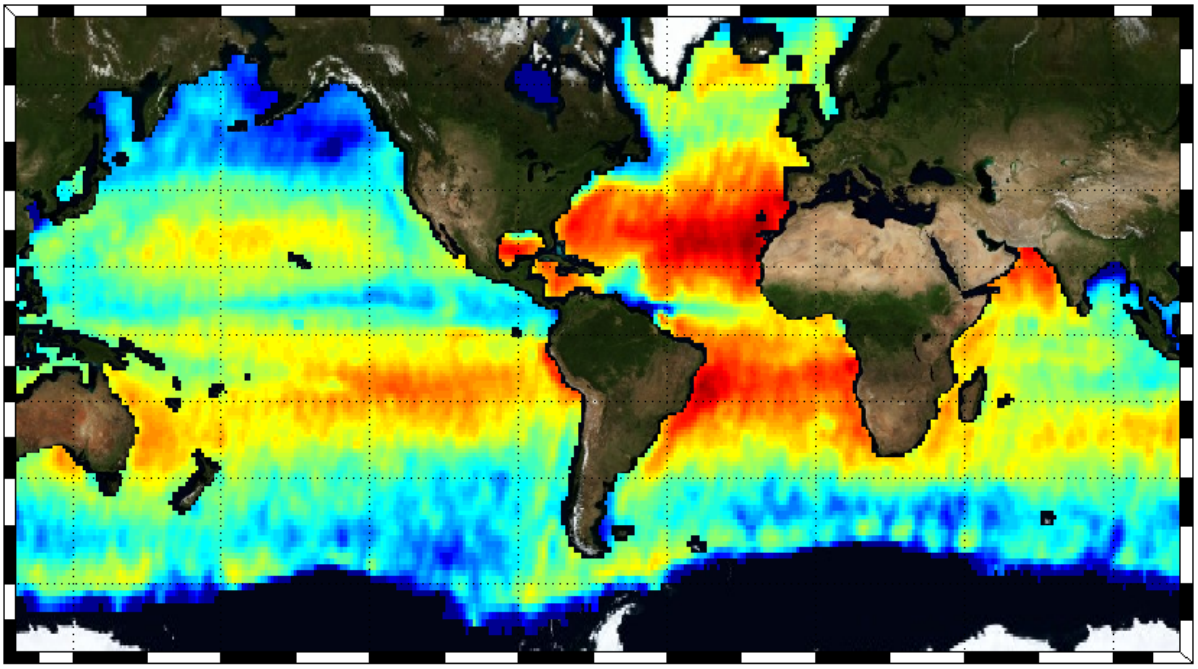
CHALMERS



Improving the emissivity model of rough water with GNSS-reflectometry

Correlation of L-band radiometric measurements with GPS-R data

Master of Science Thesis

Caroline MANCEL

Department of Earth and Space Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2011

Improving the emissivity model of rough water with GNSS-reflectometry.
Correlation of L-band radiometric measurements with GPS-R data.

Caroline MANCEL

Supervisor:
Dr. Juha Kainulainen
Department of Radio Science and Engineering
School of Electrical Engineering, Aalto University
02150 Espoo
Finland
Telephone +358 (0)9-470-24778

Examiner:
Dr. Leif Eriksson
Department of Earth and Space Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
Global map of sea surface salinity processed by Ifremer, from August 2010 SMOS data
http://www.salinityremotesensing.ifremer.fr/activities/smos/data/l3

Department of Earth and Space Sciences
Göteborg, Sweden 2011

Abstract

Remote sensing of sea surface salinity (SSS) is a challenging task. Variations of salinity in oceans are very small, but even minor changes influence the global circulation of water, thus the environment and climate. Those variations are therefore to be monitored with the highest sensitivity possible, and one way to do this is by means of radiometry. Unfortunately, the thermal radiation that is emitted by the sea, and from the measurement of which salinity can be estimated, does not change dramatically with SSS gradients, hence the difficulty to map the global salinity distribution with high accuracy. What's even more important is the roughness of the sea: increasing roughness leads to additional thermal radiation from the sea, but in a way that is not known and not properly accounted for. As a result, the theory used to model the emissivity of water needs empirical corrections, which can be made by investigating in which manner, and to which extent, the emission of the sea changes with geophysical parameters that characterize this roughness. Those parameters are brought by studying how the signals of the Global Navigation Satellite System (GNSS) are modified when they are reflected at the surface, and methods used to do so define the field of GNSS-reflectometry (GNSS-R).

This thesis investigates how the additional emission of the sea due to its roughness is possibly correlated to GNSS-R observables. The data considered were collected on two different airborne campaigns, for which the instrumentation onboard was however the same (same radiometer and same GNSS reflectometer). Another campaign was organized during the thesis (different instrumentation onboard though); data of the latter will be processed in the close future, meaning that the work done during this thesis will lead to further research on the same topic. As for the results obtained from the other campaigns, they show that the emission of the sea that is due to the surface roughness is always a linear function of this roughness when correlation is observed. This observation was made for five GNSS-R observables, out of seven investigated in total. Among those, one is computed from the combination of GNSS-R measurements with a statistical description of the sea waves: the Mean Squared Slope (MSS). Although correlation is observed for five different observables, it turns out that it is clearest when MSS estimates are considered, indicating that the roughness of the sea is better characterized when GNSS-R measurements are combined with statistical information.

Foreword

This thesis was written for my Master's degree in Radio and Space Science at Chalmers University of Technology (Gothenburg, Sweden). The research was made at Aalto University (Helsinki, Finland), at the department of Space and Radio Engineering, as part of a project involving several international institutes. In this thesis, I demonstrate how the remote sensing of the sea surface salinity can be enhanced by means of GNSS-reflectometry techniques (GNSS acronym stands for Global Navigation Satellite System). GNSS-reflectometry consists of studying how affected the signals of the Global Navigation Satellite System are when they are reflected at the sea surface. Such information help to better model the emissivity of water at L-band – a model which is quite solid when the sea surface is smooth, but unreliable when conditions are rougher. This project thus involves, in addition to Aalto University, the European Space Agency (ESA), which launched the Soil Moisture and Ocean Salinity satellite (SMOS) in 2009, and the Catalonian Institute for Space Studies (IEEC), which do research on GNSS-reflectometry techniques.

I would like to thank the following people for their help and time. My sincere gratitude goes to my supervisor Juha Kainulainen for his guidance, suggestions, and trust, thorough the whole thesis. More generally, I would like to thank my colleagues of Aalto University for making me feel at ease in this new environment from the very beginning. I would also like to thank people from the Finnish Meteorological Institute (FMI) and IEEC, whether they gave me information, or helped me in any way, and especially Fran Fabra, from IEEC, for his continuous practical support with GNSS-reflectometry data. Last but certainly not least, I would like to thank Leif Eriksson, my supervisor at Chalmers, for his support, advice, and devoted time.

Table of Contents

Table of figures

Acronyms

C/A	Coarse Acquisition
CMB	Cosmic Microwave Background radiation
DDM	Doppler Delay Map
DM	Delay Map
DTU	Danmarks Tekniske Universitet
ESA	European Space Agency
FMI	Finnish Meteorological Institute
GNSS	Global Navigation Satellite System
GNSS-R	GNSS-Reflectometry
GOLD-RTR	GPS Open Loop Differential Real-Time Receiver
GPS	Global Positioning System
GPS-R	GPS-Reflectometry
HPBW	Half Power Beam Width
IEEC	Institut d'Estudis Espacials de Catalunya
LTE	Local Thermodynamic Equilibrium
MSS	Mean Squared Slope
PRN	Pseudo Random Noise
RAD	RADio and space engineering
RFI	Radio Frequency Interference
SMOS	Soil Moisture and Ocean Salinity
SNR	Signal-to-Noise Ratio
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
TIR	Thermal Infrared Radiometer

1. Introduction

Remote sensing of sea surface salinity (SSS) is a challenging task. Variations of salinity in oceans are very small, but even minor changes influence the global circulation of water, thus the environment and climate. Those variations are therefore to be monitored with the highest sensitivity as possible, by means of radiometry. Unfortunately, the thermal radiation that is emitted by the sea, and from the measurement of which salinity can be estimated, does not dramatically change with SSS, hence the difficulty to map the global salinity distribution with high accuracy. What's more is the roughness of the sea: increasing roughness leads to additional thermal radiation from the sea, but in a way that is not known and not properly accounted for. Therefore, the theory used to model the emissivity of water needs empirical corrections, which can be made by investigating in which manner, and to which extent, the emission of the sea changes with geophysical parameters that characterize its roughness. Those are brought by studying how the signals of the Global Navigation Satellite System (GNSS) change when they are reflected at the surface, and are obtained from so-called GNSS-reflectometry (GNSS-R) techniques.

In chapter 2, a very brief outline of the thesis' objectives is given. In chapter 3, theoretical background and general methodology about GNSS-R and radiometry are explained. Chapters 4 and 5 detail what the starting point of the thesis actually was. In chapter 6, I present the instruments used by to study how radiometry and GNSS-R can be combined together, while in chapter 7, I describe the flight experiments that were consequently organized. In chapter 8, the methodology used to process the data is described and results are presented and discussed in chapter 9. Last, I conclude on the work done during this thesis in chapter 10.

2. Aims of the thesis

In this thesis, the contribution of roughness to L-band radiometric measurements of sea surface was to be studied. To reach that goal, the idea was to investigate possible correlation between the thermal emission of the sea and a couple GNSS-R observables that describe the roughness conditions of the sea.

In parallel to the pure research part of the thesis, a radiometric field experiment is to be organized in the Gulf of Finland, with Aalto's Skyvan plane. This more experimental part of the thesis is not negligible since it permits to seize the various aspects of radiometry as a whole, namely:

- to actually see what is measured, and how it is measured,
- to take part in the experiment,
- to handle, process, and analyze the various ground reference data, which is required to process the radiometric measurements.

3. Theory

3.1.GNSS-reflectometry

3.1.1. Principle of GNSS-reflectometry

Signals of the Global Navigation Satellite System (GNSS) are signals that are emitted by a system of satellites in order to provide geo-spatial positioning and reference timing for end-users on Earth. There are four systems of that kind, either completed – such as GPS (United States of America) and GLONASS (Russia) - or in development – such as GALILEO (European Union) and COMPASS (China). In order to ensure global coverage, around 30 satellites are usually used for one system. Frequencies used for GNSS signals belong to L-band (1 to 2GHz). More specifically, the GPS system, which is the most developed and used system, consists of 31 satellites, each broadcasting signals at two frequencies: L1 (1.57542 GHz) for civilian utilization, and L2 (1.2276 GHz) for military utilization. Without entering into too many details, each satellite can be identified by the Pseudo-Random-Noise (PRN) of the received signal, which is a sequence encoded in the message data. Although mostly used for positioning, navigation or timing, GNSS signals can be used for other purposes; remote sensing of the Earth is one example.

The basic idea of GNSS-reflectometry is to analyze on an elevated platform (aircraft, balloon, satellite, etc.) the signal emitted by a GNSS satellite after it has been reflected off at the Earth's surface, and compare it with the direct signal received from this same satellite (see Figure 1). If distortion of the reflected signal is analyzed, one can learn about the geophysical characteristics of the surface; and if ray-path lengths of the signals are compared, one can perform altimetric measurements.

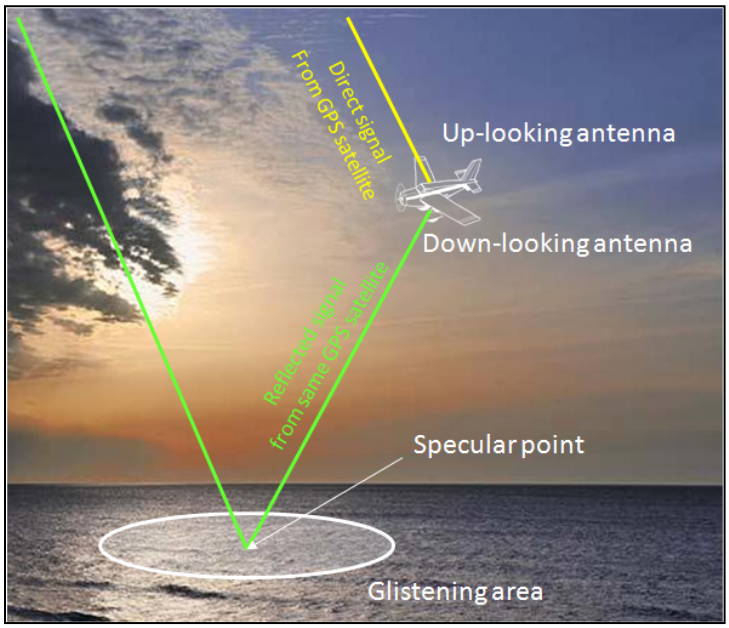


Figure 1: Illustration of GNSS-reflectometry setup

The concept of GNSS-reflectometry was first presented in 1993, by Manuel Martin-Neirà on behalf of the European Space Agency (ESA). Initially, the idea was to use GNSS-R as a low-cost substitute to satellite altimetry. Since then, the amount of possible applications has increased. Soil moisture, sea surface state (salinity, roughness and temperature), or wind speed, are only a few of the geophysical parameters where the normal remote sensing observations can be enhanced, when combined with GNSS-R. Several reflectometry experiments, whether they are organized for demonstration, calibration, rehearsal or validation campaigns, have taken place since 1993, and already two satellites using GNSS-reflectometry have been launched (SMOS in 2009, and Aquarius in 2011).

3.1.2. Into more details

The GNSS-reflectometry technique that is used in this master's thesis is based on the analysis of the so-called waveform, which is the squared magnitude of the correlation function of the GNSS signal, reflected or direct, with its model replica (expected signal). The main distinction that can be made between direct and reflected GPS signals is indeed the shape of that function. If the direct signal is correlated with a clean replica, the waveform resembles the triangle function. If the reflected signal is correlated with its expected model (all atmospheric and propagation delays taken into account), the waveform spreads out, because the scanned surface is not a perfect flat surface. To generate proper model replicas, the GNSS satellite from which the signals come from has to be perfectly identified (PRN, frequency…).

If the water surface is flat, the reflection is of the specular type. By definition, the specular point is the point of shortest ray-path length, which corresponds to a transmitter-receiver time delay t_0. Let's consider the correlation $\rho_{s,prn}$ of the received (reflected) signal s with its model replica prn, for an expected specular delay t_0. It writes:

$$\rho_{s,prn} = \int_0^{T_i} s(t') \cdot prn(t_0 + t') \cdot e^{-2j\pi f_c t'} dt'$$

- f_c is the carrier frequency of the signal
- T_i is the coherent integration time

Contributions that arrive at delays t such as $t - t_0 > \tau_c$ are filtered out in the computation. This condition draws a narrow ellipse centered on the specular point, the first iso-range ellipse (or first Fresnel zone).

If the surface is not flat, as it is often the case, the signal is scattered over a more or less wide area: the glistening zone. The receiver does not collect contributions from the specular point solely, but from the whole glistening zone, at longer relative time delays than in the case of a flat surface (see Figure 2). If the glistening area is wider than the first iso-range ellipse, contributions from outside this ellipse are filtered out, as explained above. To compensate for this, and also to increase the Signal-To-Noise (SNR) ratio, correlation is computed for

several time-shifted replicas. The squared modulus of the correlation function $Y(\delta t)$ builds the so-called waveform, where:

$$Y(\delta t) = \rho_{s,prn}(t_0 + \delta t) = \int_0^{T_i} s(t') \cdot prn(t_0 + \delta t + t') \cdot e^{-2j\pi f_c t'} dt'$$

- $prn(t_0 + \delta t + \tau)$ is the time-shifted replica of the GPS L1 coarse acquisition (C/A) code

A waveform is the assembly of several correlation coefficients $\{Y(\delta t)\}_{\delta t \in \mathrm{T}}$, computed at different time delays $\delta t \in \mathrm{T}$, each carrying the contribution of one particular annulus of "width" τ_c centered on the specular point (see Figure 3). A number of parameters can be produced from its analysis, among which the peak power, the delay of the peak with respect to the expected delay, and many more (more details are given in section 6.2.3). Those GNSS-R observables refer to the so-called Delay Maps (DM), as opposed to those referring to the so-called Delay Doppler Maps (DDM).

If the surface is flat, the shape of the correlation function is rather close to that of the ideal triangle function: the peak of the waveform is quite sharp (see

Figure 2, and please note that in reality, even the direct signal does not correlate as perfectly as what is presented on the upper graph of the figure). If the surface is not flat, the peak of the waveform is lower and the overall shape is flattened over time. The term "spreading" is commonly used in such a case (and it is used also in the thesis), although it is actually wrong; power is not really spread, it is instead the result of additional contributions to the waveform.

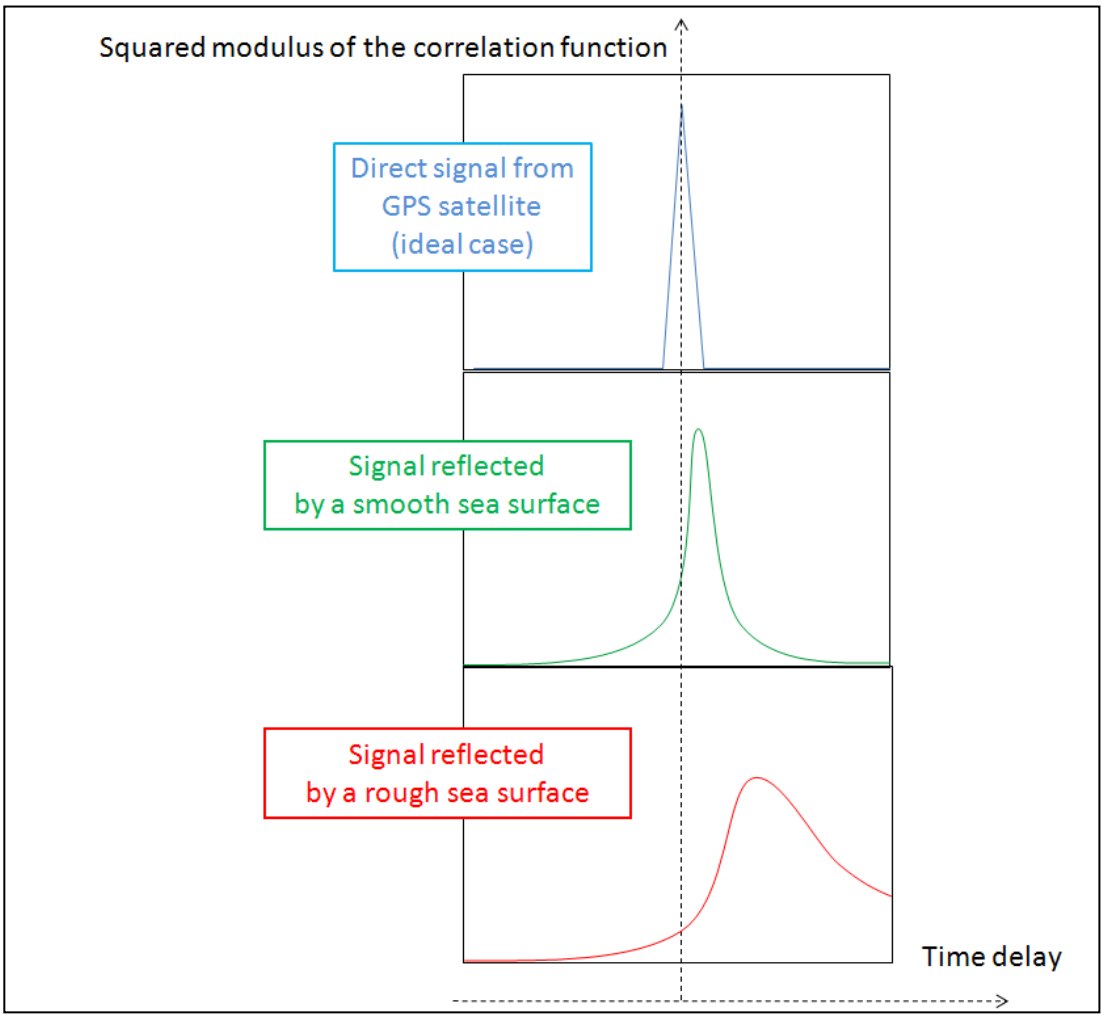


Figure 2: Distortion of the waveform depending on the signal considered

Because of the motion of the transmitter and receiver, the Doppler effect also comes into play. However, incoming reflected signals that arrive with frequencies $f_c + \delta f$ such that $\delta f > 1/T_i$ are filtered out by the processor. By computing the correlation function for several Doppler frequency shifts δf, the GNSS-R processor compensates for the Doppler effect and can better "track" the signal. The Delay Doppler Map (see Figure 4) is produced from the assembly of the waveforms $\{Y(\delta t, \delta f)\}_{(\delta t \in \mathrm{T}, \delta f \in \mathrm{F})}$, computed at different time delays $\delta t \in \mathrm{T}$ and Doppler frequency shifts $\delta f \in F$, each carrying the contribution of one particular cell on the sea surface, defined by the intersection of one iso-range annulus and two iso-Doppler stripes (see Figure 3).

$$Y(\delta t, \delta f) = \int_0^{T_i} s(t') \cdot prn(t_0 + \delta t + t') \cdot e^{-2j\pi(f_c + \delta f)t'} dt'$$

