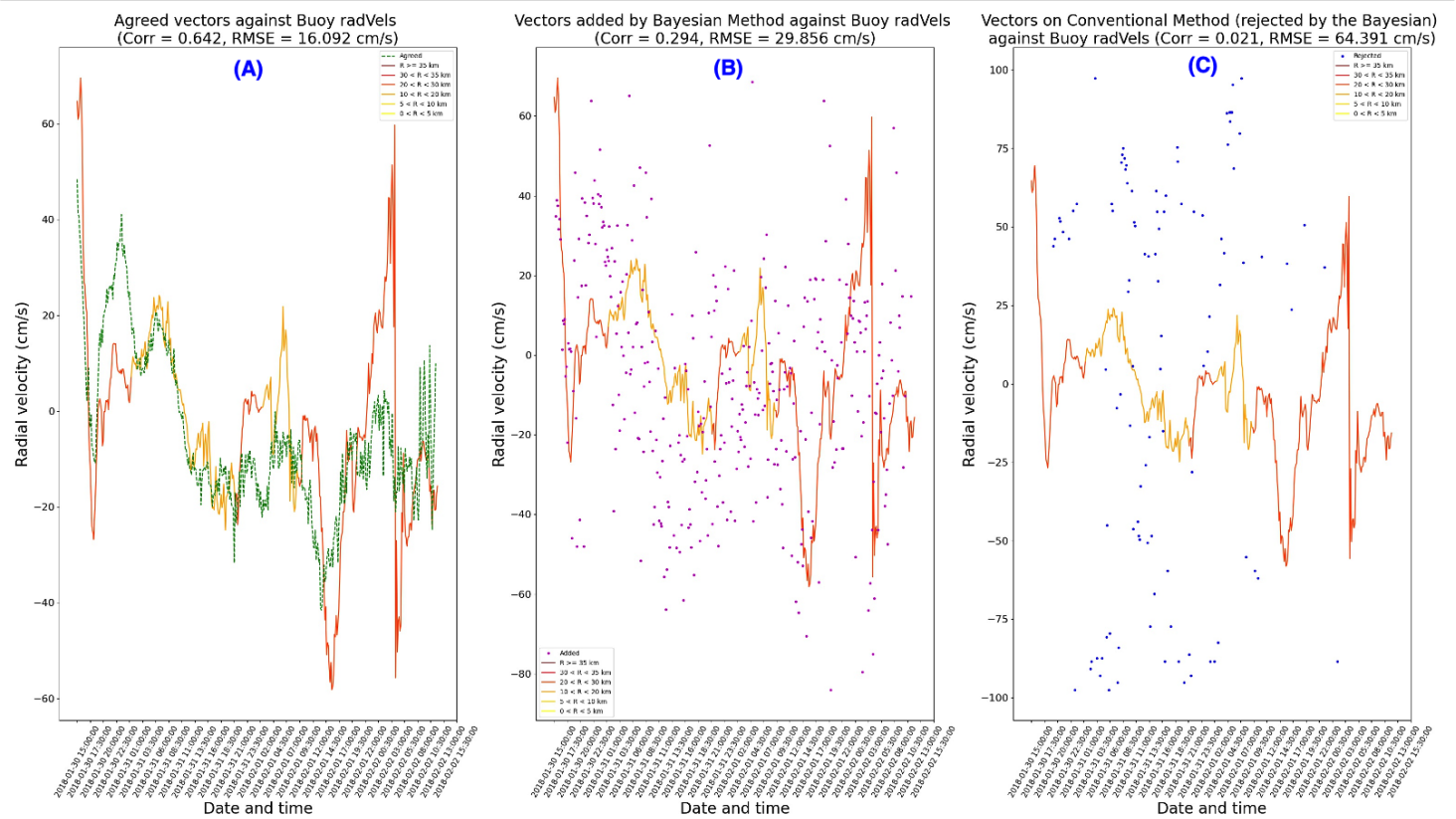

Figure 9.7: Comparison of radial ocean surface currents obtained from the VATK HF radar site, against the corresponding radial velocities from the UBC Drifters from Deployment 21 for the 4th week of January 2018.

Chapter 10

Summary, Conclusions, Contributions to Knowledge and Topics for Future Research

In this final chapter we present a summary of our study and sum up the success of using our proposed multi-probabilistic method especially in challenging scenarios. In addition, we review our contributions to knowledge, draw conclusions and propose some future research topics.