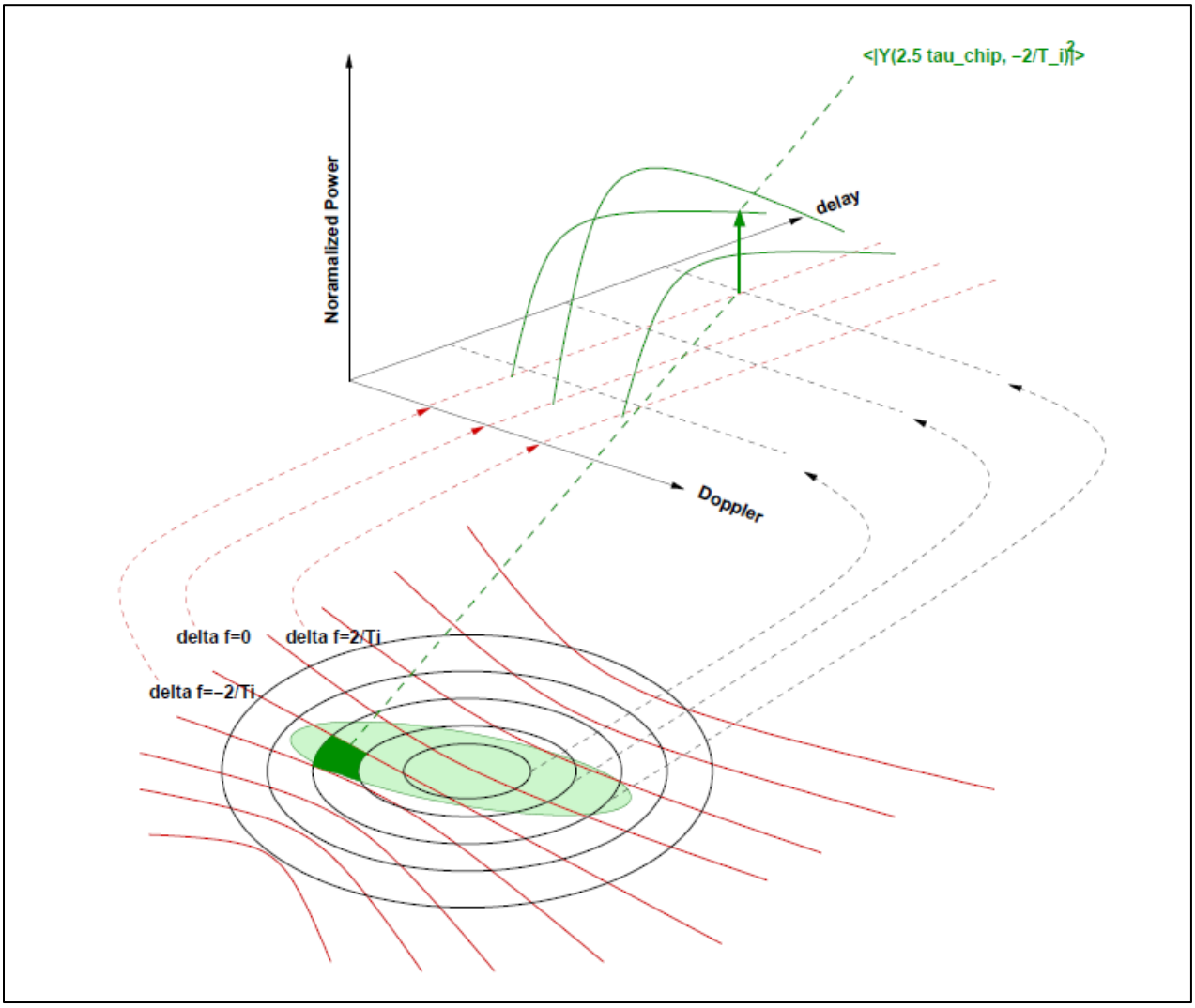


Figure 3: Gridding of the glistening area in both delay and Doppler shift domains ([9])

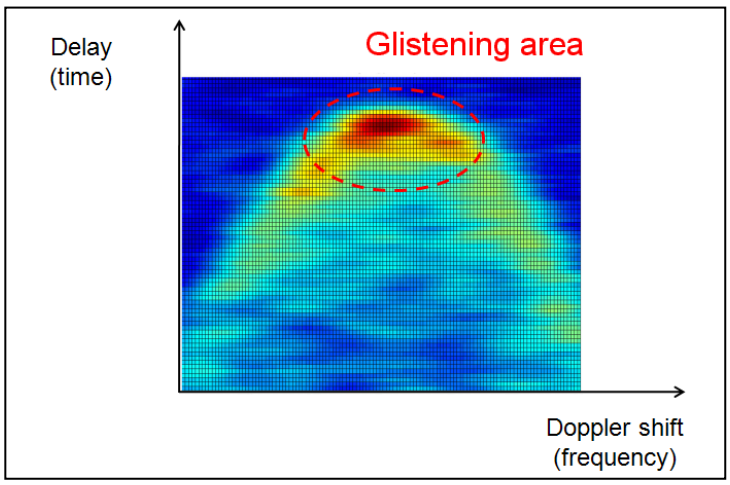


Figure 4: Example of DDM ([12]), arbitrary color scale

3.2.Radiometry

3.2.1. Principle

Remote sensing techniques can be either active or passive. Active remote sensing consists of illuminating a medium with an artificial source of radiation and measuring the amount of radiation that is reflected back (or scattered back). Passive remote sensing, on the other hand, consists of measuring the natural radiation that is emitted by the medium itself, or that is emitted by another source of radiation (the Sun for instance) and reflected by the medium, or both. It relies on the fact that every object above absolute zero does emit thermal radiation. Radiometry is a passive technique that belongs to the first category.

3.2.1.1. *Physics involved*

A blackbody is a theoretical medium which absorbs the entirety of any incident electromagnetic radiation, and in turn emits maximum possible radiation in all parts of the electromagnetic sprectrum. Given a blackbody at physical temperature $T > 0\mathrm{K}$, Planck's law describes the radiation that it emits per unit surface, in the normal direction, per unit solid angle and unit frequency f. This so-called spectral radiance writes:

$$B_f = \frac{2 \cdot h \cdot f^3}{c^2} \cdot \frac{1}{e^{hf/kT} - 1} \quad [\mathrm{W.m^{-2}.sr^{-1}.Hz^{-1}}]$$

- c is the speed of light, $c = 3.10^8 \mathrm{m.s^{-1}}$
- h is the Planck constant, $h = 6.626068.10^{-34}\ \mathrm{m^2.kg.s^{-1}}$
- k is the Boltzmann constant, $k = 1.3806503.10^{-23} \mathrm{m^2.kg.s^{-1}.K^{-1}}$

For L-band radiometric measurements of the Earth's sea surface, $hf/kT \sim 2.10^{-4} \ll 1$, so the Planck's law simplifies to the Rayleigh-Jeans approximation. The spectral radiance simplifies to linear function of the physical temperature T.

$$B_f = \frac{2 \cdot h \cdot f^3}{c^2} \cdot \frac{1}{hf/kT} = \frac{2 \cdot k \cdot f^2}{c^2} \cdot T$$

The emissivity of a medium is defined as the ratio of the radiation emitted by this medium and the radiation emitted by the medium if it were a blackbody. If I_{f_0} is the intensity coming from a medium at temperature T and observed at the frequency f_0, then:

$$e = \frac{I_{f_0}}{B_{f_0}(T)}$$

In radiometry, the concept of brightness temperature is used rather than that of intensity. If the medium is in thermal equilibrium (conditions of Local Thermodynamic Equilibrium, LTE), the brightness temperature T_b is the physical temperature at which the same medium would have to be if it were a blackbody, in order to emit the same radiation. Put in other words, we have:

$$T_b = e \cdot T \quad [\mathrm{K}]$$

Moreover, if the surface is flat, the analytical expression of the emissivity *e* is known (see 3.2.4), as opposed to the case of a rough surface.

The Rayleigh criterion is the most common way to determine whether a surface is flat or rough. Given the wavelength λ_0, the angle of incidence θ, and the significant wave height SWH of the sea surface, the surface is flat if it fulfills:

$$SWH < \frac{\lambda_0}{\alpha \cdot \cos\theta}$$

where α is a factor between 8 and 25.

In the present case, $\lambda_0 = \frac{c}{f_0} = \frac{3.10^8}{1.4.10^9} = 21.43\mathrm{cm}$ (L1 GPS signals have a frequency of 1.57542GHz, but radiometers used in the thesis operate at the frequency of 1.4GHz) and $\theta = 0$ (normal incidence). The most loose criterion ($\alpha = 8$) then leads to $SWH < 2.68\mathrm{cm}$, which is rarely fulfilled at open sea.

3.2.1.2. *Retrieval of geophysical parameters*

The end-goal of remote sensing is usually to retrieve geophysical parameters from the measured brightness temperatures. This retrieval is done through inversion of the forward model. The idea is to build a theoretical model – the forward model – which computes the brightness temperature, being given as input a certain amount of geophysical parameters, and taking into account all side-effects (instrument's sensitivity for instance). Then, the model is inverted, going from the measured brightness temperature to the unknown geophysical parameter. The most common inversion algorithms combine measured and *a priori* information, making assumptions about the various geophysical parameters (for instance, the estimation of SSS is made according to the typical salinity level of the area considered), and use statistics to determine the most probable solution.

To build such a model, one shall rely on both physical principles (radiation theory, instrument's noise, Cosmic Microwave Background radiation (CMB), etc.) and empirical relations (established from previous field experiments for instance) because the more accurate the forward model is, the more realistic the retrieved parameter is.

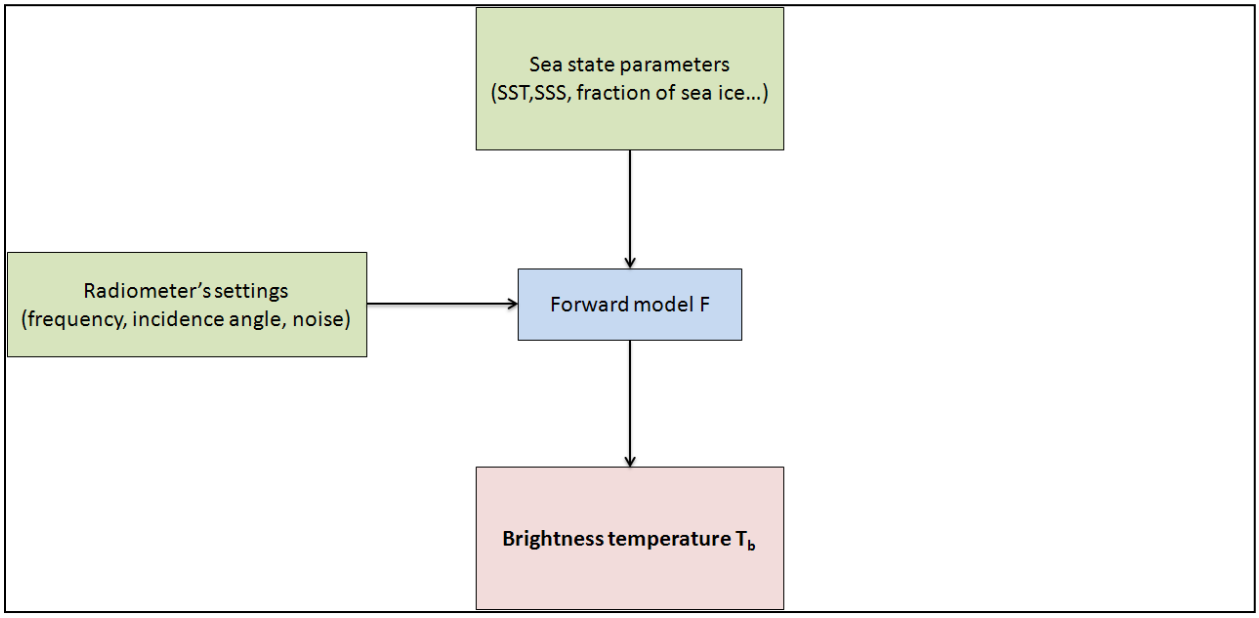


Figure 5: The forward model: from the sea state parameters to the sea emission

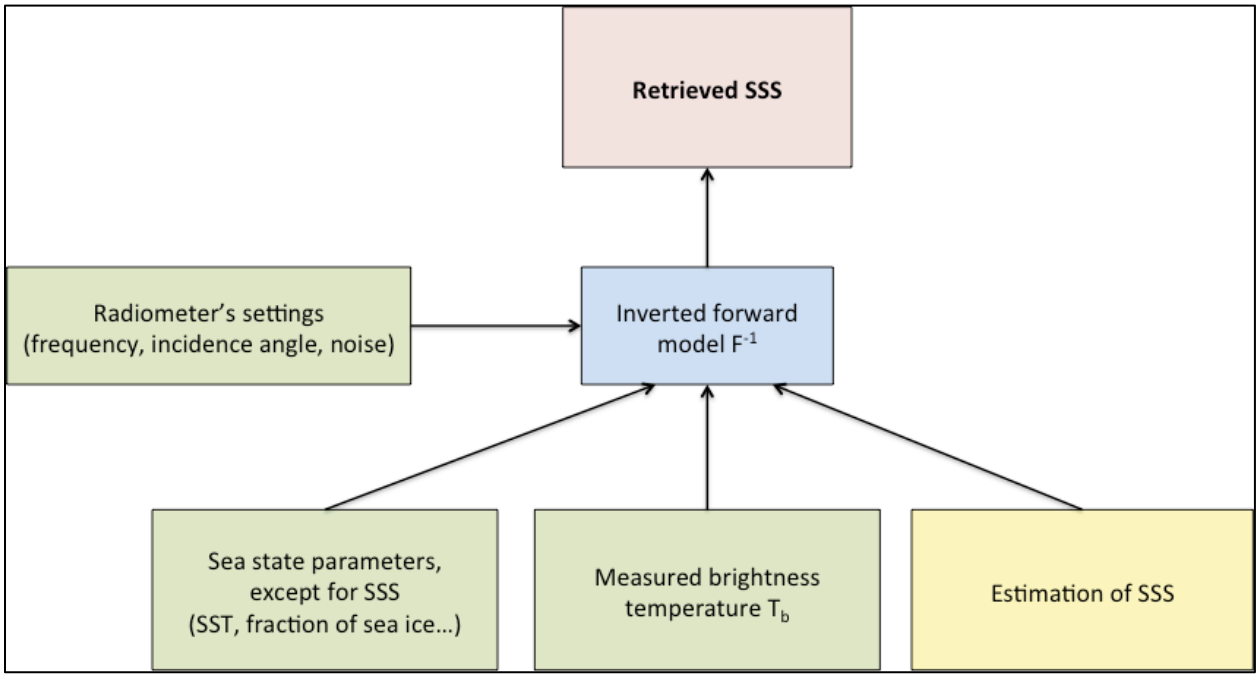


Figure 6: Inversion of the forward model: from the measured sea emission to the unknown sea state parameter

Throughout this thesis, only nadir-pointing radiometry is considered, meaning that the radiometer points at the medium with a normal incidence. Theoretically, the brightness temperatures for vertical and horizontal polarizations are then strictly equal, and investigation into only one of them is sufficient. However, airborne experiments cannot guarantee a perfect nadir pointing, due to small changes in the aircraft's attitude. If retrieval of geophysical parameters is done by inverting only one component of the brightness temperature,

there will be irrevocably errors in the solution. Therefore, in the present thesis, the first Stokes parameter is considered instead (see Figure 7, which presents the Klein & Swift predictions for the emission of a flat sea surface at $SST = 20°\mathrm{K}$ and $SSS = 4\mathrm{psu}$). Being the sum of the two opposite polarized brightness temperatures, this parameter somehow cancels out the possible discrepancies in the measurements.

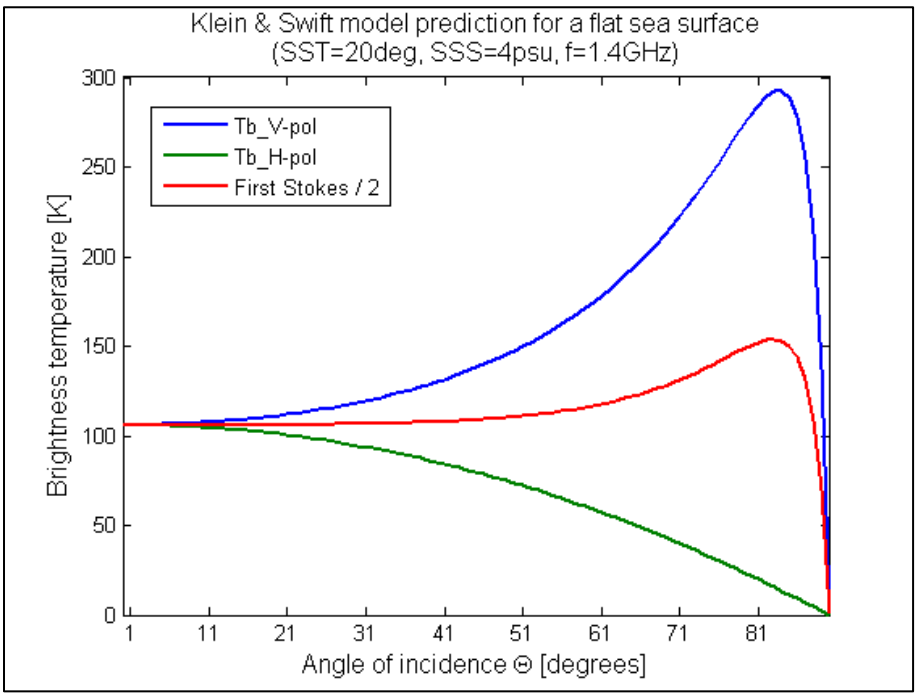


Figure 7: Klein & Swift prediction for the emission of sea surface ($SST = 20°, SSS = 4psu, f = 1.4GHz$)

3.2.2. Choice of L-band

Although gradients of salinity in oceans are very low, they play a major role for the circulation pattern of water worldwide, and thus have major effects on the whole environment system. That is why it is so important to measure SSS with the highest sensitivity possible.

Whereas the choice of passive remote sensing to measure oceans' salinity distribution is expected, the choice of using L-band may not be. There are actually three reasons for this. First, radiometric measurements are the most sensitive to salinity at frequencies around 1Ghz. Second, the protected frequency band [1.4; 1.427] GHz is dedicated to radio-astronomy utilization. Last, the radiation of the sea surface does not undergo any absorption from the cloud layer at L-band (more exactly, this absorption is neglected); likewise, thermal emission of the cloud layer can be neglected. Therefore, by making L-band radiometric measurements of the sea surface, one makes sure that the measured brightness temperatures really mirror the salinity distribution and are also clean from man-made emission signals (also called RFI, Radio Frequency Interference). Using low frequencies results in poor resolutions of the radiometers, but salinity gradients are quite large-scale in open sea, so a low resolution is not problematic (barely +/-2psu over thousands of kilometers).

3.2.3. Motivation for a better model of apparent sea emissivity

L-band radiometry has proved to be very promising for the remote sensing of sea salinity. However, two issues of prior importance make the proper retrieval of this parameter difficult.

First, although sensitivity of radiometric measurements to salinity is best at L-band frequencies, it is still quite low. Therefore, it demands that the forward model used to retrieve SSS is thoroughly accurate. Figure 8 presents the sensitivity of sea emission at 1.4GHz to SSS gradients at different SST values (conditions of the Gulf of Finland are $SST = 15°\text{C}$ and $SSS = 5\text{psu}$). Salinity in the open ocean is typically between 32psu and 38psu, but salinity in the Baltic Sea is really low, and sensitivity of the first Stokes is accordingly very little.

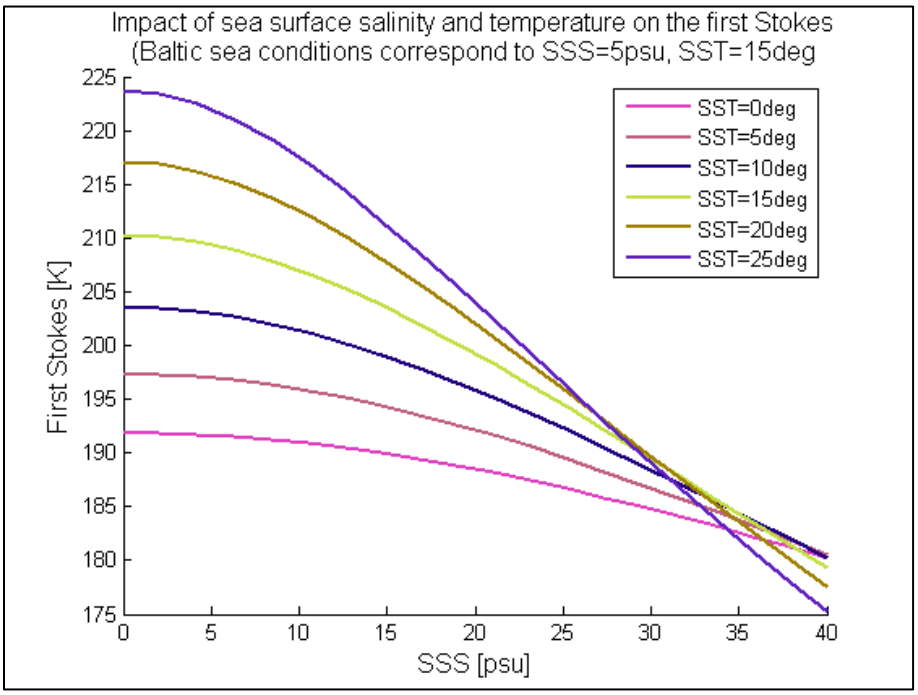


Figure 8: Impact of salinity and temperature on the sea emission

Secondly, it is unclear how, and to which extent, the roughness affects radiometric measurements. For flat water conditions, there exists a reliable forward model, using Klein & Swift model for the dielectric constant of water (see section 3.2.4). However, when sea surface roughness comes into play – as it is case at those wavelengths – there is some extra contribution from the roughness. Although the Klein & Swift model is still valid for the permittivity of water at such conditions, corrections of the forward model have to be made, and none of the models suggested so far is truly consolidated.

Once the forward model for flat surface emission is implemented, the difference between its output value and what is measured by the radiometer can be isolated, and identified as the contribution of the roughness to the emission. Of course, it is clear that in this thesis no salinity values are retrieved (hence no inversion algorithm is needed), since we solely study the effect of the roughness on the sea emission. The idea is instead to isolate the fraction of the first Stokes parameter that is the result of a non-flat sea surface, and

possibly correlate it with GNSS-R observables. This also means that all the inputs of the forward model must be known and given, notably the sea state parameters SST and SSS.