First we presented the principles of high-frequency radars and its applications to oceanography and particularly surface current mapping in **Chapter 2**.

A few fundamental radar calculations were discussed in Section 2.2 while some relevant concepts about HF radar for oceanography were presented in Section 2.3:
· On the one hand, we learned that HF radar systems estimate surface currents by analyzing the spectra of the Bragg resonant scattering region of the radar echo – returned signal–, which is also known as the first-order Bragg region.

· In addition, we reviewed Doppler frequency shift and how each Doppler bin within a constant-range bin in the radar's field of view represents the amount of energy received at that particular Doppler frequency and range.
· On the other hand, we also presented the MUSIC algorithm which determines the azimuth direction of the signals scattered off the ocean surface in every range and Doppler bin. We note that our research applies to the radar data processing stream before application of the MUSIC direction finding stage.

Next, **Chapter 3** focuses on the main concept that we study in this thesis: the first-order Bragg region boundaries and why their determination is so important. We first introduce the idea of Bragg scattering when an HF radar illuminates the ocean surface in Section 3.1 and then we describe the importance of separating the stronger first-order Bragg region from the second-order and the noise and interference in Section 3.2. Right afterwards, we discuss the challenge of having the first-order Bragg region split into multiple components (Section 3.3 – *split Braggs* phenomenon) and how that is one of the main challenges for our study described in Section 3.4 on Research Approach.

From there, we started to detail the necessary milestones to achieve our main research objective: determination of the boundaries of the first-order Bragg region in the Doppler spectrum from a constant range arc on the ocean surface. To get an overview of the process we suggest the reader review the List of Algorithms on pages xxv and xxvi of the introductory sections of this document.

I) First, in **Chapter 4** we start describing the null-finding method necessary to separate the first-order and the second-order regions of the Bragg backscatter.

The List of Algorithm Steps in Section 4.1 provides a summary of this method and we remind the reader with the list below.

1. Extract the backscatter spectrum of the monopole antenna
2. Normalize the spectrum
3. Apply a noise threshold
4. Smooth the spectrum
5. Limiting the spectral range on both sides
6. Find the possible nulls (*initial* estimates of the first-order boundaries)
7. Search for *split Bragg* regions
8. Apply Bayesian tracking in range and time
9. Extract the *final* estimates of the first-order boundaries

After all those steps are detailed in Sections 4.3 through 4.9, we present the implementation of the *conventional* null-finding method in Section 4.10 and the complete implementation of the *proposed* multi-probabilistic null-finding method in Section 4.11. We develop the conventional null finding method to use it as a standard of comparison to help evaluate our new method.

II) Secondly, in **Chapter 5** we describe the Bayesian estimation techniques applied in our tracking problem. The list Overview of Our First-Order Boundaries Tracking Problem in Section 5.1 provides an in depth review of the subject and it is provided next.

1. What are we tracking?

   The location of the first-order Bragg line boundaries i.e. what is the Doppler shift corresponding to the radial velocity of the maximum ocean current present at that time and at that range bin (Section 3.2). The four first-order boundaries (boundaries with Doppler shifts above and below both the Bragg lines associated with approaching and receding ocean waves) are tracked independently, thus we are dealing with a *single object tracking in clutter* problem (Section 5.1.2) that we will run four independent times, using data from radar observations of the Doppler spectra from each single range bin.

2. How do we track them?

   We face a state estimation problem where the object states – first-order boundaries – are to be estimated from noisy measurements.

   The tracking system consists of the *object* to be tracked, a *sensor* (the SeaSonde radar) which measures some aspect of the object, and also provides us with some *noisy measurements* (i.e. the velocity of maximum ocean current present at that time and at that range bin). Therefore we will use a Kalman Filter (Section 5.3) for this tracking problem.

3. What are our observations or measurements?

   From our *sensor* (the SeaSonde radar) we will be able to process a series of backscatter spectra files, which correspond to the specific time we are studying. We will go range bin by range bin using their respective spectra files, and find the initial estimates for each of the four first-order boundaries for each range bin spectra file. Those initial estimates are the input *noisy measurements* we will be sending to our Kalman Filter (Section 5.3).