3.2.4. Klein & Swift model for the dielectric constant of water and sea emissivity model

In 1977, Lawrence Klein and Calvin Swift established a solid model for the dielectric constant of water ε_{water} at L-band, by means of a straightforward expression which depends on SST, SSS and frequency ([6]). The emissivity of water, whether flat or rough, is in turn related to the incidence angle and this dielectric constant, in a different manner depending on the polarization considered. For a normal incidence, it can be written as (the subscripts v and h refer to vertical and horizontal polarizations):

$$e_v = 1 - \left(\frac{\sqrt{\varepsilon_{water}} - \varepsilon_{water}}{\sqrt{\varepsilon_{water}} + \varepsilon_{water}} \right)^2$$

$$e_h = 1 - \left(\frac{1 - \sqrt{\varepsilon_{water}}}{1 + \sqrt{\varepsilon_{water}}} \right)^2$$

When Local Thermodynamic Equilibrium is reached – and it is assumed so here – the brightness temperature of flat sea water at temperature SST is then:

$$T_{b,v} = \left[1 - \left(\frac{\sqrt{\varepsilon_{water}} - \varepsilon_{water}}{\sqrt{\varepsilon_{water}} + \varepsilon_{water}} \right)^2 \right] \cdot SST$$

$$T_{b,h} = \left[1 - \left(\frac{1 - \sqrt{\varepsilon_{water}}}{1 + \sqrt{\varepsilon_{water}}} \right)^2 \right] \cdot SST$$

Since the cloud layer does not have any significant impact on the radiation of the sea surface (neither absorption nor emission), the brightness temperature measured by the instrument is precisely what the equations above predict.

As mentioned previously, the end-product of the forward model is actually the first Stokes parameter because it smoothes out the possible variations of the angle of incidence. Among the radiation that is measured by the radiometer, the contribution of the flat sea surface is then:

$$I_{flat} = T_{b,v} + T_{b,h} = \left[2 - \left(\frac{\sqrt{\varepsilon_{water}} - \varepsilon_{water}}{\sqrt{\varepsilon_{water}} + \varepsilon_{water}} \right)^2 - \left(\frac{1 - \sqrt{\varepsilon_{water}}}{1 + \sqrt{\varepsilon_{water}}} \right)^2 \right] \cdot SST$$

3.2.5. Apparent emissivity of rough sea surface

When the surface cannot be considered flat, the brightness temperature measured by the radiometer is higher than what the forward model yields. The radiation measured by the instrument, in terms of first Stokes parameter, can be split into two components:

$$I = I_{emission\ from\ sea} + I_{sky}$$

What happens is that the "sky radiation" (CMB radiation, galactic radiation, sun glint) is reflected, or more exactly scattered, by the multiple facets of the sea surface, and a part of it is picked up by the radiometer. It is not a component of the first Stokes, which is related to the emissive nature of the sea, but rather to the geometry of its rough surface. For given roughness conditions (including the case of absence of roughness, that is: flat sea), this quantity is constant over the time of the experiment since the illumination is also constant. For rougher and rougher conditions, this fraction is expected to increase; however, those variations are minor. [1] gives the modeled contribution of the reflected sky radiation, using two different models (WISE and L2), for Campaign A. One can see from Figure 9 that I_{sky} variations are very little over the test site (around 0.2K from the estuary (low longitude values) to the open sea (high longitude values)) and from one day to the other (almost null at the estuary, around 0.1K at the open sea). Models to quantify such effects are beyond the scope of the thesis, and – in this thesis – the fraction I_{sky} is considered constant during the time of the experiments.

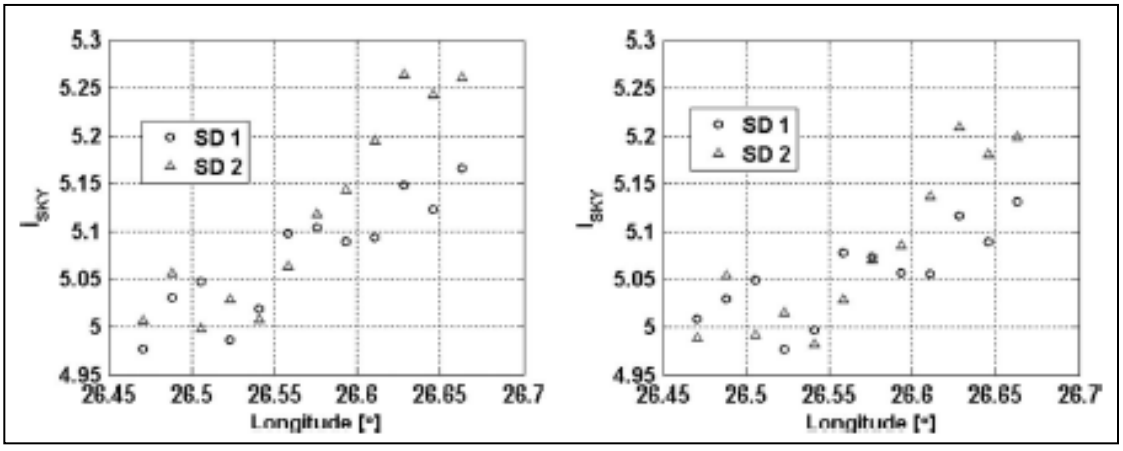


Figure 9: Modeled contribution of I_{sky}, using WISE (left) and L2 (right) models ([1])

The radiation that is due to the emissive nature of the sea can now be split into two components:

$$I_{emission\ from\ sea} = I_{flat} + I_{rough}$$

The component I_{flat} is described and modeled by the exact same means as those previously presented (see section 3.2.4). The Klein & Swift model for the dielectric constant is still valid, no matter how rough the surface is. As a matter of fact, the model that gives I_{flat} shall not be modified. Instead, the extra contribution $I_{rough} + I_{sky}$ shall be removed from the measurements prior to the retrieval of any parameter. By doing so, inversion of the flat-sea model will yield proper solutions.

As for the model of I_{rough}, it is the very purpose of this thesis. It is not totally right to claim that this contribution is due to the emissive nature of the sea; it is actually due to the local scattering of this emission. This radiation is reflected by multiple facets, all positioned at different angles from the others – although locally – leading to its scattering, and therefore to an extra radiation picked up by the radiometer. Although the emissivity of water does not change, the apparent emissivity does, and one could identify the effect of the roughness as an additional term $\Delta T_{b,rough}$ in the expression of the brightness temperature:

$$T_b = e \cdot SST + \Delta T_{b,rough}$$

It is very important to note at this point that the component that is identified in the thesis, and investigated for possible correlation, is not I_{rough} itself but the sum $I_{rough} + I_{sky}$ (as explained previously, I_{sky} is considered constant for each experiment). As a result, I_{rough} may be overestimated (there will be an offset in its values), but its variations with the surface conditions (hence the GNSS-R observables) are preserved.

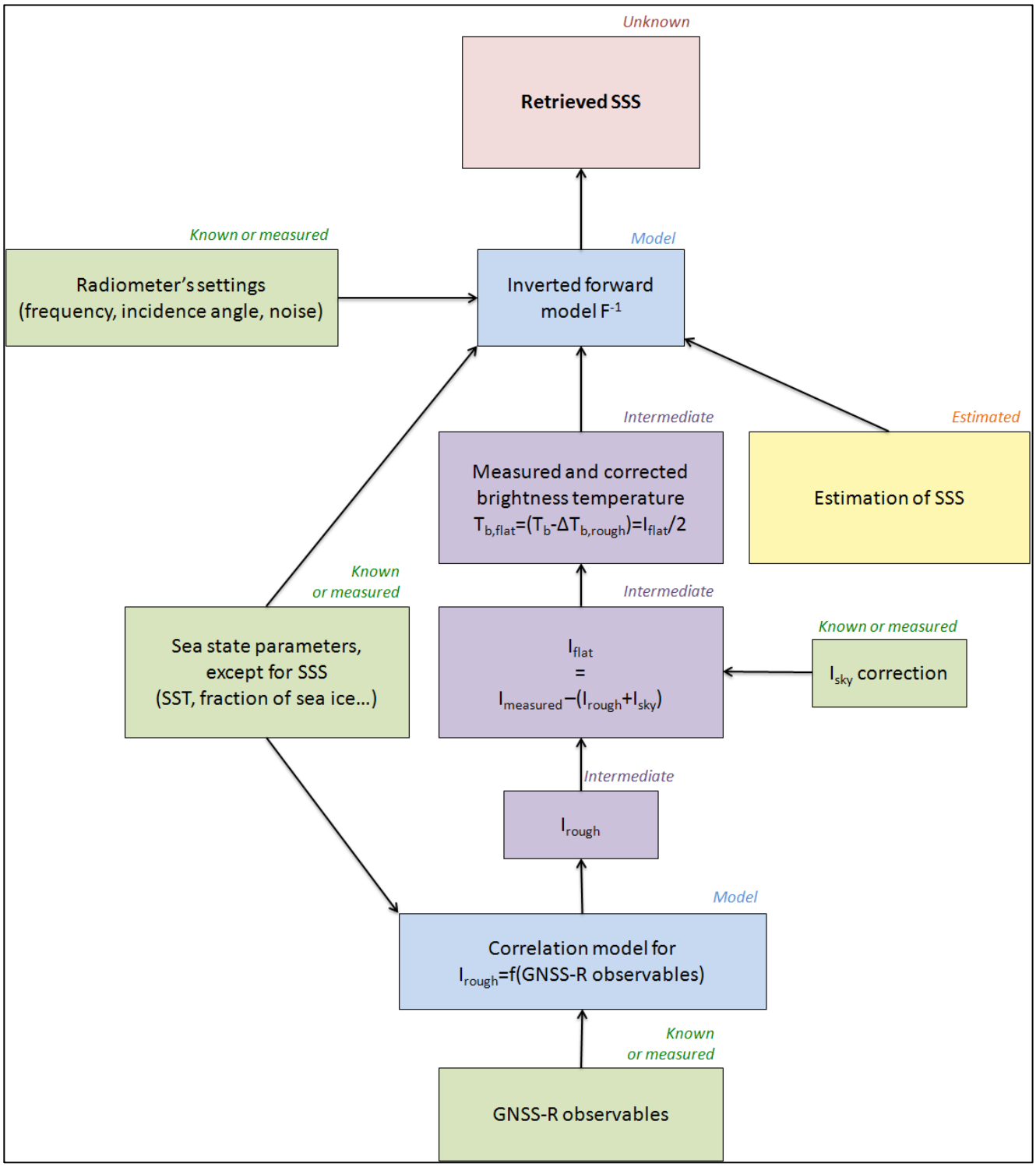


Figure 10: Proper retrieval of *SSS*

3.2.6. Attempts to model the influence of roughness

Approaches about how to model the contribution of the roughness are numerous. For instance, SMOS Level 2 Salinity Processor alone uses three of them. Most of them rely on expressions of the component I_{rough} in terms of SWH, MSS (Mean Squared Slope, see definition below), or U_{10} (wind speed 10m above sea level). It is not clear which of those expressions is the most accurate though, for their validity depends on the overall roughness condition of the experiment, and no model has been proclaimed better than the others.

One could argue that looking for a correlation between I_{rough} and a GNSS-R observable only will not help correcting the forward model, for this correlation might also depend on SST, SSS, or some other parameter. However, it is most likely not the case.

Some scientific articles, most of them published recently, did announce possible correlation of I_{rough} with some GNSS-observables, but they rarely provided any consolidated model or any analytical expression. Following is a brief review of the GNSS-R observables that were considered of possible interest according to the recent scientific publications.

Scatterometric delay

The scatterometric delay gives the relative delay of the waveform peak with respect to the specular point. The waveform peak is more and more delayed with increasing roughness conditions, whereas it coincides with the contribution coming from the specular point in flat sea conditions. Therefore, the higher this delay is, the rougher the surface is.

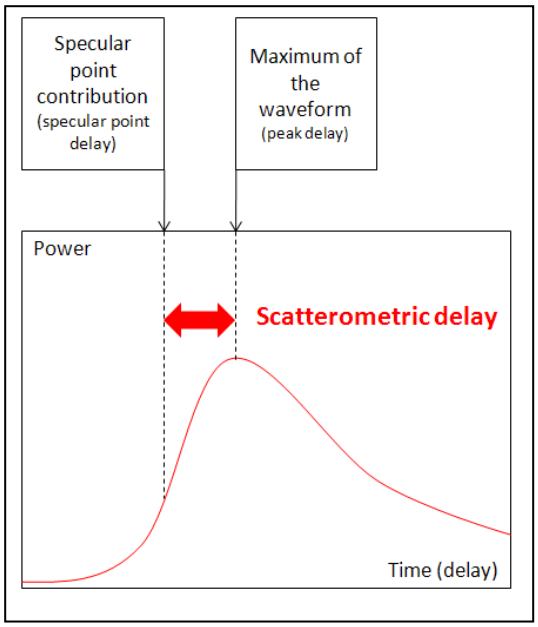


Figure 11: Description of the scatterometric delay

Waveform's decay time

The decay time of the waveform is the time that the waveform takes to decay from its maximum value to the maximum value divided by e . Motivation to study this parameter comes from the fact that the waveform "spreads" more and more with increasing roughness conditions. Therefore, this time length is a good indicator of the roughness.

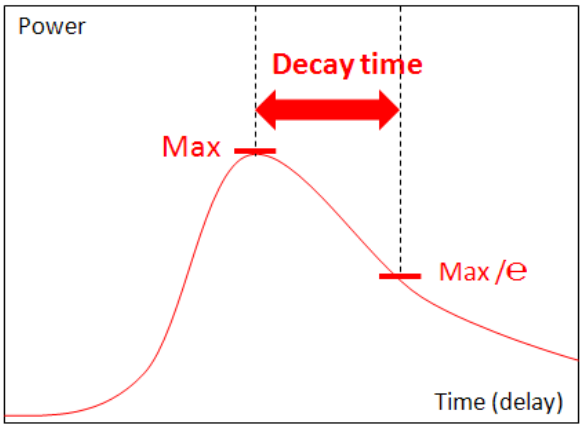


Figure 12: Description of the decay time

Surface Mean Square Slope estimates

The surface Mean Square Slope (MSS) is a variable that combine data concerning the delay of the waveform peak with statistical information about the sea surface roughness. It is computed from the wave height spectrum of the sea, and relies on a couple assumptions about the statistical distributions of waves and slopes (see [14] for more details). The higher this estimate is, the rougher the surface is.

DDM volume

The DDM volume is the result of the triple integration of the DDM, after setting a certain threshold in the power domain in order to minimize noise (the power is sometimes normalized to its maximum value). The value set for this threshold is a trade-off between the GNSS reflectometer's noise and its sensitivity to sea state variations. This variable is expected to increase with the roughness since the size of the glistening area increases with the roughness.

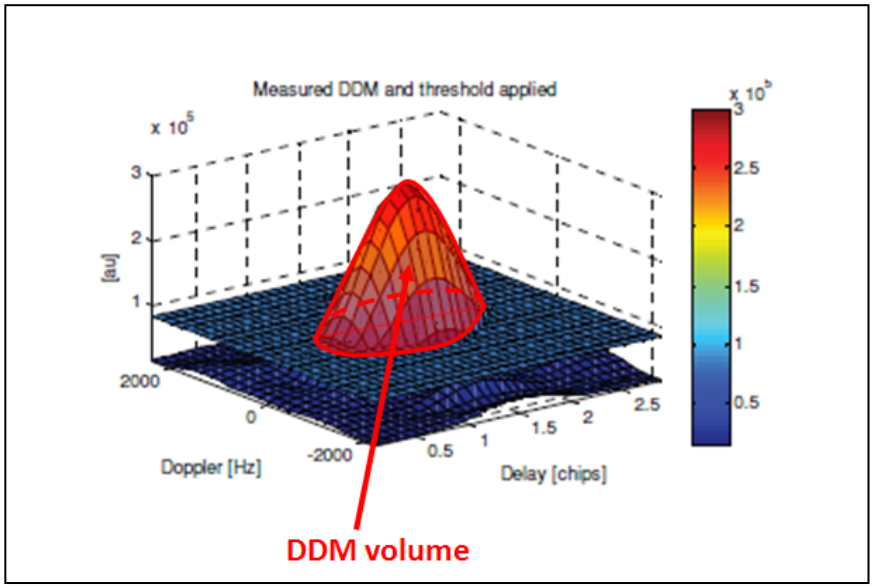


Figure 13: Description of the DDM volume ([13])

DDM area

The DDM area is the area of the section of the DDM with a certain threshold. Similarly to DDM volume case, the value of the threshold has to be carefully chosen.

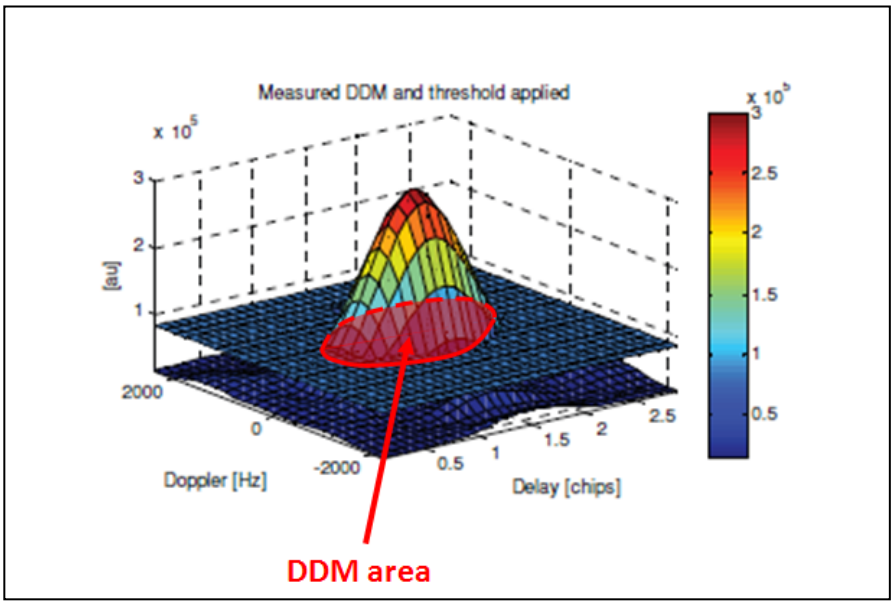


Figure 14: Description of the DDM area ([13])

4. Current status of research at RAD

Research about how GNSS-R techniques could be used to enhance the remote sensing of salinity has been going on for a while at RAD Department (Department of RADio and space engineering, at Aalto University, School of Electrical Engineering). The whole collaboration was first initiated in 2000, by the European Space Agency (ESA), in the framework of the SMOS satellite project (Soil Moisture and Ocean Salinity). The mission of SMOS is to provide the scientific community with regular and global maps of salinity in the surface waters of oceans (and similar maps of soil moisture in land areas), by performing L-band radiometric measurements. SMOS was launched in 2009, but next-generation satellites are on their way already, and scientific community commonly agrees that retrieval of parameters could be enhanced with the help of GNSS-R techniques.

Since 2004, several institutes have worked together with Aalto University on that topic (Aalto was called differently before 2010). Among them are: the Finnish Meteorological Institute (FMI), ESA, the Catalonian Institute for Space Studies (IEEC, for Institut d'Estudis Espacials de Catalunya), and the Technical University of Denmark (DTU). Since 2007, and before this thesis started, two flight experiments had been conducted with Aalto's research aircraft Short SC-7 Skyvan (see Figure 15). The overall idea is to operate an L-band airborne radiometer whose principle is analogous to the one used on board of the SMOS satellite, while simultaneously operating a GPS reflectometer.

- Campaign A was organized over the Gulf of Finland, in August 2007
- Campaign B was organized over the Baltic Sea, from Helsinki to Lübeck (Germany), in March 2008

For both campaigns, L-band radiometric measurements were performed with 2 independent instruments: HUT-2D, which belongs to Aalto University, and EMIRAD, which was borrowed from TUD. However, HUT-2D data cannot be used for Campaign B as the rainy conditions made the instrument fail during the campaign. More details about the campaigns are given in chapter 7.

Prior to the thesis, only EMIRAD data of Campaign A were used to possibly establish a correlation with one of the GNSS-R observables.

Figure 15: Aalto's research aircraft Short SC-7 Skyvan

5. Limitations

5.1.Limitation on the datasets to handle

Priority is given to data collected during Campaign A. The two main reasons for this is the reliability of the radiometric measurements for Campaign B, which were made at a quite high altitude (more land contamination of the signal and less gain), and the reliability of radiometric measurements for Campaign C, which were made with HUT-2D solely, a much more noisy instrument than EMIRAD.

5.2.Limitation on the GNSS-R observables to investigate

Priority is given to simple parameters when investigating the set of available GNSS-R observables. In particular, the key parameters of the so-called delay maps are studied prior to any other parameter, which means that it has been agreed at the beginning of the thesis that the key parameters of the so-called Doppler delay maps would be studied only if time allows it. More details about those parameters are found in section 6.2.3.

6. Instruments

6.1.Radiometers

6.1.1. HUT-2D

Aalto's radiometer HUT-2D is a two-dimensional interferometric radiometer usingaperture synthesis (dual polarization). It consists of 36 individual receivers arranged in a U-shape configuration, each branch of the U measuring two meters, the whole piece being positioned underneath the Skyvan aircraft (see Figure 16 and Figure 17). Its manufacturing and testing were completed in 2001, and many flight experiments have been organized since then, among which SMOS calibration and validation campaigns.

Figure 16: HUT-2D, unmounted

Figure 17: HUT-2D, mounted onboard (underneath) the Skyvan

HUT-2D generates maps of brightness temperatures, making snapshots of the ground every 250 milliseconds, with a ground resolution of roughly one ninth the flying altitude. Because HUT-2D is not a fully polarimetric radiometer, but works instead in a dual polarization mode, it does not measure the vertical and horizontal components of the brightness temperature, but instead the X and Y components in the antenna frame, by means of a X-probe and Y-probe. However, since the polarization rotation angle is known by the instrument, the First Stokes parameter can be computed easily. Theory indeed ensures that the radiation emitted by a point M at the Earth's surface is:

$$T_{b,X}(\varphi_X) + T_{b,Y}(\varphi_Y) \underbrace{=}_{If, and\ only\ if\ \varphi_X = \varphi_Y} T_{b,v} + T_{b,h} = I$$

The polarization for which the measurements are made is switched to its counterpart every second: the X-probe and Y-probe alternate. Among the four measurements made for a polarization each second, the two first ones are most likely corrupted because of transient effects, thus discarded. Eventually, there remain two measurements for each polarization every two seconds, that is: one First Stokes parameter per second.

It is worth pointing out that since the incidence angle actually ranges from 0 to 25 degrees (95% of the pixels fall into that bin), the pointing is not strictly at nadir. However, the range is narrow enough to consider that all measurement points are made at nadir, because the first Stokes parameter remains rather constant over this range of angles (see Figure 7).

Last, there is a radiometric bias of HUT-2D, which introduces offsets to the X-probe and Y-probe brightness temperatures. Possible reasons for this bias are the systematic errors in antenna directivities, and losses and noise in the receiver. However, those offsets can be considered constant for the same type of surface (sea in the present case), if not over years, at least during the time of the experiment. They are usually around +4K for the X-probe and -6K for the Y-probe. As for the radiometric resolution of HUT-2D, the finite integration time and thermal noise of the instrument cause the sensitivity to be around 6K.

6.1.2. EMIRAD

TUD's radiometer EMIRAD is a fully polarimetric real aperture radiometer, which means that it measures the brightness temperatures for both vertical and horizontal polarizations simultaneously. It has been developed especially for the remote sensing of sea salinity. Therefore, considering the weak impact of salinity gradients on the brightness temperature, it has been designed so that its sensitivity is very high (less than 0.1K with a one-second integration time). This instrument has been used on several airborne campaigns already.

EMIRAD is actually a two-receiver system: one antenna points at nadir, the other one points at an angle 40 degrees off nadir, both with a Half Power Beam Width (HPBW) of approximately 35 degrees (see Figure 18). However, only data from the nadir-pointing receiver are used for the thesis, leaving the rest of the data for future studies.

Figure 18: EMIRAD horn antennas mounted on the Skyvan

6.2.GNSS reflectometers

6.2.1. GOLD-RTR

The GPS Open Loop Differential Real-Time Receiver GOLD-RTR was designed and assembled by IEEC, and ready-to-use in 2005.

There are three antennas from which the signals are collected: one up-looking antenna positioned on top of the Skyvan (to collect the right-hand circularly polarized (RHCP) direct GPS signals), and two down-looking antennas (to collect the reflected GPS signals which, due to the reflection at the sea surface, are left-hand circularly polarized (LHCP)) positioned underneath the Skyvan. The GPS reflectometer consists of 640 complex correlators distributed in ten channels of 64 lags each. Channels can be programmed in various ways, either all using the same GPS satellite and antenna at the same time, or not. Also, they can either correlate signals at zero Doppler shift (DM maps), or use instead a frequency-shifted model signal (DDM maps). The dataset of the ten waveforms of 64 lags each, which are generated every millisecond, forms the raw data. However, GOLD-RTR also delivers integrated geo-referenced data, after incoherent summation of every set of 1000 waveforms computed each second, and those data are the ones investigated in this thesis.

For all campaign flights A, B, and C, the sampling rate f_r in the correlation channels was set to 20MHz. This corresponds to a spacing of $\frac{c}{f_r} = 15\text{m}$ between two subsequent waveform samples (time delays and distances are used indifferently, since time delays are just distances travelled at light speed), One channel being made of 64 lags, it covers the domain $(N_{lags} - 1) \cdot {}^{1}/_{f_r} = \frac{63}{20 \cdot 10^6} = 3.15\mu\text{s} \approx 3.2\tau_c$ in terms of chip length, or 945m in terms of distance.

Let's consider one of the ten-correlator channels of GOLD-RTR, aiming at computing a waveform at zero Doppler shift for the reflected signal received at the down-looking antenna number 2. Figure 19 illustrates how the lag number i (out of the 64 lags of the channel) computes the power of the waveform at delay $(i - 1) \cdot {}^{1}/_{f_r}$.

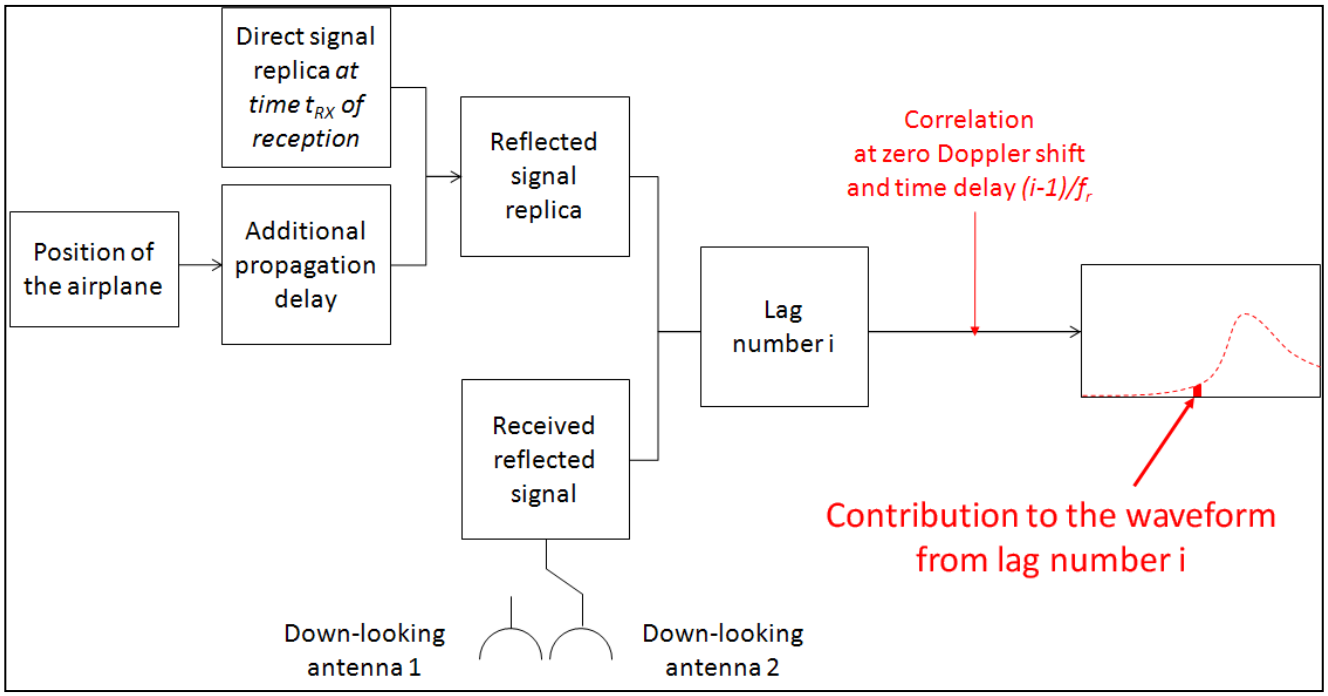


Figure 19: Diagram illustrating the work of one single correlator

The direct signal replica is computed from the theoretical GPS signal, to which a correction for the various delays that the actual received direct signal encountered is made. Those delays can be geometric (additional travelled path), atmospheric (due to the ionosphere in particular), systematic (multiple time scales for the tagging of the data), or simply noise (receiver-antenna downlink-cable delays).

Figure 19 helps to understand the general principle of each correlator. However, it is wrong to associate the values $(i-1)\cdot {}^{1}/_{f_r}$ to actual time delays with respect to the specular point. As a matter of fact, GOLD-RTR starts correlating the reflected signal with its replica some time *before* the estimated delay of the specular point. Moreover, the estimation of this delay does not match the actual value (see section 6.2.3 for more details).

The power values computed by each of the 64 lags form one waveform. However, what is used by GOLD-RTR is the coherent average of the waveforms computed within each millisecond. In addition, GOLD-RTR performs an incoherent averaging of those waveforms, with an integration time of one second (that is: 1000 coherently integrated waveforms), in order to provide users ready-to-use data.

If the correlation is to be performed at a non-zero Doppler shift frequency, the according shift is made to the time of computation, but the overall principle is the same.

6.2.2. PIRA

PIRA is an enhanced version of GOLD-RTR. Instead of using the direct signal replica to compute the reflected signal replica, the processor uses the received direct signal (after compensating for the additional path traveled by the reflected signal, and other noise and system delays). This method makes the correlation much more accurate. Yet, one drawback about the instrument is that it uses M-code GPS satellites only, those being quite scarce for the moment.

6.2.3. Observables

Many variables are included in the integrated geo-referenced data, but only a few of them are worth investigating for the purpose of this thesis. Also, only data concerning the reflected signal are of interest here (waveforms of the direct signal are valuable for altimetric purposes).

The variables all refer to a certain record index, which identifies each measurement made (one measurement being the incoherent average of coherently integrated waveforms, as explained in the previous section).

The variables that quantify delays are all given in distance unit. Most of them are defined with respect to a reference range: $NominalDelay$. For the reflected signal waveform, $NominalDelay^{reflected}$ means the expected delay of the specular point (or rather an *a priori* estimation of it) with respect to the direct signal. For the direct signal waveform, $NominalDelay^{direct}$ is actually the origin of time (zero value).

Delay Maps

For each record considered, GOLD-RTR provides the following variables (see Figure 20 for better understanding of their meaning):

- $NominalDelay$,
- $Waveform$, which is the power of the waveform, at each correlation lag of each channel,
- $MaxWav$, which is the peak power of the waveform,
- $MaxWavDelay$, which is the time - with respect to $NominalDelay^{reflected}$ - at which the waveform is maximum,
- $MaxDerDelay$, which is the time - with respect to $NominalDelay^{reflected}$ - at which the first derivative of the waveform is maximum. As it is explained in [18], the first derivative's maximum situates the actual specular reflection, so the knowledge of $MaxDerDelay$ permits to compute observables defined with respect to the *real* specular point (as opposed to its estimate $NominalDelay^{reflected}$).

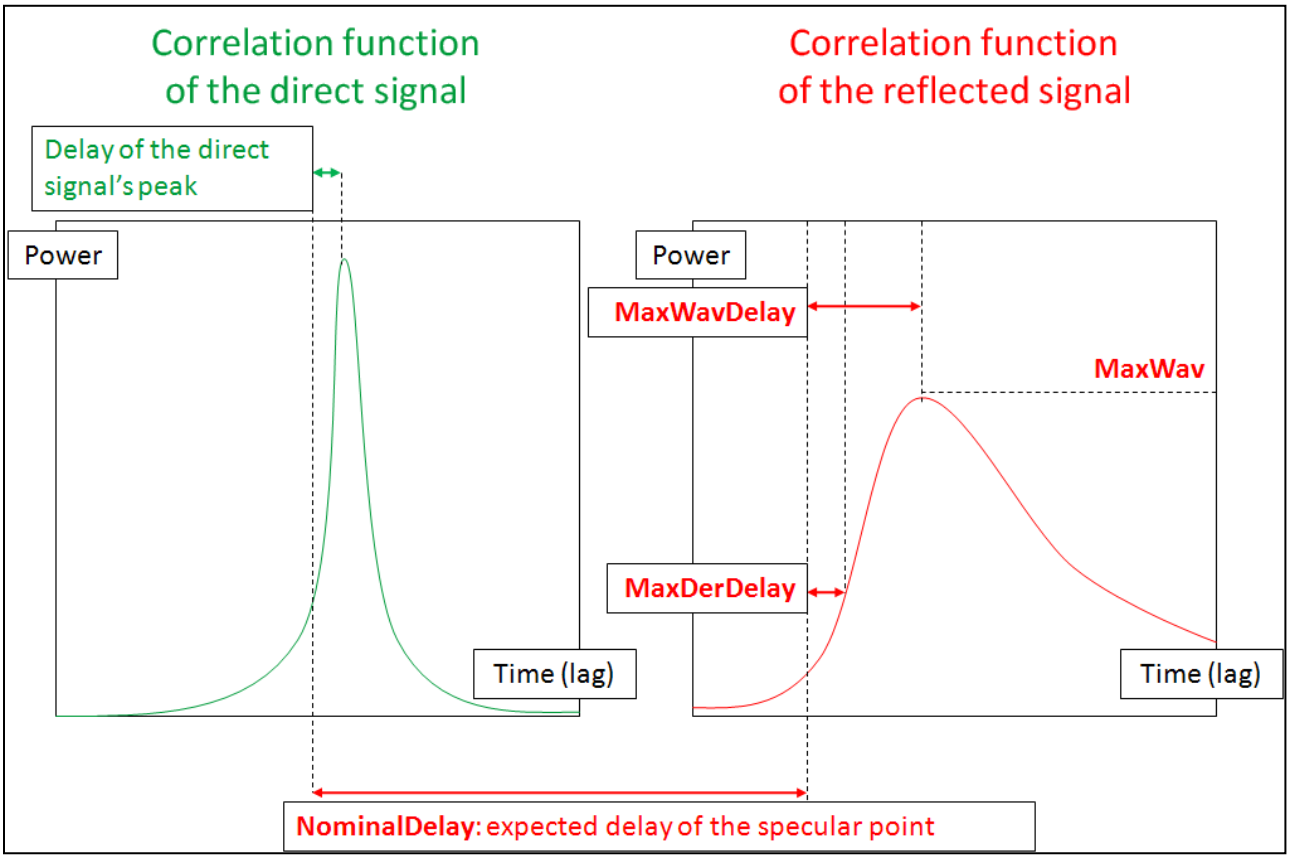


Figure 20: GOLD-RTR observables

From this set of variables, one can compute some other basic observables for the reflected signal's waveforms (see Figure 21 for better understanding of their meaning). Some of them are defined with respect

to the actual specular reflection, while the others are defined with respect to the time of reception of the direct signal. For the latter, data have to be identified for each PRN code separately, because they are inherently related to the elevation of the satellite in the sky; otherwise, no good use of the data is possible. Information contained in all those variables are meaningful, in the sense that they do not suffer from possible errors in the estimation of $NominalDelay^{reflected}$. However, IEEC indicates that the computation of the waveform might fail at low altitudes, and/or when separation of the different PRN codes cannot be made for some reason, and/or when reflected and direct signals overlap. Whereas those corrupted measurements appear like additional noise in variables that are not defined with respect to the direct signal (all PRN averaged together), they can completely distort those that shall consider one single PRN code.

Following are the variables that are studied in the thesis. Figure 21 and Figure 22 helps to better understand their meaning.

- The peak delay ($PeakDelay$), which is the delay of the peak of the waveform with respect to the direct signal.

  $$PeakDelay = NominalDelay^{reflected} + MaxWavDelay^{reflected} - MaxWavDelay^{direct}$$

  The term $-MaxWavDelay^{direct}$ can be neglected in the computation.

- The waveform's tail time delay ($TailDelay$), which is the time (though it is given in distance unit) that the waveform takes to decay from its maximum value to a factor exp(-1) of it. The reason why this parameter is a good indicator of the roughness is explained in section 3.2.6.

  $$TailDelay = \min_{t>t_{MaxWav}} \left(t; waveform(t) = \frac{MaxWavDelay^{reflected}}{e} \right)$$

- The specular delay ($SpecDelay$), which is the delay of the specular point with respect to the direct signal.

  $$SpecDelay = NominalDelay^{direct} + MaxDerDelay^{reflected} - MaxWavDelay^{direct}$$

- The scatterometric delay ($ScattDelay$), which is the delay of the peak with respect to the specular point. The reason why this parameter is an indicator of the roughness is explained in section 3.2.6.

  $$ScattDelay = MaxWavDelay^{reflected} - MaxDerDelay^{reflected}$$

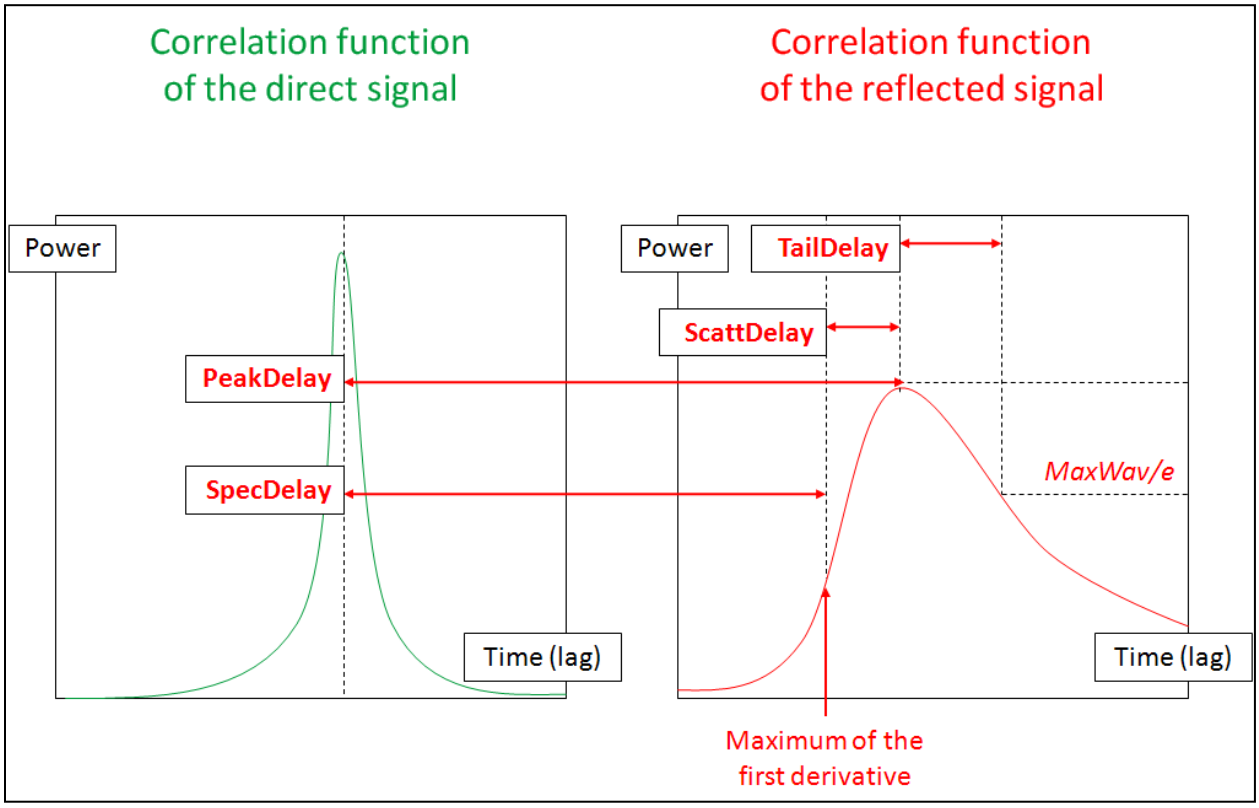


Figure 21: Second-level GOLD-RTR observables ($MaxWavDelay^{direct}$ is neglected)

- The waveform area ($WavArea$), which is the area under the normalized waveform once a certain threshold has been applied (to reduce the impact of noise). Motivation to explore this parameter is [15], especially since the latter article studies data of Campaign A (for the first flight experiment only). The choice of the threshold 0.2 is also made according to [15], in order to compare achieved results.

$$WavArea = \int_{\frac{Waveform}{Waveform_{MAX}} \geq 0.2} \frac{Waveform}{Waveform_{MAX}}$$

  Since the waveform is normalized, the unit of this area is the same as the one of the delay domain (meter in the present work).

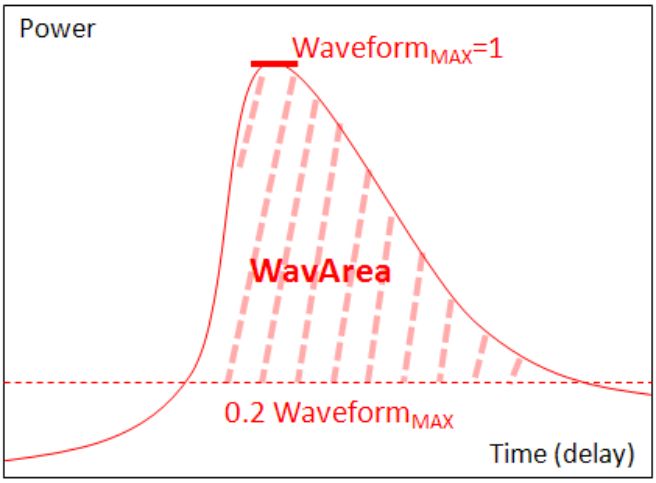


Figure 22: Illustration of $WavArea$

In addition, IEEC provides MSS estimates with an integration time of 10 seconds (see section 3.1.2 for the description of MSS). MSS estimates are given – and used – as-they-are, without further explanation about the algorithm used for the computation.

Doppler Delay Maps

Doppler Delay Maps can be generated from complex DM. However, the data made available by IEEC and used in the thesis already contain some DDM "slices", computed by IEEC itself. The latter contain little information about the glistening area because the Doppler spread that those DDM cover is quite narrow ([-100Hz; +100Hz], with steps of 50Hz). Since generating DDM from complex DM is much beyond the scope of the thesis, those DDM datasets are utilized as-they-are, and there are doubts that they contain valuable information.

6.3.Thermal infrared radiometer

For campaigns A and B, a narrow-beam Thermal Infrared Radiometer (TIR) was installed in the nose compartment of the Skyvan. It provides temperature readings with an integration time of 1 second, the data being synchronized with the GPS time. In the case of Campaign B, those recordings are used as back-up data for in-situ SST. Indeed, no ground-truth data are available for this experiment, and data extracted from the model simulation have poor resolution and low reliability; it is then better to use IR data for SST.