4. What if we have multiple initial estimates for a single boundary location?

   Sometimes we will have multiple estimates for a single boundary location at a specific time and range. As we will see, we will apply the concepts of a Bayesian filter receiving *multiple simultaneous measurements* (Section 5.4) to choose between them:

   - The Nearest Neighbor Filter (first introduced in Section 5.4.2 and described in detail in Algorithm 31 in Section 6.5) would be a simple option to decide between multiple single boundary location possibilities.
   - The Probabilistic Data Association Filter (Section 5.4.3 and Algorithm 32 in Section 7.5) is a more complex filter applicable to our problem which will statistically weight and estimate a boundary location somewhere in between the multiple single boundary location possibilities we receive as measurements.

5. How do we filter the four boundaries once we have chosen the complete set of initial estimates for them?

   We realized that tracking in time and range made sense for the first order line problem, physically the maximum ocean current present should not change many cm/s (or Doppler bins) when advancing in range (about 5 km step) or moving forward in time (about 30 minutes time). This two-dimensional tracking in range and time is fully described in Chapter 7 and it is a novel application of Bayesian tracking for this first-order boundaries determination problem.

III) Combine the null-finding algorithm and the diverse Bayesian estimation techniques into a Python script that determines the first-order boundaries in **Chapter 6**. Please refer to the List of Algorithms in pages xxv and xxvi for a full chronological overview of the complete algorithm steps.

IV) Bayesian tracking application for our first-order boundaries problem, which is described in **Chapter 7**. The estimation steps are listed at the start of Chapter 7 and a review is provided next.

Overview of the Estimation Steps:

1. Set up the estimation model.
2. Prediction.
3. Gating: Not necessary for our problem since the measurements for every boundary are received separately.
4. Data Association: Step determining which of the received observations is the best match, and therefore its value will be used on the Update step. Probabilistic Data Association Filter (PDAF) to resolve multiple options for a boundary. Bayesian tracking in time and range to provide the final estimates for the boundaries.
5. Update.

Then, once the proposed method has been described in Steps I through IV, **Chapter 8** proceeds with a series of statistical comparisons to demonstrate the validity of our new multi-probabilistic method versus the conventional method. We present comparisons of determined first-order boundaries on the radar backscatter spectra as well as comparisons of the corresponding results mapped into surface current radial maps.

Finally, in **Chapter 9** we look at a validation of our method from a different aspect using HF radar and drifting buoy data collected in the Straits of Georgia. In this study we compare the radial current vectors from both line boundary determination methods against a series of radial velocities obtained from drifters that were deployed by the University of British Columbia in the Straits of Georgia. In both these test our multi-probabilistic method is as good as or better than the conventional method. In challenging situations our new method proves to be superior.

Contributions to Knowledge:

Next, we review the contributions to knowledge and draw some conclusions.

- Current surface current measurements using compact HF radars are presently handicapped, especially under challenging conditions, by inclusion of radar data that results in some erroneous radial vectors and not making use of radar data that can produce useful radial current vectors. These errors are often caused by incorrect determination of the first order Bragg Line (FOL) boundaries in the Doppler spectrum of a given range bin arc. We make two major innovations:

  - Development and application of a new null finding algorithm and a series Bayesian estimation techniques.
  - Application of probabilistic data association filtering (PDAF) to resolve multiple options in setting the first-order Bragg line (FOL) boundaries.

  Using these developments we show that we can better define the first order Bragg line (FOL) boundaries, especially under adverse circumstances (see Chapter 2) and thus include more useful surface current vectors while at the same time excluding data that results in erroneous surface current vectors. Our improvements are most useful for cases when Doppler spectra show peculiar characteristics e.g. in challenging conditions, such as

  - Split Braggs, e.g., from islands in the radar coverage area
  - Second order scatter at high sea states
  - Interference, e.g., ionospheric echoes

- We demonstrate the validity of our techniques by comparison with the conventional techniques using both surrounding current measurements and drifter measurements of surface currents. Our multi-probabilistic method added vectors that match better with the radial vectors extracted from the drifters, as opposed to the rejected vectors that present a higher average RMSE. Even with no split Braggs, we observe that the added vectors present better correlation with their neighbors, while the vectors that the proposed Bayesian method rejects have indeed a lower correlation.

- The impact of our improvements mean that compact antenna, HF radars can make measurements in some adverse conditions where useful measurements were previously impractical. Our improvements mean that oceanographic and practical applications are less susceptible to problems caused by the erroneous radial vectors distorting results. For example, surface currents over large areas for a very useful surface boundary condition for ocean circulation models. The more accurate the boundary condition, the more accurate will be the circulation model.

Conclusions:

The objective of this thesis is to devise and improved methods for selecting that portion of the radar echo data that allows valid radial current vectors to be retrieve from the radar observations of compact antenna, decameter wavelength, ground wave radars. The aspect of this issue that we address is the correct identification of the Doppler boundaries of the first-order Bragg line echoes from a range bin, arc on the ocean surface.

We conclude that the application of our multi-probabilistic method does significantly improve the correct identification of these boundaries and hence, improves the quality of the surface current, radial vectors, especially in challenging situations. Such situations are most frequent for radar observation with islands in the radar footprint, flows in straits, tidal flows, coastal upwelling and similar situations.

Future work:

Below are several topics that we think are worth pursuing to improve and validate the results of work as well as moving forward with improvements to the method.

- Further testing of our multi-probabilistic algorithm on a broader range of radar sites where problematic issues like split Braggs or islands or high sea states are present. For each different site we would set up some of the basic parameters for the estimation models corresponding to the circumstances of the site location. We think this activity could lead to a pathway with further details on how to choose the first-order configuration specifications for the radar, depending on the site and the issue it deals with.

- Investigation of our update model that is currently very simple. Although supported by general knowledge of ocean surface current variability, a more thorough investigation into regional variations could allow improvement in our update model.

- Invetigators at Codar Ocean Sensors and other users of compact ocean sensing radars continue to improve the methods for retrieval of ocean surface currents. Investigation of these improvements could lead to improvements in the multi-probabilistic method developed in this thesis. Radar data processing, such as being able to study and compare the relative phases, the relative amplitudes, the angle of arrival and the pattern distance values between Doppler bins. We anticipate that an additional layer of comparisons could help in our first-order determination when they are applied into the Bayesian correction and tracking stage.

- Application Bayesian estimation to the MUSIC direction finding step in the processing might lead to a reduction of the misplacement in azimuth angle of the radial current in a given Doppler bin. Investigation of this issue could lead to the improvement of radial current retrievals.

Bibliography

[1] D.E. Barrick. Dependence of second-order Doppler sidebands in HF sea echo upon sea state. *Antennas and Propagation Society International Symposium*, 1971.

[2] D.E. Barrick. First order theory and analysis of MF/HF/VHF scatter from the sea. *IEEE Trans. Ant. and Prop.*, AP-20:2–10, 1972.

[3] D.E. Barrick. Remote sensing of sea state by radar. *Remote Sensing of the Troposphere*, 1972.

[4] D.E. Barrick. Extraction of wave parameters from measured HF radar sea-echo Doppler spectra. *Radio Sciences*, 12:415–24, 1977.

[5] D.E. Barrick. The ocean wave height non-directional spectrum from inversion of the HF sea-echo Doppler spectrum. *Remote Sensing of the Environment*, 6:201–27, 1977.

[6] D.E. Barrick. Accuracy of parameter extraction from sample-averaged sea-echo Doppler spectra. *IEEE Transactions on Antennas and Propagation*, 28(1):1–11, 1980.

[7] D.E. Barrick. Comparison of Direction-Finding and Beam-Forming in HF Radar Current Mapping. *National Oceanic and Atmospheric Administration*, Phase 1 SBIR Final Report. Contract No. 50-DKNA-5- 00092. Rockville, MD., 1996.

[8] D.E. Barrick, M.W. Evans, and B.L. Weber. Ocean Surface Currents Mapped by Radar. *Science*, 198(4313):138–144, 1977.

[9] D.E. Barrick, J.M. Headrick, R.W. Bogle, and D.D. Crombie. Sea backscatter at HF: Interpretation and utilization of the echo. *Proceedings of the IEEE*, 62(6):673–680, 1974.

[10] D.E. Barrick and B.J. Lipa. Evolution of Bearing Determination in HF Current Mapping Radars. *Oceanography*, 1997.

[11] D.E. Barrick and B.J. Lipa. Radar angle determination with MUSIC direction finding. *US Patent 5,990,834-A*, 1997.

[12] D.E. Barrick and W.H. Peake. Extraction of wave parameters from measured HF radar sea-echo Doppler spectra. *Scattering from surfaces with different roughness scales: Analysis and interpretation*, Battelle Res. Rep. BAT-197A-10-3, Columbus, Ohio., 1967.