7. More about the campaigns

7.1.Experiment setup

The experiment setup could be represented as Figure 23. The interval δt between the time of GOLD-RTR measurement and time of radiometric measurement is negligible, so the pair {GPS-R observable – first Stokes} actually forms one single measurement, made at time t and over the point of coordinates $(lat, long)$.

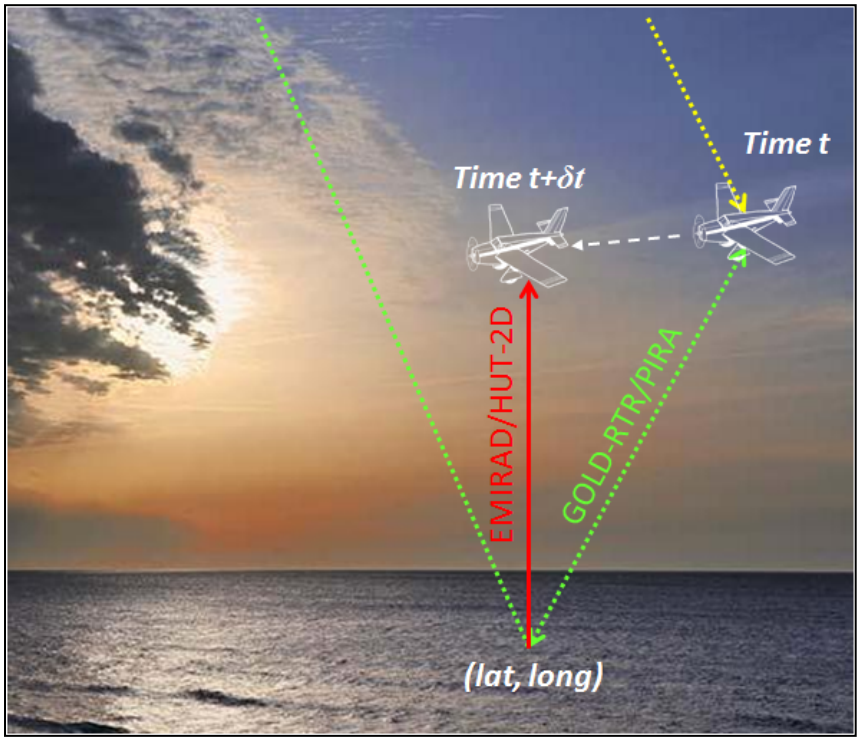


Figure 23: Experiment setup

Table 1 presents the different instrumentation setups of the Skyvan. It is a summary of the different instruments that were on board the Skyvan for campaigns A, B and C (more details about each campaign are given in section 7.3).

Table 1: Instrumentation of the Skyvan used for campaigns A, B and C

Campaign / Instrument	A Gulf of Finland 2007	B Baltic Sea 2008	C Gulf of Finland 2011
EMIRAD	X	X	
HUT-2D	X	X	X
GOLD-RTR	X	X	X
PIRA			X
IR radiometer	X	X	

7.2.Selection of the test site

The three campaigns that have taken place so far were not organized for the same unique purpose. For instance, the main motivation for Campaign A was to demonstrate that the utilization of GNSS-reflectometry together with L-band radiometry could capture a salinity gradient. Campaign B, on the other hand, aimed at rehearsing the validation phase of SMOS (SMOS was launched one year after this campaign). Last, Campaign C aimed at testing the measuring capabilities of the enhanced version of IEEC's GPS reflectometer, called PIRA.

However, all three campaigns have been organized so that they meet of the four common requirements described in the following sections.

7.2.1. Ground reference data

The campaigns were all test flights, so there was a need for in-situ data. In the case of demonstration or rehearsal especially, only ground reference data can discriminate a successful test from a failure. For instance, in order to know whether the instruments could detect a salinity gradient in Campaign A, one needs to know the real distribution of this gradient. As for the thesis itself, it is clear that no model can be inferred from the measurements if no data about the sea state are collected. The selection of the flight route is then made – to some extent – according to the availability of in-situ measurement stations in the vicinity of the flight area (either marine weather stations operated by FMI, or buoys deployed specifically on the occasion of the flight experiment). Still, because of the scarcity of these stations, data are also extracted from the computational

numerical weather forecast models in collaboration with FMI. These models are the operational hydrological model for the Baltic Sea (HIRLAM), and the BalEco model).

7.2.2. Different roughness conditions

The existence of different roughness conditions within the same area of flight is desirable. The change in the roughness should not be too important, but it should be noticeable enough for the performance of the model/retrieval/processor/etc. to be assessed. Especially for the campaign to be organized, the new GPS reflectometer is different from the one used in the previous campaigns, so the main objective is to test whether it works or not, and to check whether the modifications actually lead to better results or not (GOLD-RTR will also be taken on board), hence the demand to test the instrument in both flat and "normal open sea" conditions.

7.2.3. Trade-offs with what is actually feasible

Trade-offs have to be made between the ideal experiment setup and the feasible one. There exists zones in the sky where flying is either restricted or forbidden (for safety or military reasons), so it is better that the flight route does not cross any of those zones. Also, there is a minimum flying altitude authorized, depending on the area, which is not the same depending on whether the flight happens during daytime or nighttime. Another constraint is the capacity of the Skyvan's gas tank, which allows experiment flights of about three hours (depending on the weigh onboard). It is thus not time-efficient to plan a route to the Gulf of Bothnia while take-off is planned in Helsinki airport. Last but not least, even though weather conditions are not taken into account when the flight route is selected, those rule over every other factor: if flying is not recommended, the experiment is cancelled.

7.2.4. Visibility of GPS satellites

The visibility of GPS-satellites has to be taken into account. One first point is that the closer to zenith they are, the better it is. The second point is that the latest GPS reflectometer PIRA only uses GPS satellites that transmit M-code signals. It is not worth going into details here, but it is worth mentioning that those satellites belong to the latest generation of satellites, hence are quite scarce, and so is their visibility in the sky. All this leaves a quite narrow time window for the selected day of experiment (see Figure 24 as an example of M-code satellites' visibility (in red) for the 22nd of June 2011).

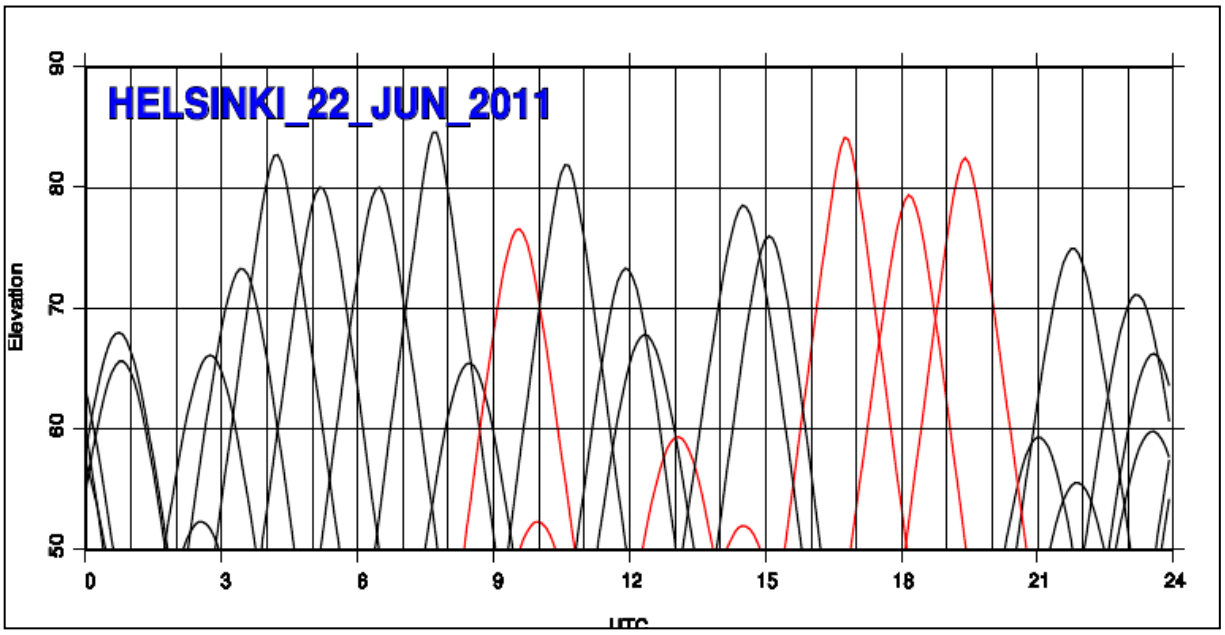


Figure 24: Visibility of GPS satellites on 22/06/2011 in Helsinki, among which M-code satellites in red

7.3.Overview of the test-sites

7.3.1. Campaign A

Test site A is in the Gulf of Finland, in the vicinity of Kotka and Loviisa, East of Helsinki. It can be divided into two distinct parts: a bay (which is the discharge area of Kymi river) and an open sea area (which is connected to the bay through a narrow passage). A gradient in the salinity level of water is therefore expected: the bay contains mostly fresh water, while the open sea area is much more saline. This campaign was organized to rehearse the validation phase of SMOS after the launch, hence the need for a salinity gradient to detect.

The flight route is about 15km long. The experiment took place on two days, the 13th and 15th of August 2007, both around 23h local time, and the test site was flown over 20 and 21 times respectively. In-situ data (SST and SSS) were collected by means of a boat. Two buoys were also deployed from this boat at the different sampling sites, in order to collect information about the roughness conditions (such as SWH, wave spectrum, etc.). Flying altitude for this experiment was 350m.

Figure 25 presents the whole flight route, while Figure 26 presents the part of interest SSS1 <-> SSS2.

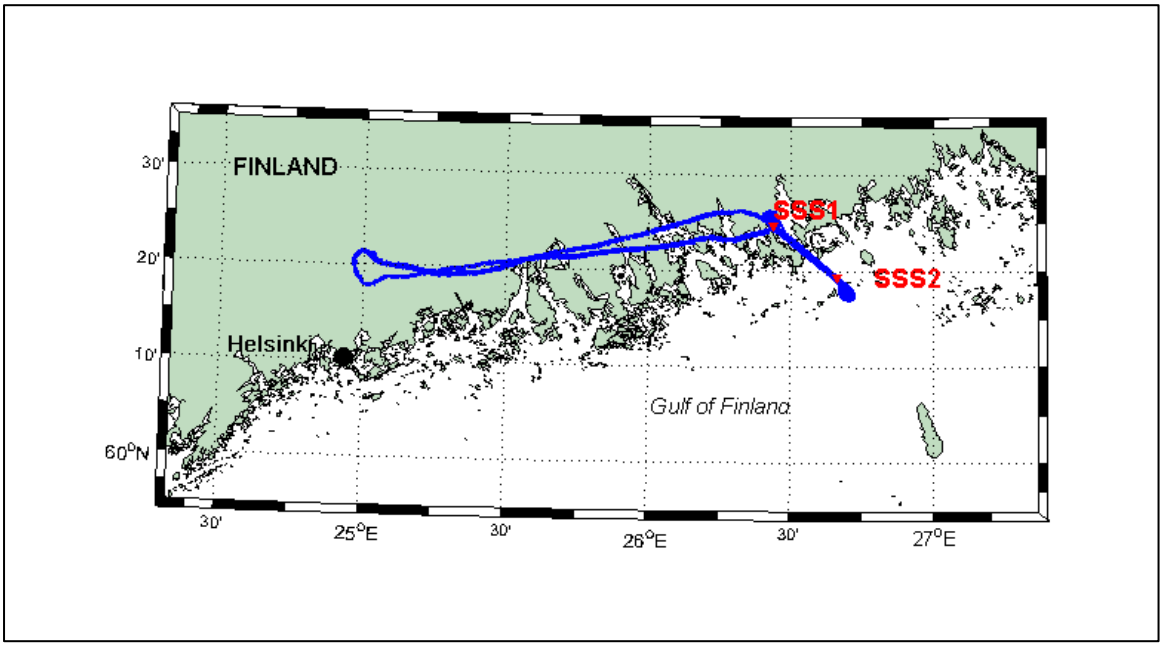


Figure 25: Flight route for Campaign A

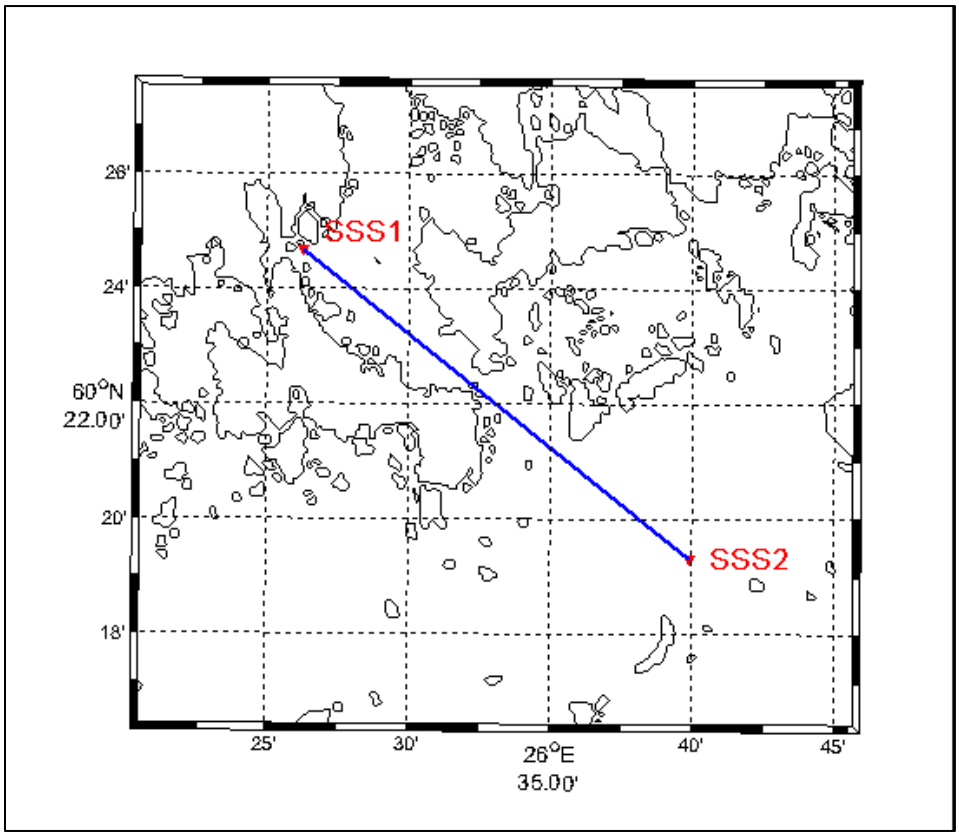


Figure 26: Area of interest for the measurements of Campaign A

7.3.2. Campaign B

Test site B is in the Baltic Sea, from Helsinki to Lübeck, in Germany, with a stop in Visby, on the Swedish island of Gotland (in order to re-fuel the Skyvan's gas tank). However only measurements made strictly over the sea are valuable, and those made too close to the shoreline are discarded (instruments were also turned on over land for the purpose of soil moisture measurements). The entire route is about 1000km (see Figure 27 with different landmarks), but the valuable part corresponds to only 400km. No in-situ information were collected for this campaign; instead, SST and SSS extracted from the computational forecast models, as well as recordings of SST by the TIR, were used.

This campaign was organized to rehearse the calibration of SMOS satellite. It took place on the 31st of March 2008, at a flying altitude of around 3000m.

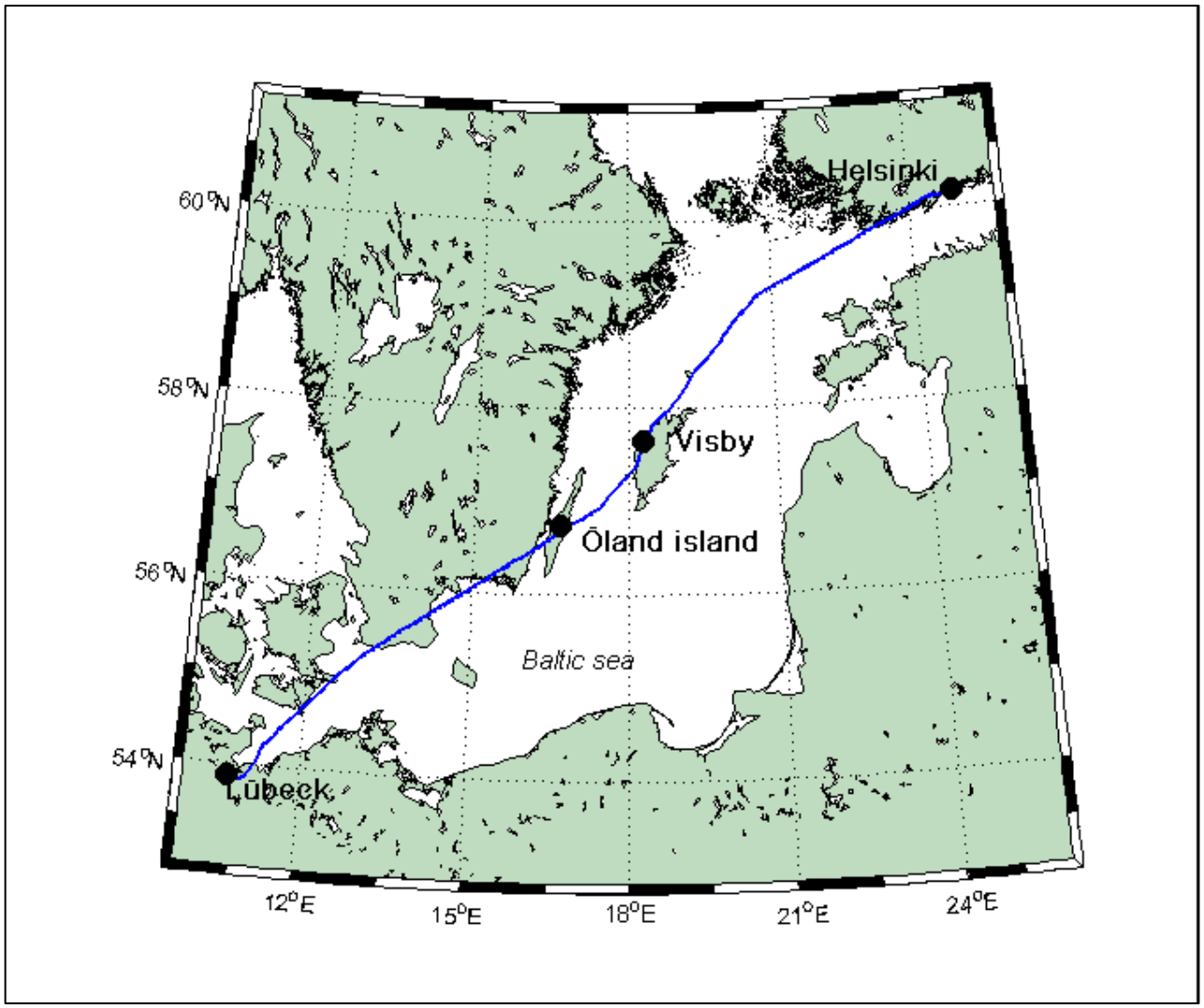


Figure 27: Flight route for Campaign B

7.3.3. Campaign C

Test site C is in the Gulf of Finland. There are two distinct parts in the flight route, both being straight trajectories. One of them consists of the same test site as in Campaign A (Kymi river estuary), expected to have a rather smooth water surface. In contrast, the second part of the route consists of an open sea area, parallel to the Finnish coast line and expected to have a rather rough water surface. This campaign was organized to test the airborne measuring capabilities of the new GPS reflectometer of IEEC – PIRA – hence the need for different roughness conditions. For comparison and validation, GOLD-RTR was also onboard. It took place on the 21st of June 2011, around 20h local time (this because of the visibility of M-code GPS satellites). As for in-situ measurements, no boat was deployed for this campaign. However, the test site has in its vicinity a few automatic weather stations operated by FMI. These stations measure wind speed and direction (one of them, Harmaja station - number 0330 - also measures SST). Their identification and location are displayed on Figure 28. What's more is the absence of the TIR in the nose of the Skyvan.

Figure 28: Identification and location of the automatic weather stations

Figure 29 presents the flight route, with its two parts of interest:

- the open sea part, SWP1 <-> SWP2, of approximate length 55km, which was flown over twice, at two different flying altitudes (500m and 150m),
- the estuary part, SSS1 <-> SSS2, of approximate length 15km, which was flown over six times in total , that is: three times back and forth from SSS2 to SSS1, at the flying altitude of 150m.