[13] D.E. Barrick and W.H. Peake. A review of scattering from surfaces with different roughness scales. *Radio Sciences*, 3:865–8, 1968.

[14] P.R. Bevington and D.K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, 2003.

[15] Encyclopedia Britannica. Electromagnetic Radiation and Radio Waves. `https://www.britannica.com/science/electromagnetic-radiation/Radio-waves`.

[16] Ocean Networks Canada. Strait of Georgia ONC Observatories. `https://www.oceannetworks.ca/observatories/pacific/strait-georgia`, 2021.

[17] S. Challa, M.R. Morelande, D. Musicki, and R.J. Evans. *Fundamentals of Object Tracking*. Cambridge University Press, 2011.

[18] R.D. Chapman, L.K. Shay, H.C. Graber, J.B. Edson, A. Karachintsev, C.L. Trump, and D.B. Ross. On the accuracy of HF radar surface current measurements: Intercomparisons with ship-based sensors. *Journal of Geophysical Research: Oceans*, 102(C8):18737–18748, 1997.

[19] D.D. Crombie. Doppler Spectrum of Sea Echo at 13.56 Mc./s. *Nature*, 175:681–682, 1955.

[20] T. De Paolo, E. Terrill, and A. Kirincich. Improving SeaSonde Radial Velocity Accuracy and Variance Using Radial Metrics. *Proceedings of the OCEANS MTS/IEEE Conference, Genova*, 2015.

[21] B. Emery, L. Washburn, and J. Harlan. Evaluating Radial Current Measurements from CODAR High-Frequency Radars with Moored Current Meters. *Journal of Atmospheric and Oceanic Technology*, 21:1259–1271, 2004.

[22] A.S. Frisch and B.L. Weber. A new technique for measuring tidal currents by using a two-site HF Doppler Radar System. *Journal of Geophysical Research*, 85:485–493, 1980.

[23] G. Gopalakrishnan and A. F. Blumberg. Surface Currents in Raritan Bay, New Jersey: Importance of HF Radar First-order Doppler Settings. *Journal of Operational Oceanography*, 4, 1-13, 2011.

[24] H.C. Graber, B.K. Haus, R.D. Chapman, and L.K. Shay. HF radar comparisons with moored estimates of current speed and direction: Expected differences and implications. *Journal of Geophysical Research: Oceans*, 102(C8):18749–18766, 1997.

[25] M. Halverson, J. Gower, and R. Pawlowicz. Comparison of drifting buoy velocities to HF radar radial velocities from the Ocean Networks Canada Strait of Georgia 25 MHz CODAR array. *Canadian technical report of hydrography and ocean sciences, 1488-5417 ; 319*, 2018.

[26] A. Harrison. *Introduction to Radar using Python and MATLAB*. Artech House, 2020.

[27] K. Hasselmann. Determination of Ocean Wave Spectra from Doppler Radio Return from the Sea Surface. *Nature Physical Science*, 229:16–17, 1971.

[28] M. Hubbard, D. Barrick, N. Garfield, J. Pettigrew, C. Ohlmann, and M. Gough. A New Method for Estimating High-Frequency Radar Error Using Data from Central San Francisco Bay. *Ocean Science Journal*, 1-12, 2013.

[29] IOOS. The U.S. Integrated Ocean Observing System. `https://ioos.noaa.gov`.

[30] M.M. Janopaul, P. Broche, J.C. De Maistre, H.H. Essen, C. Blanchet, G. Grau, and E. Mittelstaedt. Comparison of measurements of sea currents by HF radar and by conventional means. *International Journal of Remote Sensing*, 3(4):409–422, 1982.

[31] Antony Joseph. *Measuring Ocean Currents : Tools, Technologies, and Data*. Elsevier, 2013.

[32] R.E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*, 82(D):35–45, 1960.

[33] S.Y. Kim. Quality Assessment Techniques Applied to Surface Radial Velocity Maps Obtained from High-Frequency Radars. *Journal of Atmospheric and Oceanic Technology*, 32-10:1915–1927, 2015.

[34] S. Kingsley and S. Quegan. *Understanding Radar Systems*. SciTech Publishing Inc., 1999.

[35] S. Kingsley and S. Quegan. *Understanding Radar Systems*. Electromagnetics and Radar. Institution of Engineering and Technology, 1999.