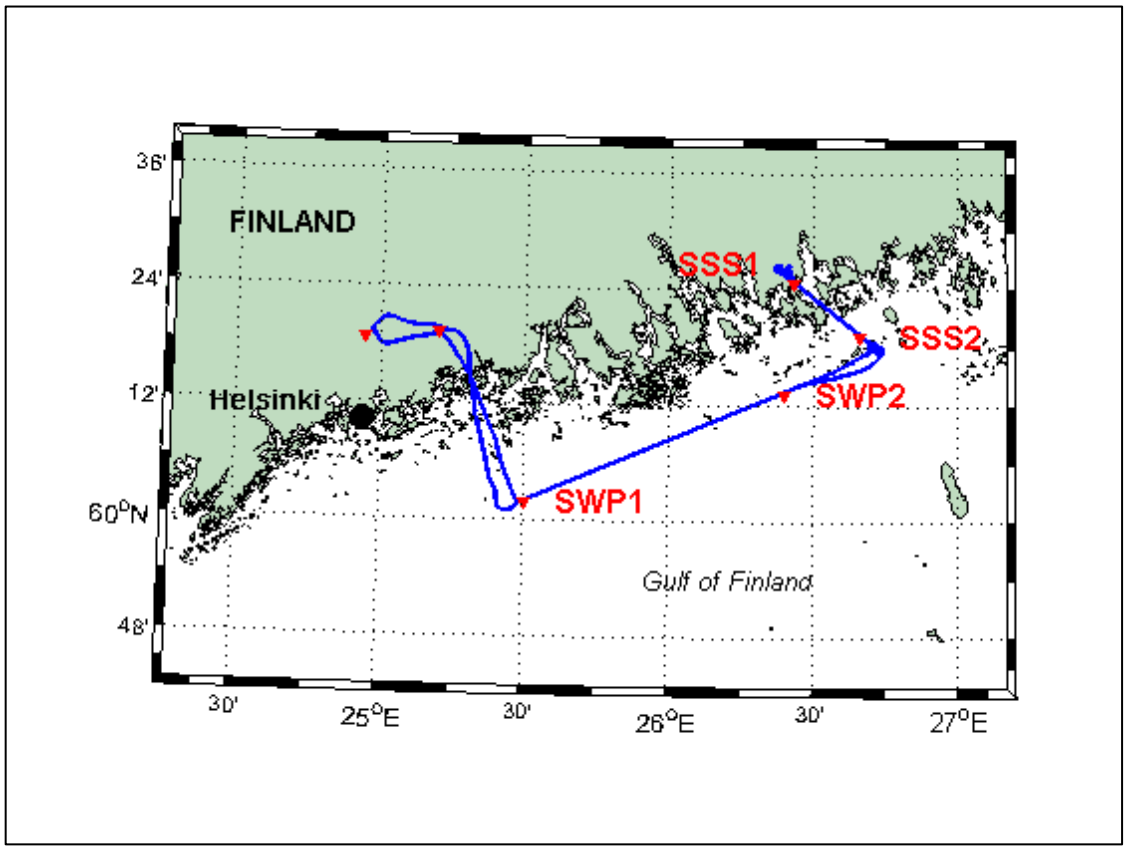


Figure 29: Flight route for Campaign C

8. Method

8.1.General method

The software program Matlab (version R2011a) was used to process radiometric and GNSS-R measurements and study possible correlations. The general method was the following:

1. First, data had to be averaged, in both time and space domains. Depending on the spatial resolution of the instruments and the deviations of the aircraft from its theoretical route, the test site was divided in a certain number of cells. In the case of campaign A for instance, cells resemble square patches of width 1km.

2. Ground reference data were interpolated at the middle point of each of those cells. The Klein & Swift model was then used to compute the component of the first Stokes parameter that is solely due to the emission of flat sea surface $I_{flat,cell}(SST_{cell}, SSS_{cell})$.

Approach A:

One approach is to remove this quantity from the radiometric measurements, leaving as residual component the emission that is due to the roughness solely (with an offset value due to the CMB radiation). This quantity is referred to as "first Stokes residual".

Approach B:

Another approach is to remove the impact of salinity and temperature gradients only. The first Stokes parameter that is computed at the first cell (and for the first overpass in case of campaigns A and C) acts as a reference value $I_{flat,reference}$. For every other cell, the additional emission $I_{flat,cell}(SST_{cell}, SSS_{cell}) - I_{flat,reference}$ is removed from the radiometric measurements, leaving as residual emission a component whose variations are due to the roughness solely.

3. Every cell (and at each overpass in case of campaigns A and C) was then assigned one value for the first Stokes and one value for each GNSS-R observable. This allowed possible correlation to be studied.

8.2.In more details

- Independent processing of the data, from scratch

The first milestone of this thesis was to investigate the same data that were already looked into at RAD department, and reproduce the results that were achieved, with an independent approach for the processing.

Correlation with MSS

One first article was recently published about Campaign A data ([1]). Using a least square fit between EMIRAD radiometric measurements and MSS data, it shows that the first Stokes is a linear function of MSS, save some offset value due to the CMB and flat surface emission (Approach B was used).

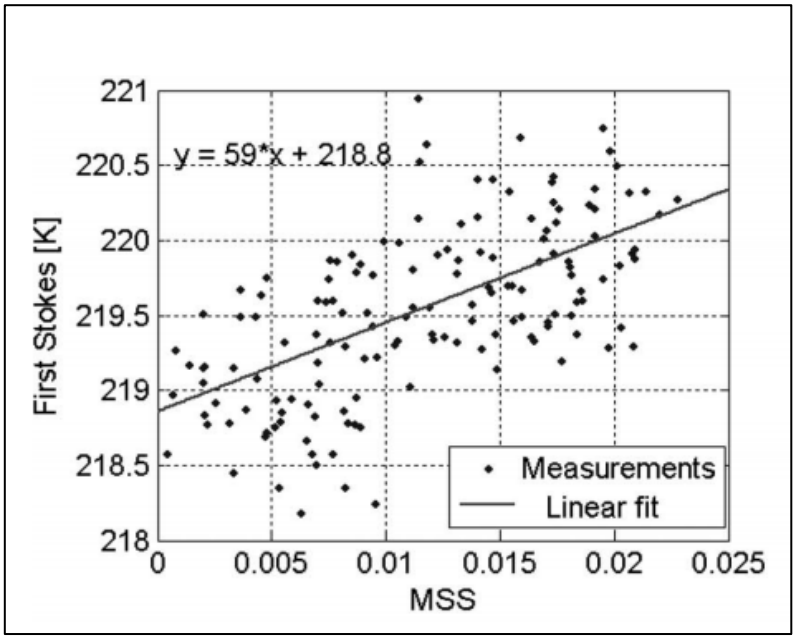


Figure 30: First Stokes parameter (EMIRAD data) as a function of MSS simultaneous measurements (GOLD-RTR data), Campaign A [1]

Table 2: Linear fitting information relative to Figure 30

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MSS + \beta_{day}$	**Correlation coefficient**
Day 1	?	0.54
Day 2	$y_{day2} = 59 \cdot MSS + 219$	0.69

Correlation with the area under the waveform

A second article was published during the thesis. It also handles Campaign A data, although only the first day experiment is considered. Results are not presented in the same way as they are in this thesis: the component I_{CMB} is taken care of prior to processing and First Stokes residuals – with a factor one-half – are plotted as a function of the increment of the waveform area (see section 6.2.3), expressed in terms of chip length. This increment is the additional value of the area with respect to the theoretical area under the direct signal's normalized waveform (with a threshold equal to 0.2). Anyhow, although even collocation of data is different in this article, results can be compared by means of variation range: residuals in [15] indeed show variations of around 1K over the whole range of waveform areas computed for the first day, with a correlation coefficient of 0.41 (see Figure 31).

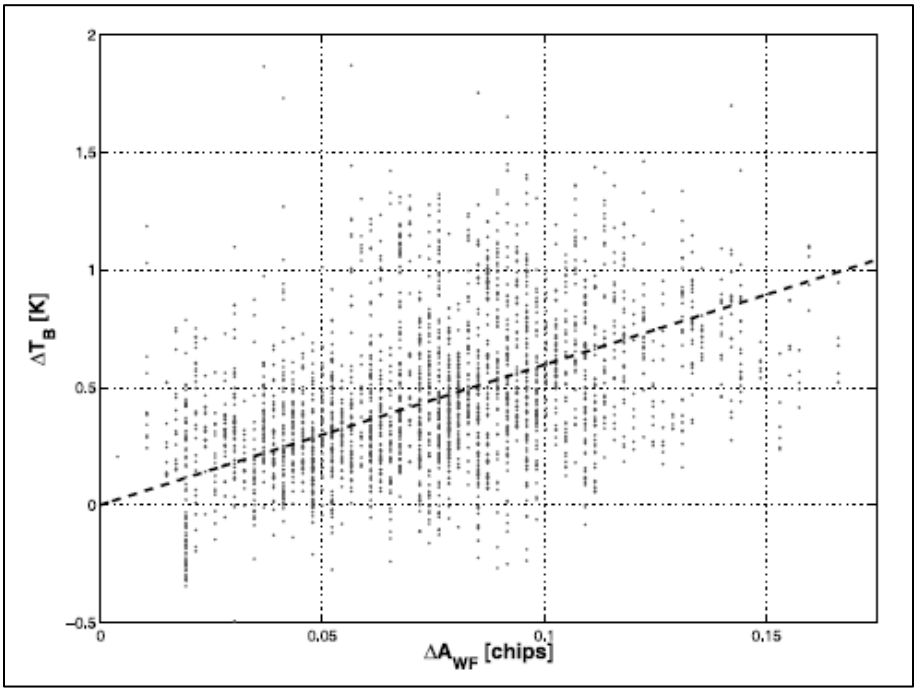


Figure 31: Half first Stokes residuals as a function of the waveform area's increment ([15])

The fact that the code was written from scratch – with no knowledge of the existing codes – guarantees unbiased results, and somehow means that data that were to be handled had to be perfectly understood and seized before being processed.

- Filtering of the corrupted data

Considering the high noise of the radiometric measurements, perfect correlation is not expected. A correlation that can be observed with eyes only, that is: a fitting for which correlation coefficient is poor although correlation is clearly observable, is satisfactory enough.

However, in order to filter out as many corrupted data as possible, measurements made over areas that are very close to the shoreline were sometimes discarded. One reason for this is the changes in the aircraft's attitude and route, which can make the radiometer pick up radiation coming from the land, instead of the sea solely. Another reason is the probability of RFI contamination of the signals, which is of course higher in populated areas (thus on land). Posterior to this filtering, measurement points that were clear outliers (as compared to the rest of the data) were furthermore discarded.

Last, a rough estimation of some variables could be made. Because GOLD-RTR can sometimes fail in the real-time processing of some variables (especially $MaxDerDelay$), many measurements turned out not to be realistic and could be discarded.

- Partitioning of the datasets

It was necessary to separate sets of data that cover a time period longer than 2 hours. If the experiment consisted of measurements made over several days (such as Campaign A), or over an entire day (such as Campaign B), data should be split into different parts prior to processing. Offsets in the radiometric

measurements are indeed expected for such campaigns: whenever the ground conditions change (e.g. the illumination by the Sun), the calibration of the instrument is annihilated, hence offsets appear. What's more is the systematic error of the measurements that can change over a long time period.

- Collocation of GNSS-R data

Both radiometric and GNSS-R data were collocated in time and space domains, as explained previously. However, when it comes to GNSS-R specifically, it was sometimes necessary to also identify which radio link every measurement belonged to, namely: which down-looking antenna was used and to which GPS satellite this signal corresponded (PRN code number). What's more is the elevation of the satellite in the sky: considering the noise of the data, it is better not to discard a too large amount of points, that's why even waveforms computed for signals of rather low-elevation GPS satellites were taken into account.

- Collocation of HUT-2D data for Campaign A

As mentioned is section 6.1.1, in the special case of HUT-2D, the First Stokes parameter has to be constructed from the X- and Y-components of the brightness temperature, measured in the antenna frame. Although very coarse, the collocation of the data was such that, for each cell c defined on the ground, all X-measurements $T_{b,X}$ falling in the bin defined by the triplet {cell c – polarization rotation angle $\delta\phi$ – incident angle [0°,10°] or [10°,25°]} were averaged, and added to their Y-counterpart $\mathrm{mean}_{\delta\phi}(T_{b,Y})$. All the resulting first Stokes $I_{\delta\phi}$ were then averaged so that cell c was assigned one single first Stokes value I. For better understanding of the method, see Figure 32. Such a method is admittedly coarse, but the usual high noise of HUT-2D measurements does not justify more time to be spent on the processing, especially since focus was to be given to possible correlation between EMIRAD data and GNSS-R observables *other than* MSS.

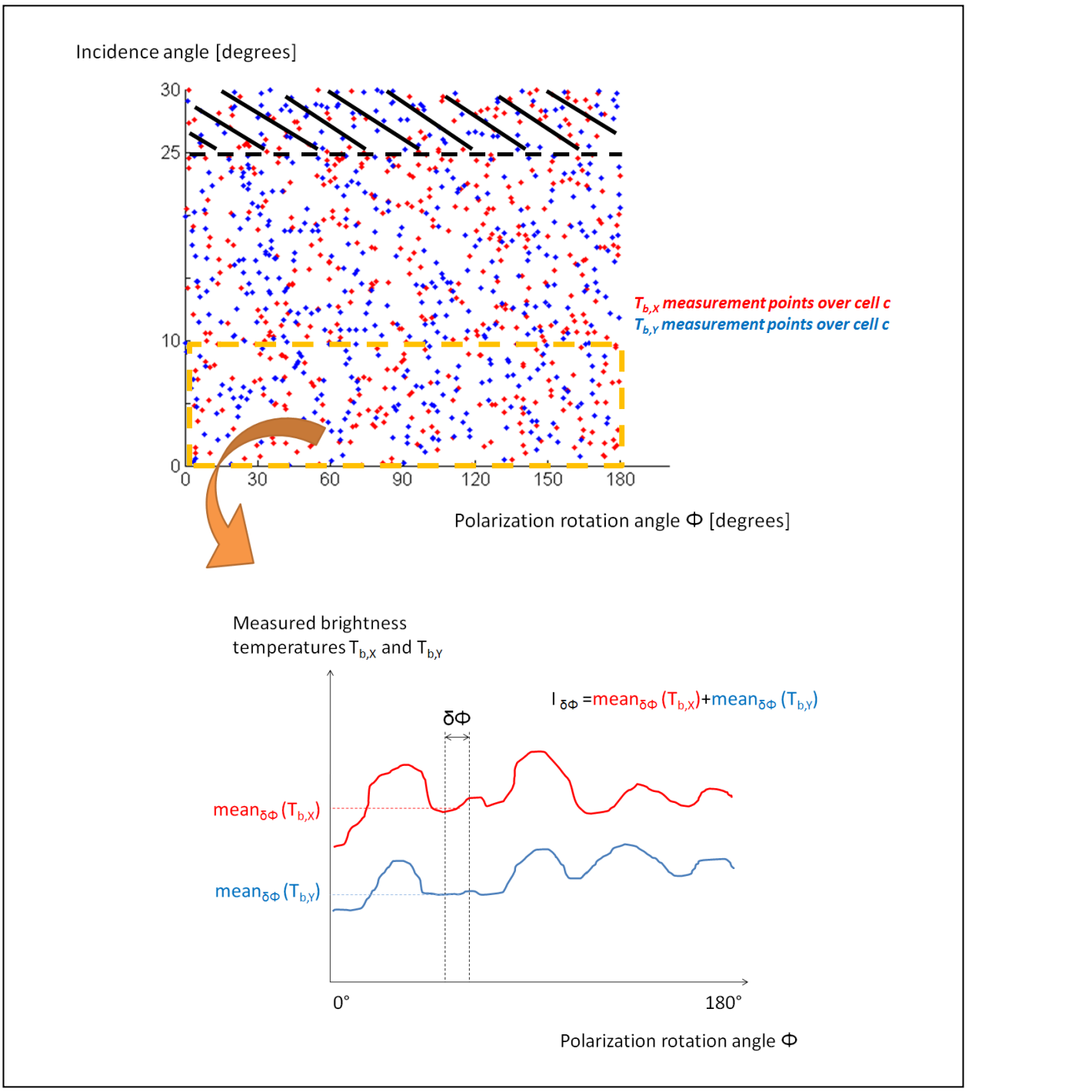


Figure 32: Collocation and averaging method for HUT-2D data

9. Presentation and discussion of results

9.1.Campaign A

9.1.1. Delay map parameters

Sections 3.2.6 and 6.2.3 give a description of the delay map parameters. The present section presents achieved results, whether they show correlation or not.

The test site is divided in 15 square cells of width 1km, centered on the planned flight route. Every data point (for both radiometric and GNSS-R measurements) is assigned the cell number it belongs to, and the overpass number it corresponds to. After doing so, points that were tagged for the same cell and the same overpass are averaged. This gives one pair {first Stokes, GNSS-R observable} per cell and overpass.

9.1.1.1. *MSS estimates*

EMIRAD radiometric measurements as a function of MSS

The linear correlation between the first Stokes parameter and MSS was reproduced, and results are even enhanced in comparison to those presented in 8.2 (corresponding correlation coefficients are higher). Throughout the whole thesis, data were processed using Approach A; yet, in order to compare the results of the two processing algorithms and better visualize the achieved enhancement, Approach B is first used in the present section.

Figure 33 presents the first Stokes parameter as a function of MSS, using approach B for the EMIRAD data. The observed variations are due to the roughness solely (as explained in section 8.1). Despite the offset between the first-day and second-day measurements, the first Stokes is clearly linear with MSS, the slopes being very similar for both days.

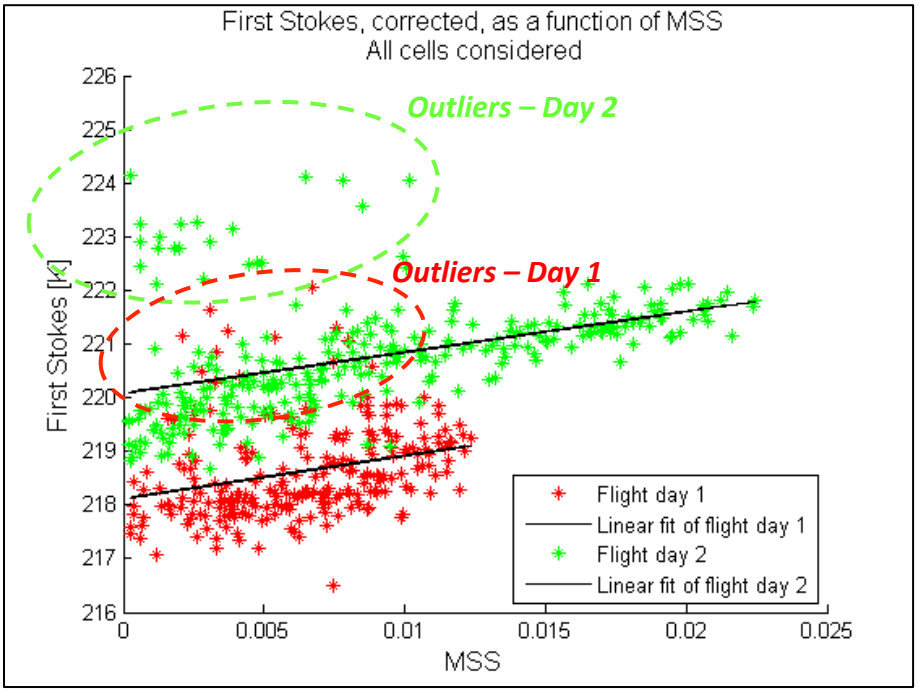


Figure 33: First Stokes parameter as a function of MSS

Table 3: Linear fitting information relative to Figure 33

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MSS + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 80 \cdot MSS + 218$	0.2787
Day 2	$y_{day2} = 76 \cdot MSS + 220$	0.4412

It is rather peculiar to observe a cloud of fairly high first Stokes measurements at low values of MSS, this on both days (Figure 33 hardly shows it, because data points of the second day cover up those of the first day, but two clouds are observed). Those outliers are assumed to be corrupted data.

When measurements made over areas that are close to the shoreline are discarded, results are even more enhanced. Figure 34 presents the 15 cells uniformly distributed along the flight route. Figure 35 is a picture of the test site, at the transition between the bay and the open sea: one can observe that this passage is really narrow, and that two small islands are even in the way (other islands, or simply the shoreline, are also very close to the route). Therefore, when data relative to cells 5 to 9 are not taken into account, correlation coefficients clearly increase, for both days, as shown in Table 4. A quick comparison of Figure 33 and Figure 36 shows that data relative to the middle cells are indeed corrupted: both clouds of outliers disappear when discarding those cells. A clear increase of the slope can also be observed, which suggests that the much lower slope computed in [1] might be a consequence of too much noise and corrupted data (although the outlier

clouds are not present on Figure 30). Hereafter in the thesis, middle cells 5 to 9 are always discarded, unless specified otherwise.