[36] A. Kirinchich, T. De Paolo, and E. Terrill. Improving Estimates of Surface Currents Using Quality Metrics, with Applications to the MVCO High-Resolution Radar System. *Journal of Atmospheric Ocean Tech.*, 29, 1377-1389, 2012.

[37] A. Kirincich. Improved Detection of the First-Order Region for Direction-Finding HF Radars Using Image Processing Techniques. *Journal of Atmospheric and Oceanic Technology*, 34(8):1679–1691, 2017.

[38] R. Labbe. Kalman and Bayesian Filters in Python. `https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python`.

[39] K. E. Laws, J. F. Vesecky, and J. D. Paduan. Error assessment of HF radar-based ocean Current Measurements: An Error Model Based on Sub-Period Measurement Variance. *Proceedings of IEEE/OES Current, Waves, and Turbulence Measurement Conference*, 1-7, 2011.

[40] B.J. Lipa. Inversion of second-order radar echoes from the sea. *Journal of Geophysical Research*, 83:959–962, 01 1978.

[41] B.J. Lipa and D.E. Barrick. Least-squares methods for the extraction of surface currents from CODAR crossed-loop data: Application at ARSLOE. *IEEE Journal of Oceanic Engineering*, OE-8:226–253, 1982.

[42] B.J. Lipa, D.E. Barrick, and C. Whelan. A Quality Control Method for Broad-Beam HF Radar Current Velocity Measurements. *Journal of Marine Science and Engineering*, 7(4), 2019.

[43] B.J Lipa, M. Daugharty, M. Fernandes, D.E. Barrick, A. Alonso-Martirena, H. Roarty, J. Dicopoulos, and C. Whelan. Developments in Compact HF-Radar Ocean Wave Measurement. *Physical Sensors, Sensor Networks and Remote Sensing*, 2018.

[44] B.J. Lipa and B. Nyden. Directional wave information from the SeaSonde. *IEEE Journal of Oceanic Engineering*, 30:221–231, 2005.

[45] B.J. Lipa, C. Whelan, B. Rector, and B. Nyden. HF Radar Bistatic Measurement of Surface Current Velocities: Drifter Comparisons and Radar Consistency Checks. *Remote Sensing*, 1(4):1190–1211, 2009.

[46] CODAR Ocean Sensors Ltd. SeaSonde Operator Manual.

[47] C. Mantovani et al. Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement,. *Frontiers in Marine Science*, 2020.

[48] Y. Mao and M.L. Heron. The influence of fetch on the response of surface currents to wind studied by Hf ocean surface radar. *Journal of Physical Oceanography 38 (5), 1107-1121*, 2008.

[49] J. McMillan and Sang Seok Lim. Data Association Algorithms for Multiple Target Tracking. *Defense Research Establishment Ottawa*, page 37, 07 1990.

[50] C. Ohlmann, P. White, L. Washburn, B. Emery, E. Terrill, and M. Otero. Interpretation of Coastal HF Radar–Derived Surface Currents with High-Resolution Drifter Data. *Journal of Atmospheric and Oceanic Technology, 24(4), 666-680*, 2007.

[51] J. Paduan, K. Kim, M. Cook, and F. Chavez. Calibration and Validation of Direction-Finding High-Frequency Radar Ocean Surface Current Observations. *Oceanic Engineering, IEEE Journal of*, 31:862 – 875, 11 2006.

[52] J.D. Paduan and H.C. Graber. Introduction to High Frequency Radar: reality and myth. *Oceanography*, 10:36–39, 1997.

[53] S. Page, C.G. Hannah, T. Juhász, D.J. Spear, and H. Blanken. Surface circulation tracking drifter data for the Kitimat Fjord system in northern British Columbia and adjacent continental shelf for April 2014 to July 2016. *Canadian Data Report of Hydrography and Ocean Sciences, 1488-5433 ; 206*, 2019.

[54] R. Pawlowicz. The Grounding of Floating Objects in a Marginal Sea. *Journal of Physical Oceanography*, 51(2):537–551, 2021.

[55] R. Pawlowicz and B. Boufford. Atlas of oceanographic conditions in the Strait of Georgia. `http://sogdatacentre.ca/atlas`.

[56] R. Pawlowicz, C.G. Hannah, and A. Rosenberger. Lagrangian observations of estuarine residence times, dispersion, and trapping in the Salish Sea. *Estuarine, Coastal and Shelf Science*, 225:106246, 2019.