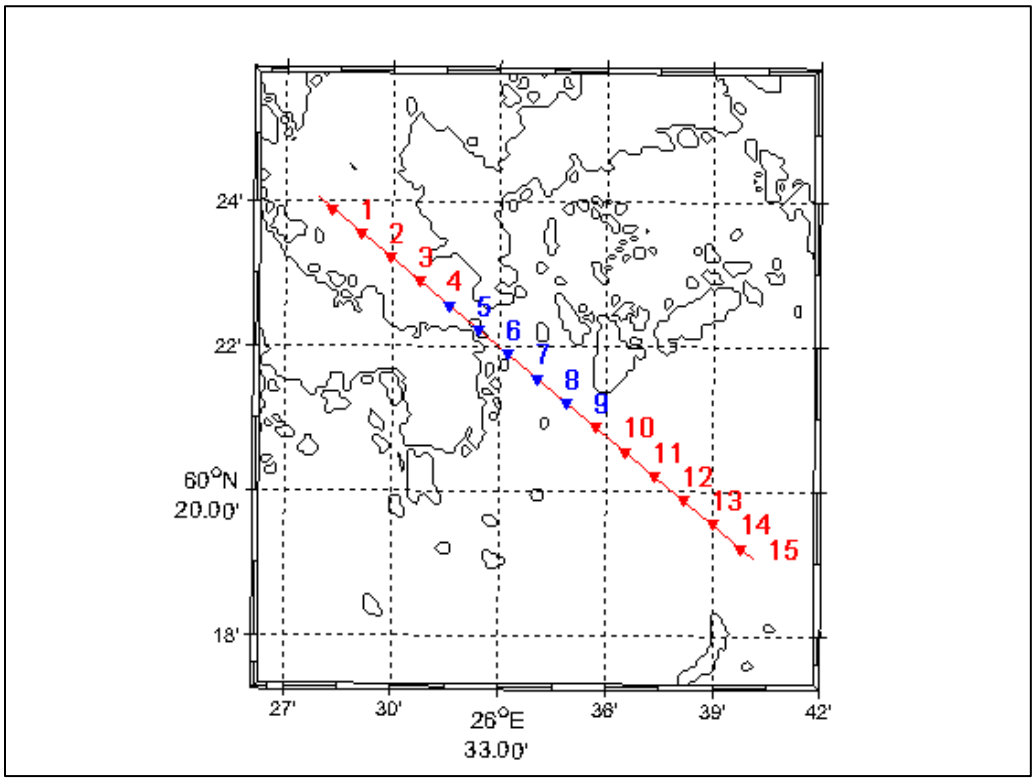


Figure 34: Campaign A: distribution of the 15 cell centers along the route

Figure 35: Transition between the bay and the open sea, with the two small islands circled

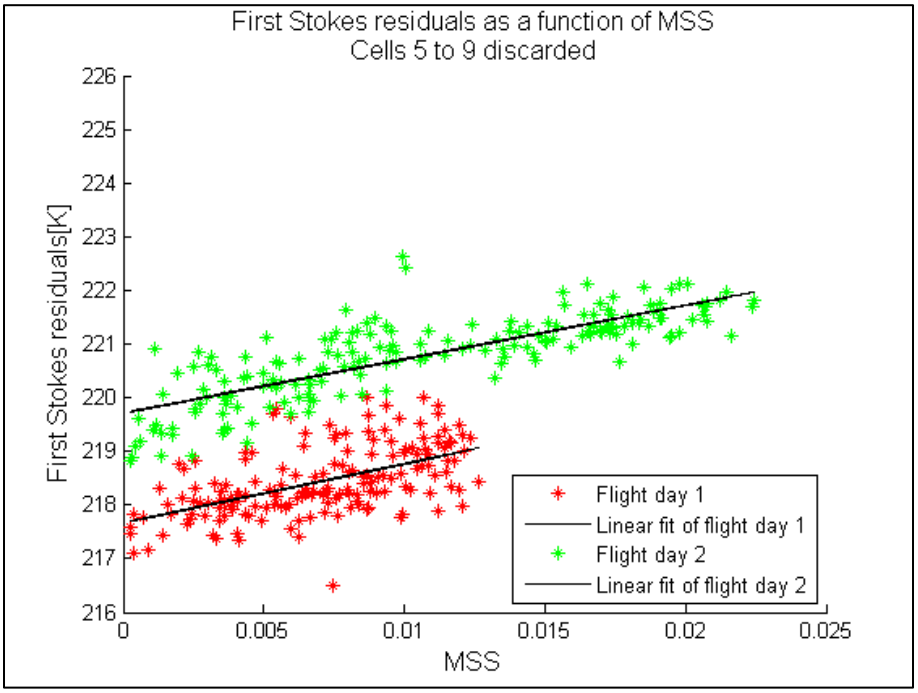


Figure 36: First Stokes parameter as a function of MSS, when middle cells 5 to 9 are discarded

Table 4: Linear fitting information relative to Figure 36

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MSS + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 110 \cdot MSS + 218$	0.5430
Day 2	$y_{day2} = 100 \cdot MSS + 220$	0.8051

A further step is to align the two sets of data, that is: to remove from all measurements made on the second day the constant value $\beta_{day2} - \beta_{day1} = 2$, and to perform a least-square fit on the whole dataset (see Figure 37 and Table 6). Considering the very similar slope values found after fitting separately the data of both days, it is naturally expected that "merging" the two datasets together will not have much effect whatsoever, save the correlation coefficient, which will be somewhere between 0.5430 and 0.8051.

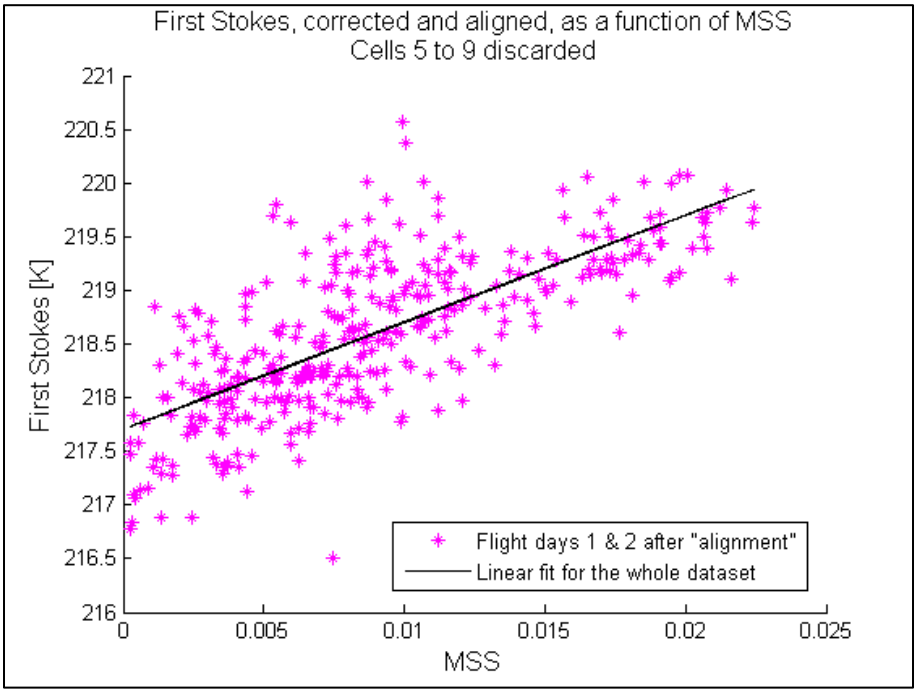


Figure 37: First Stokes parameter as a function of MSS, fitting both days altogether

Table 5: Linear fitting information relative to Figure 37

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MSS + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1 & 2	$y_{day1\&2} = 100 \cdot MSS + 218$	0.7269

As expected, results achieved after "alignment" of the radiometric measurements lead to very similar correlation (save the correlation coefficient is a bit lower), confirming the robustness of the linear model. In the rest of the thesis though, there will always be clear distinction between day 1 and day 2. The same thing goes for other campaigns, which can contain parts that are distant from each other in space and/or time.

Lastly, the results are presented in terms of first Stokes residuals, for it is Approach A which is actually used thorough the thesis (see Figure 38 and Table 6). Correlation coefficients are the same as those presented in Table 4, since results obtained with Approach A differ from those obtained with Approach B by only a constant ($I_{flat,reference}$, see section 8.1).

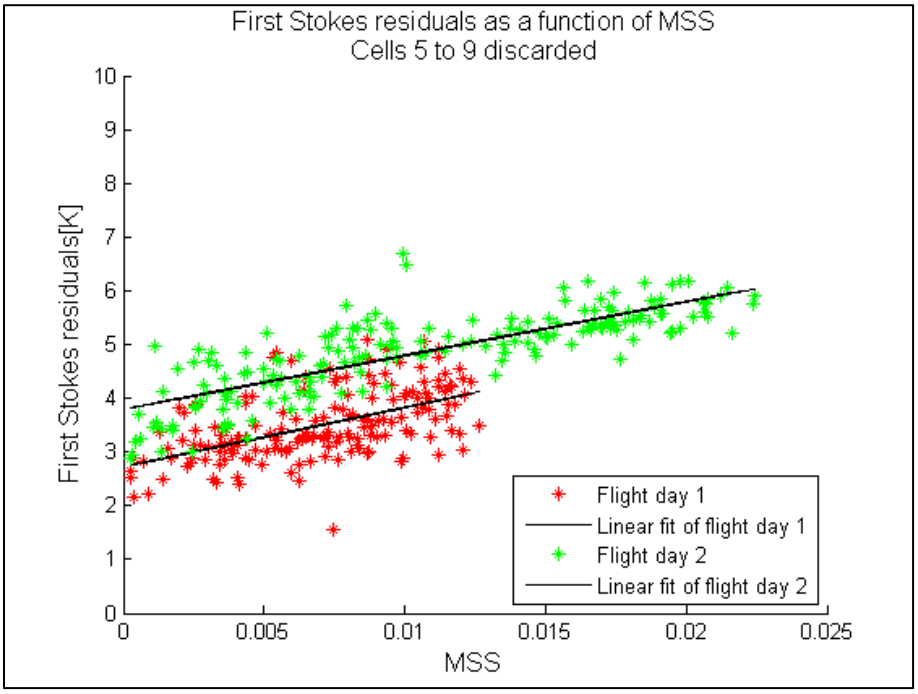


Figure 38: First Stokes residuals as a function of MSS

Table 6: Linear fitting information relative to Figure 38

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MSS + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 110 \cdot MSS + 2.8$	0.5430
Day 2	$y_{day2} = 100 \cdot MSS + 3.8$	0.8051

One can observe that the offset between radiometric measurements of the first and second days (around 1K) seems lower than what is observed on Figure 36 (around 2K). The reason for this is the different sea state conditions SST and SSS from one day to the other. Both SST and SSS values are such that the modeled emission of the flat sea is higher on the second day than on the first (higher SST, lower SSS, see the consequent change of the first Stokes on Figure 8). As a result, more radiation I_{flat} is to be removed from the second day measurements when going from $I_{flat} + I_{rough}$ to residuals I_{rough}, and this clearly explains the difference in the offset.

As for the offset itself, it can be explained, to some extent, by the change of the component I_{sky} from one day to the other. As can be seen from Figure 9, models predict an average increase in I_{sky} of around 0.1K on the second day. The remaining 0.9K offset may be the consequence of a poor calibration of the radiometer, but this remains an open issue.

HUT-2D radiometric measurements as a function of MSS

The collocation of HUT-2D data is explained in section 8.2. Figure 38 presents the achieved results, which are quite satisfactory as regards the second day. As for the first day, if the cloud of low First Stokes values around $MSS = 0.01$ is discarded, and if the resulting linear fit is made with the eyes, it becomes similar to all correlations observed so far. Figure 40 presents the standard deviation values when the measurements are put into MSS bins of width 0.0025, confirming the high noise level of the measurements.

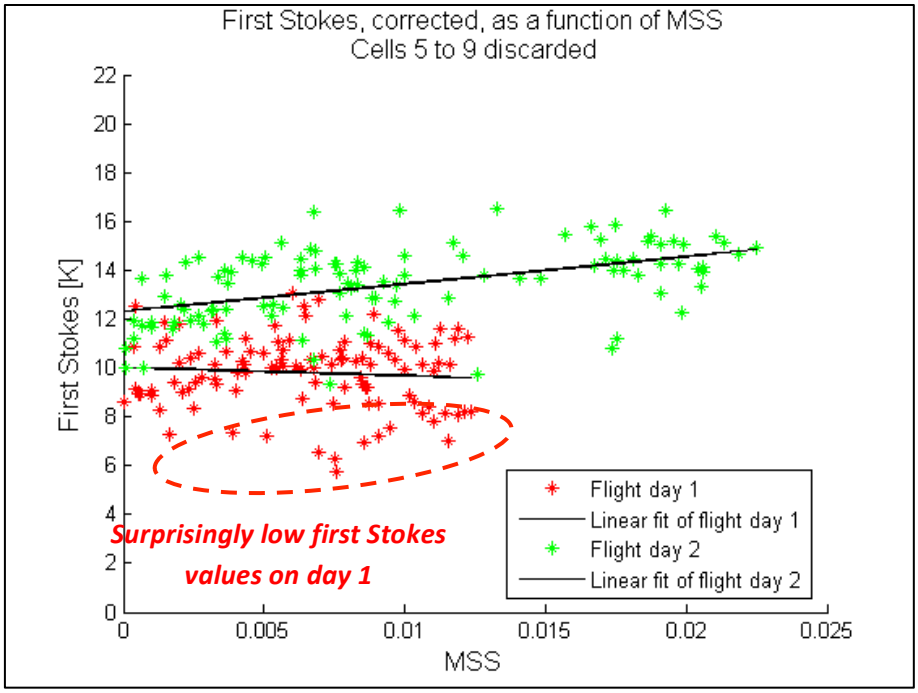


Figure 39: First Stokes residuals (HUT-2D data) as a function of MSS

Table 7: Linear fitting information relative to Figure 39

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MSS + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = -33.5 \cdot MSS + 10$	-0.0797
Day 2	$y_{day2} = 111.8 \cdot MSS + 12.3$	0.4762

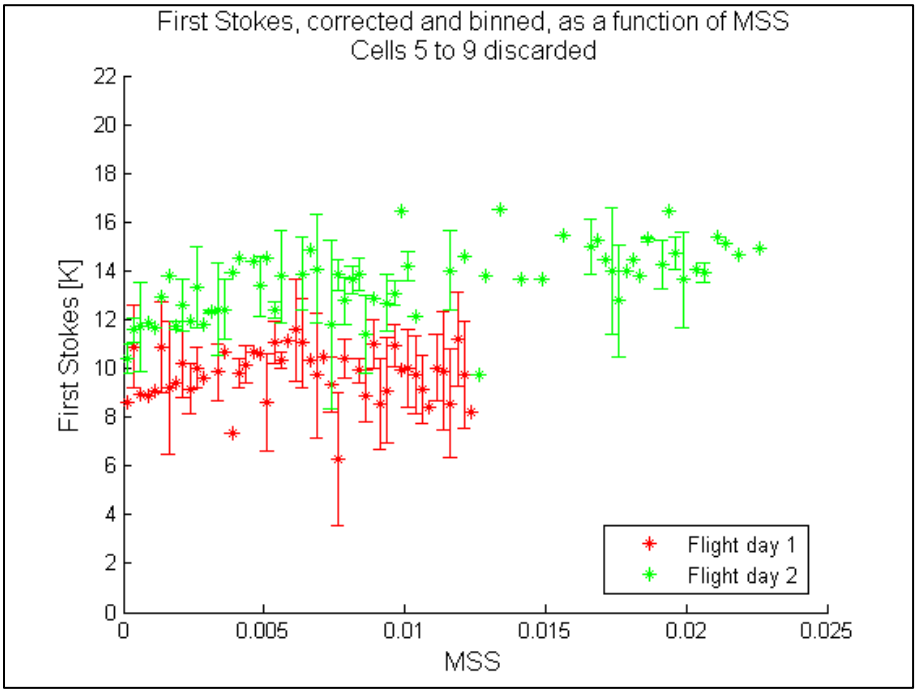


Figure 40: Median values of the first Stokes (HUT-2D) after data were binned

9.1.1.2. *Delay of the waveform's peak, or PeakDelay*

As mentioned in section 8.2, it is essential to process $PeakDelay$ for each PRN code separately, because this variable is defined with respect to the time of reception of the direct signal. The delay between the direct and reflected signals is affected by the position of the satellite in the sky (even though all selected satellites have quite high elevation), thus by the PRN code.

Correlation of the first Stokes with $PeakDelay$ is not evident. Only results for which there is a reasonable amount of data points are presented below (depending on its elevation, a GPS satellite may be selected by GOLD-RTR for a much longer time than another one); this corresponds to five datasets out of seven.

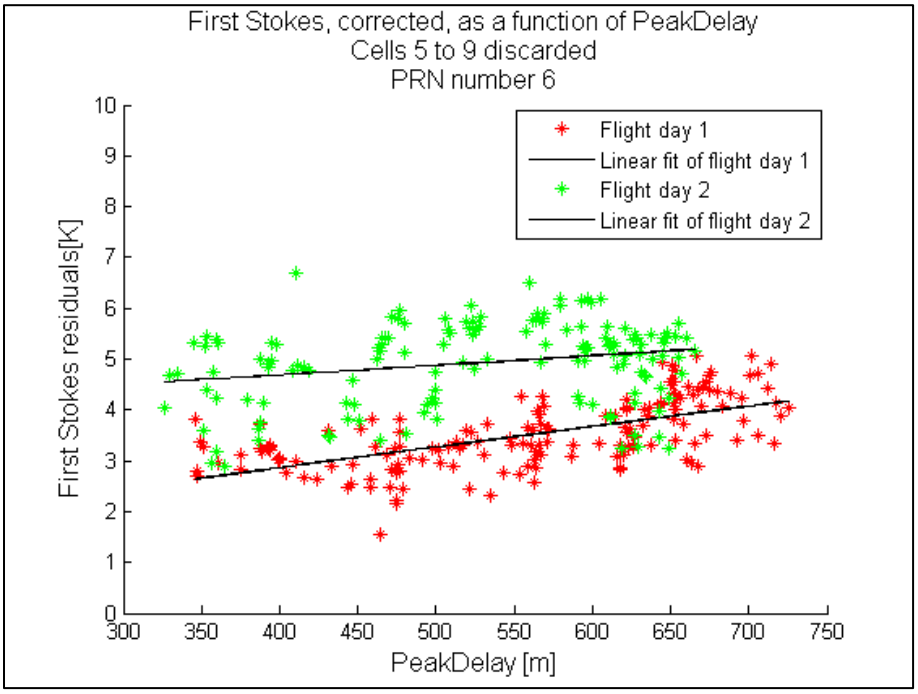


Figure 41: First Stokes residuals as a function of $PeakDelay$, PRN code number 6

Table 8: Linear fitting information relative to Figure 41

	Linear fit equation: $y_{day} = \alpha_{day} \cdot PeakDelay_{PRN\,6} + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 0.4 \cdot 10^{-2} \cdot PeakDelay_{PRN\,6} + 1.3$	0.6299
Day 2	$y_{day2} = 0.2 \cdot 10^{-2} \cdot PeakDelay_{PRN\,6} + 4.0$	0.2297

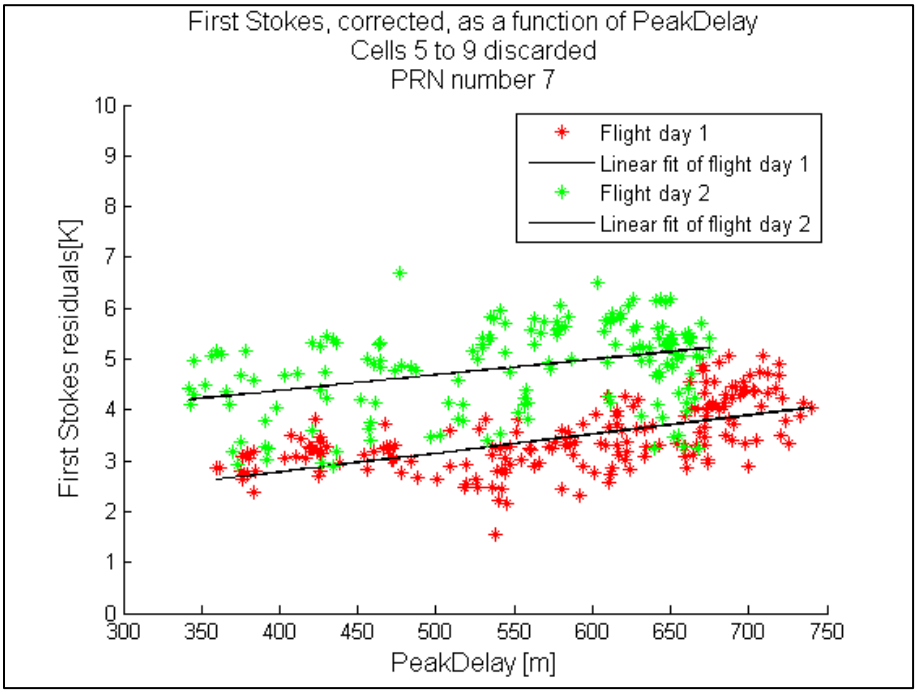


Figure 42: First Stokes residuals as a function of $PeakDelay$, PRN code number 7

Table 9: Linear fitting information relative to Figure 42

	Linear fit equation: $y_{day} = \alpha_{day} \cdot PeakDelay_{PRN\,7} + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 0.4 \cdot 10^{-2} \cdot PeakDelay_{PRN\,7} + 1.3$	0.6065
Day 2	$y_{day2} = 0.3 \cdot 10^{-2} \cdot PeakDelay_{PRN\,7} + 3.2$	0.3622

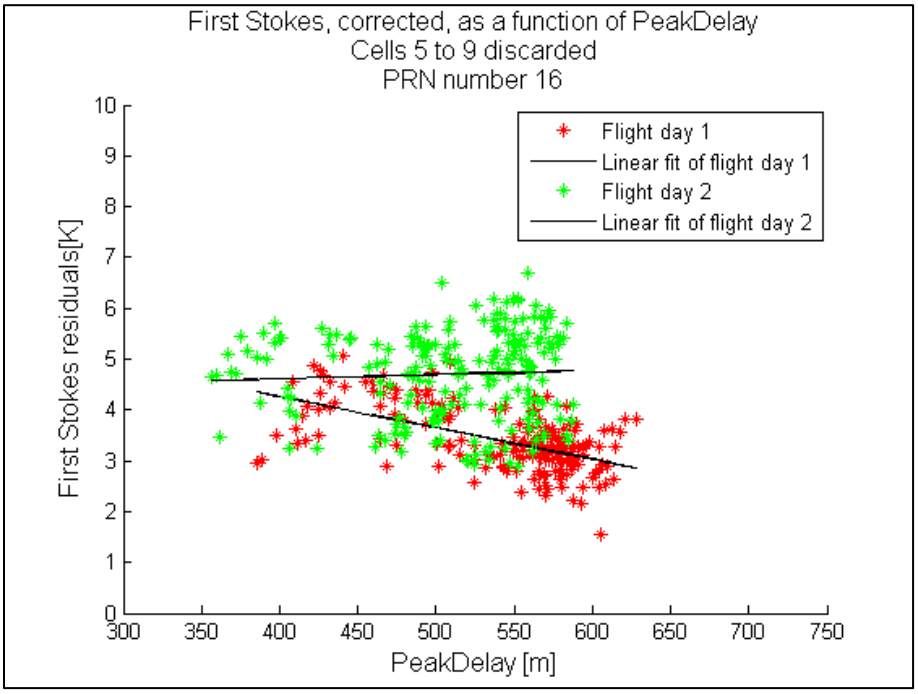


Figure 43: First Stokes residuals as a function of $PeakDelay$, PRN code number 16

Table 10: Linear fitting information relative to Figure 43

	Linear fit equation: $y_{day} = \alpha_{day} \cdot PeakDelay_{PRN\ 16} + \beta_{day}$	**Correlation coefficient** r_{day}
Day 1	$y_{day1} = -0.6 \cdot 10^{-2} \cdot PeakDelay_{PRN\ 16} + 6.7$	-0.5696
Day 2	$y_{day2} = 0.08 \cdot 10^{-2} \cdot PeakDelay_{PRN\ 16} + 4.3$	0.0525