[57] A.M. Peterson, E.E. Teague, and G.L. Tyler. Bistatic radar observation of long-period directional ocean-wave spectra with LORAN-A. *Science*, 170:158–61, 1970.

[58] H. Roarty, J. Kerfoot, J. Kohut, S. Glenn, C. Whelan, and M. Hubbard. Improving the Measurements of High Frequency Radar: Reduced Averaging Times and Bistatics. *2013 MTS IEEE OCEANS Bergen*, 2013.

[59] H.J. Roarty, E. Rivera Lemus, E. Handel, S.M. Glenn, D.E. Barrick, and J. Isaacson. Performance Evaluation of SeaSonde High-Frequency Radar for Vessel Detection. *Marine Technology Society Journal*, 45(3):14–24, 2011.

[60] Scarlett Rose. Why is Python Programming a perfect fit for Big Data? https://medium.com/@scarlett8285/why-is-python-programming-a-perfect-fit-for-big-data-5ac54ee8f95e.

[61] R.O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34:276–280, March 1986.

[62] F.A. Schott, S.A. Frisch, and J.C. Larsen. Comparison of surface currents measured by HF Doppler radar in the western Florida Straits during November 1983 to January 1984 and Florida Current transports. *Journal of Geophysical Research: Oceans*, 91(C7):8451–8460, 1986.

[63] L.K. Shay, H.C. Graber, D.B. Ross, and R.D. Chapman. Mesoscale Ocean Surface Current Structure Detected by High-Frequency Radar. *Journal of Atmospheric and Oceanic Technology*, 12(4), 1995.

[64] S.W. Stevens, R. Pawlowicz, and S.E. Allen. A study of the intermediate water circulation in the Strait of Georgia using tracer-based, Eulerian, and Lagrangian methods. *Journal of Physical Oceanography, 51 (6) 1875-1893*, 2021.

[65] R.H. Stewart and J.W. Joy. HF radio measurements of surface currents. *Deep Sea Research and Oceanographic Abstracts*, 21(12):1039–1049, 1974.

[66] E.E. Teague, G.L. Tyler, J.W. Joy, and R.H. Stewart. Synthetic aperture observations of directional height spectra for 7 s ocean waves. *Nature Phys. Sci.*, 224:98–100, 1973.

[67] William J. Teague, Dong S. Ko, Gregg A. Jacobs, Henry T. Perkins, Jeffrey W. Book, Scott R. Smith, Kyung-Il Chang, Moon-Sik Suk, Kuh Kim, Sang Jin Lyu, and Tswen Y. Tang. Currents Through the Korea/Tsushima Strait: A Review of LINKS Observations. *Oceanography*, 2006.

[68] J.R. Wait. Theory of HF ground wave backscatter from sea waves. *Journal of Geophysical Research*, 71:4839–42, 1966.

[69] C. Whelan, M. Hubbard, D. Trockel, and H. Parikh. Benefits of Multiple Antenna Pattern Measurement Methods for Maintaining a Regional HF Radar Network. *Proceedings of the ORCA Conference, Okinawa, Japan*, 2018.

[70] C. Wolff. Radar Tutorial. https://www.radartutorial.eu/index.en.html.

[71] L.R. Wyatt. The measurement of the ocean wave directional spectrum from HF radar Doppler spectra. *Radio Science*, 21:473–485, 1986.

[72] L.R. Wyatt. Ocean wave parameter measurements using a dual-radar system: a simulation study. *International Journal of Remote Sensing*, 8:881–891, 1987.

[73] L.R. Wyatt. Significant waveheight measurement with HF radar. *International Journal of Remote Sensing*, 9:1087–1095, 1988.

[74] L.R. Wyatt. A relaxation method for integral inversion applied to HF radar measurement of the ocean wave directional spectrum. *International Journal of Remote Sensing*, 11:1481–1494, 1990.

[75] L.R. Wyatt. HF radar measurements of the ocean wave directional spectrum. *IEEE Journal of Oceanic Engineering*, 16:163–169, 1991.

[76] L.R. Wyatt. The effect of fetch on the directional spectrum of Celtic Sea storm waves. *Journal of Physical Oceanography*, 25:1550–1559, 1995.

[77] L.R. Wyatt. The ocean wave directional spectrum. *Oceanography*, 10:85–89, 1997.