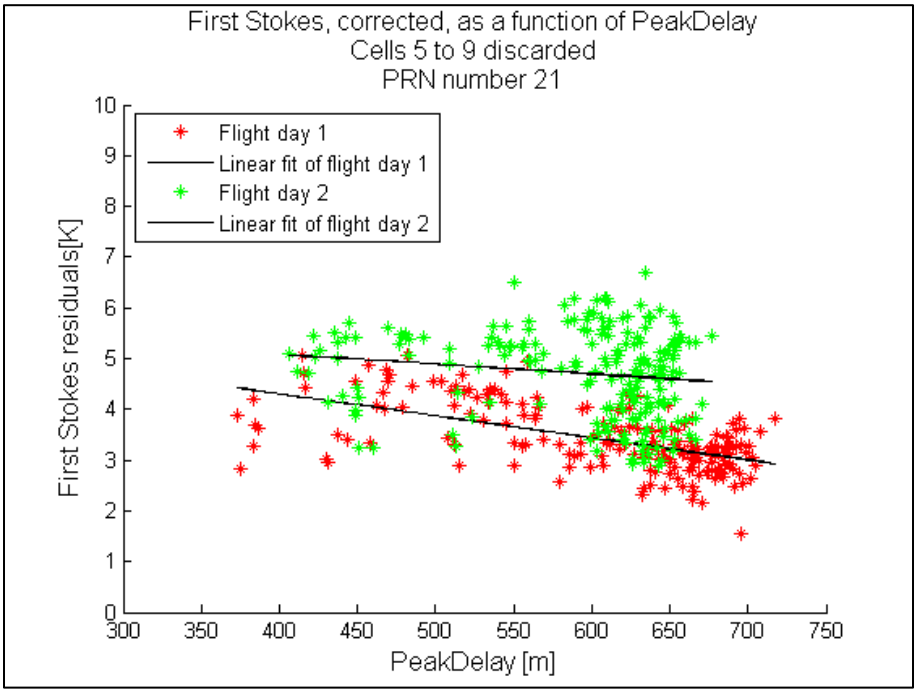


Figure 44: First Stokes residuals as a function of $PeakDelay$, PRN code number 21

Table 11: Linear fitting information relative to Figure 44

	Linear fit equation: $y_{day} = \alpha_{day} \cdot PeakDelay_{PRN\ 21} + \beta_{day}$	Correlation coefficient r_{day}
Day 1	$y_{day1} = -0.4 \cdot 10^{-2} \cdot PeakDelay_{PRN\ 21} + 6.0$	-0.5833
Day 2	$y_{day2} = -0.2 \cdot 10^{-2} \cdot PeakDelay_{PRN\ 21} + 5.9$	-0.1529

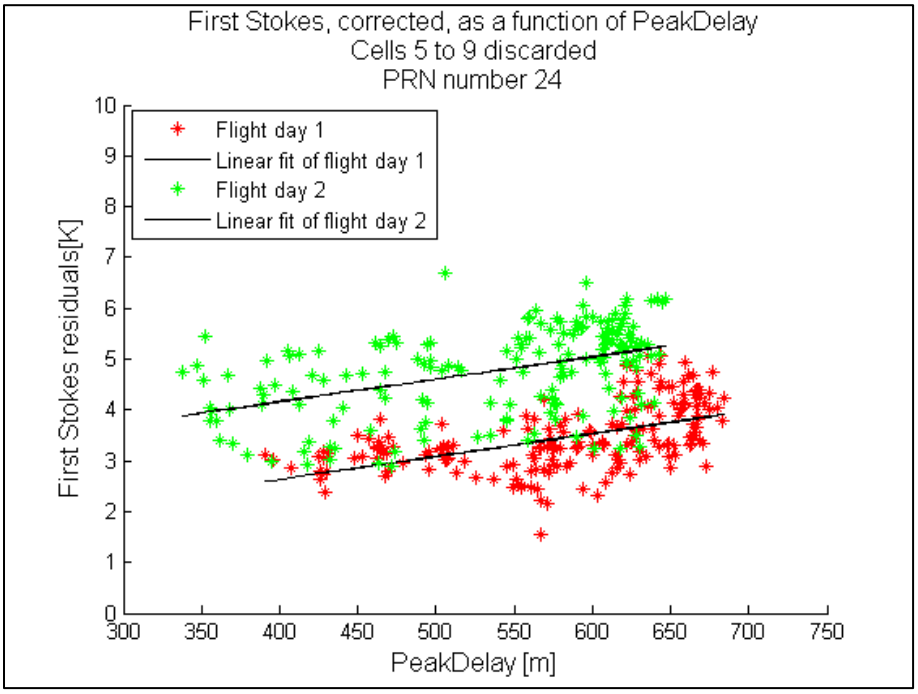


Figure 45: First Stokes residuals as a function of $PeakDelay$, PRN code number 24

Table 12: Linear fitting information relative to Figure 45

	Linear fit equation: $y_{day} = \alpha_{day} \cdot PeakDelay_{PRN\ 24} + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 0.4 \cdot 10^{-2} \cdot PeakDelay_{PRN\ 24} + 0.8$	0.5324
Day 2	$y_{day2} = 0.4 \cdot 10^{-2} \cdot PeakDelay_{PRN\ 24} + 2.4$	0.4481

Linear fittings which yield positive – and rather reasonable – Pearson correlation coefficients have admittedly similar slopes (PRN codes 6, 7 and 24); however, no clear conclusion can be made for two important reasons:

- It is strange not to observe a similar fitting outline for all other PRN codes. There can be offsets in the range of $PeakDelay$ values (due to the different positions of the satellites), but there should not be major differences in the overall outline (the first Stokes is expected to increase with the roughness, meaning $PeakDelay$).
- IEEC did indicate that the computation of the waveform might sometimes fail for the conditions of Campaign A, so the information contained in the waveforms are not valuable. Some PRN signals cannot be separated properly, and data are somehow mixed one with each other.

Instead of claiming that a correlation between the first Stokes and $PeakDelay$ exists, precious information can be withdrawn from the data. It clearly looks like signals corresponding to the PRN codes 16

and 21 were not separated or processed properly (decreasing emission of the sea with $PeakDelay$), as opposed to PRN codes 6, 7 and 24 (increasing emission of the sea with $PeakDelay$). Therefore, it is reasonable, from now on, to select data from the PRN codes 6, 7 and 21 only to process the other variables.

9.1.1.3. *SpecDelay*

$SpecDelay$ is another parameter referring to the direct signal, hence needs processing for each PRN code separately. Results achieved are pretty similar to those achieved with $PeakDelay$, in the sense that correlation seems to be observed for some of the PRN codes (the same as with $PeakDelay$), but no conclusion can be made whatsoever. Results are only presented for the PRN codes 6, 7 and 21, as justified in section 9.1.1.2.

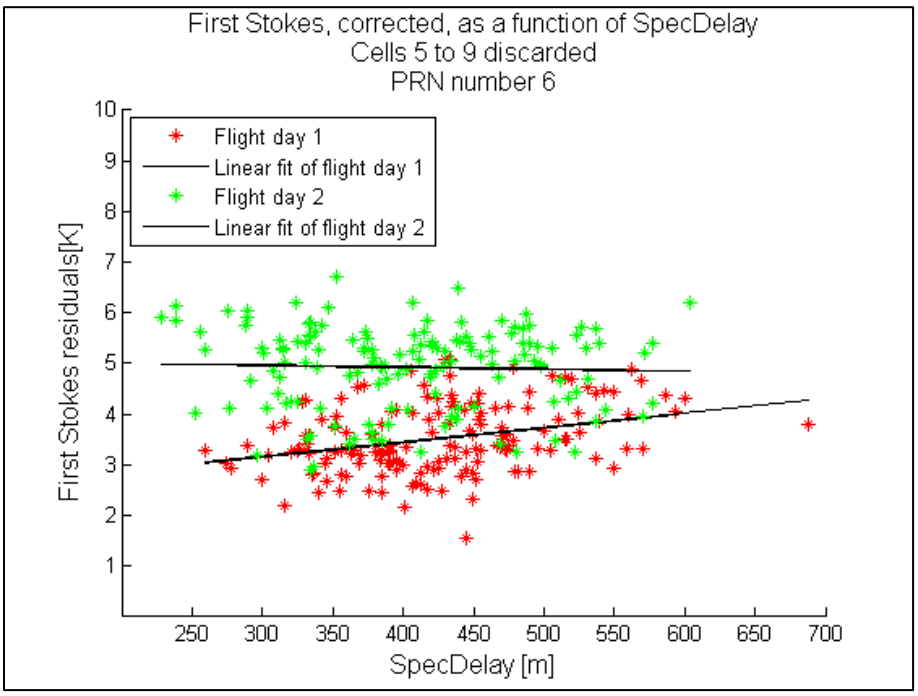


Figure 46: First Stokes residuals as a function of $SpecDelay$, PRN code number 6

Table 13: Linear fitting information relative to Figure 46

	Linear fit equation: $y_{day} = \alpha_{day} \cdot SpecDelay_{PRN\,6} + \beta_{day}$	Correlation coefficient r_{day}
Day 1	$y_{day1} = 0.3 \cdot 10^{-2} \cdot SpecDelay_{PRN\,6} + 2.3$	0.3360
Day 2	$y_{day2} = -0.04 \cdot 10^{-2} \cdot SpecDelay_{PRN\,6} + 5.1$	-0.0358

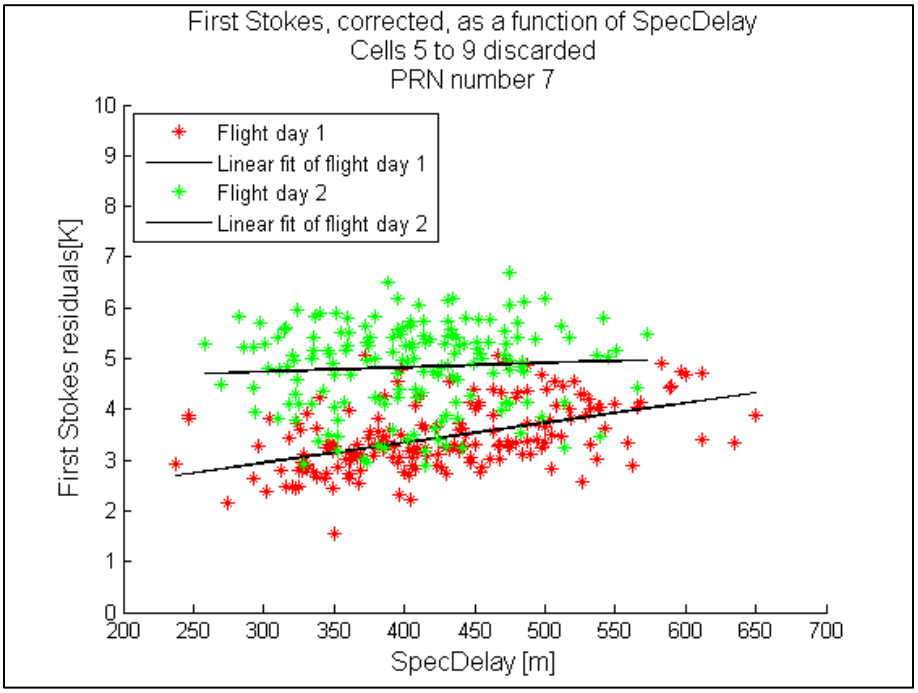


Figure 47: First Stokes residuals as a function of $SpecDelay$, PRN code number 7

Table 14: Linear fitting information relative to Figure 47

	Linear fit equation: $y_{day} = \alpha_{day} \cdot SpecDelay_{PRN\,7} + \beta_{day}$	**Correlation coefficient** r_{day}
Day 1	$y_{day1} = 0.4 \cdot 10^{-2} \cdot SpecDelay_{PRN\,7} + 1.8$	0.4939
Day 2	$y_{day2} = 0.08 \cdot 10^{-2} \cdot SpecDelay_{PRN\,7} + 4.5$	0.0669

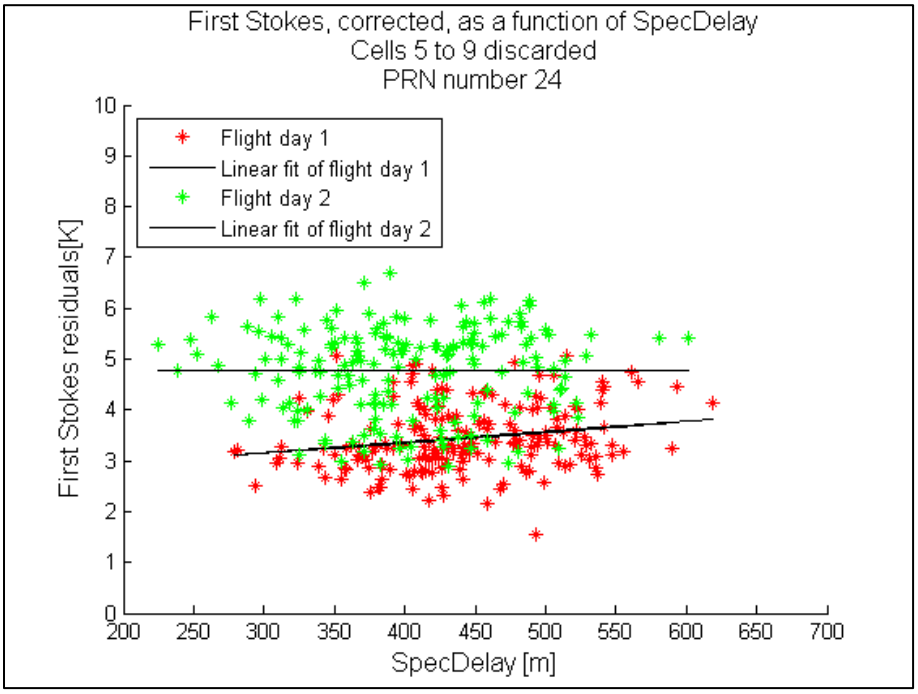


Figure 48: First Stokes residuals as a function of $SpecDelay$, PRN code number 24

Table 15: Linear fitting information relative to Figure 45

	Linear fit equation: $y_{day} = \alpha_{day} \cdot SpecDelay_{PRN\ 24} + \beta_{day}$	**Correlation coefficient** r_{day}
Day 1	$y_{day1} = 0.2 \cdot 10^{-2} \cdot SpecDelay_{PRN\ 24} + 2.5$	0.2077
Day 2	$y_{day2} = 0.05 \cdot 10^{-2} \cdot SpecDelay_{PRN\ 24} + 4.8$	0.0037

As can be observed, correlations are quite poor, and fitting functions have quite low slopes ($SpecDelay$ is expected to increase with the sea surface roughness, so the first Stokes residuals should clearly increase with $SpecDelay$).

9.1.1.4. *MaxWavDelay*

The collocation of measurements was made in time and space domains only: after investigation, it turned out that $MaxWavDelay$ data do not depend on the radio link considered (results are similar for all PRN codes 6, 7 and 24, and for both down-looking antennae). Therefore, $MaxWavDelay$ measurements that refer to the same cell and the same overpass are averaged altogether.

Figure 49 presents the achieved results.

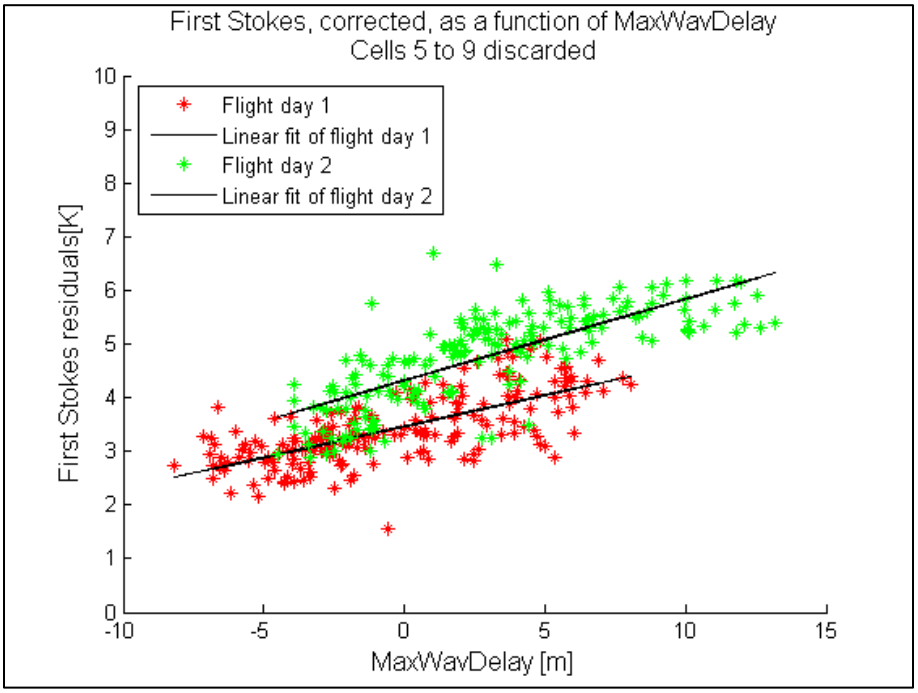


Figure 49: First Stokes residuals as a function of $MaxWavDelay$

Table 16: Linear fitting information relative to Figure 49

	Linear fit equation: $y_{day} = \alpha_{day} \cdot MaxWavDelay + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 12.6 \cdot 10^{-2} \cdot MaxWavDelay + 3.5$	0.7010
Day 2	$y_{day2} = 15.2 \cdot 10^{-2} \cdot MaxWavDelay + 4.3$	0.7610

Linear correlation is clearly observed between $MaxWavDelay$ and the first Stokes residuals, and correlations coefficients are quite satisfactory. However, the fact that $MaxWavDelay$ is defined with respect to the estimated specular point makes its meaning quite tangible, because it depends on the simplified models used by GOLD-RTR to estimate this specular delay. A better indicator is then $ScattDelay$, which gives the delay of the peak of the waveform, with respect to the actual specular point.

9.1.1.5. *ScattDelay*

$ScattDelay$ is computed from $MaxWavDelay - MaxDerDelay$. Although $ScattDelay$ is a better indicator of the roughness than $MaxWavDelay$, its computation often fails due to the difficult real-time computation of the maximum of the first derivative of the waveform ($MaxDerDelay$). As a result, many data

points are corrupted and shall be filtered out. A realistic value for $ScattDelay$ is a positive value below 100 meters. After this first filtering, it turns out that two very distinct clouds of points are present: one with values between 30 and 45 meters, the other with values between 65 and 85 meters. This observation cannot be physically explained, but points falling in the second cloud are more numerous and realistic. A scatterometric delay around 30 meters is clearly not possible, for it corresponds to a fraction of only 0.2 of the chip length, which does not even match the ideal theoretical case of a direct signal. As a result, this cloud of points is discarded and data are processed only with $ScattDelay$ values between 65 and 85 meters. Results achieved are presented on Figure 50.

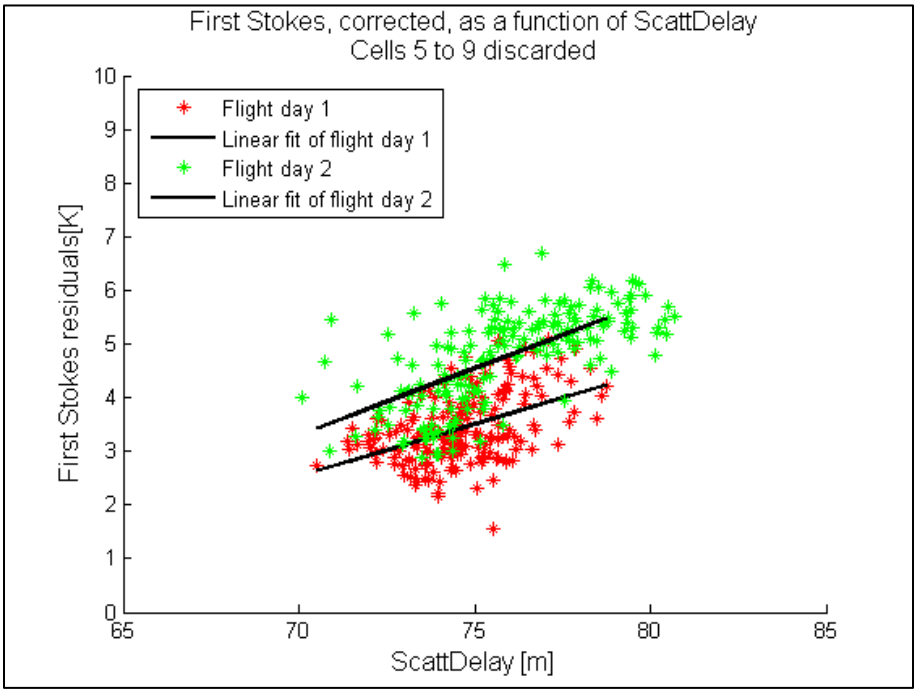


Figure 50: First Stokes residuals as a function of $ScattDelay$

Table 17: Linear fitting information relative to Figure 50

	Linear fit equation: $y_{day} = \alpha_{day} \cdot ScattDelay + \beta_{day}$	**Correlation coefficient** r_{day}
Day 1	$y_{day1} = 20 \cdot 10^{-2} \cdot ScattDelay - 11$	0.4612
Day 2	$y_{day2} = 25 \cdot 10^{-2} \cdot ScattDelay - 14$	0.6703

Pearson correlation coefficients are pretty satisfactory, and it can be concluded that the first Stokes residuals are a linear function of $ScattDelay$. Some could argue that in the processing of the other variables, data points that correspond to the measurements discarded in $ScattDelay$ shall also be discarded. However, it would be wrong to do so; first because filtering too many data so that final results look good is dishonest,

second because some of those measurements are actually valuable (the wrong estimation of $MaxDerDelay$ does not affect, for instance, the variables $TailDelay$ or $WavArea$).

9.1.1.6. *Waveform's tail time length, or TailDelay*

Here again, the collocation of measurements was made in time and space domains only: after investigation, it turned out that $TailDelay$ data do not depend on the radio link considered (neither on antenna, nor on PRN code).

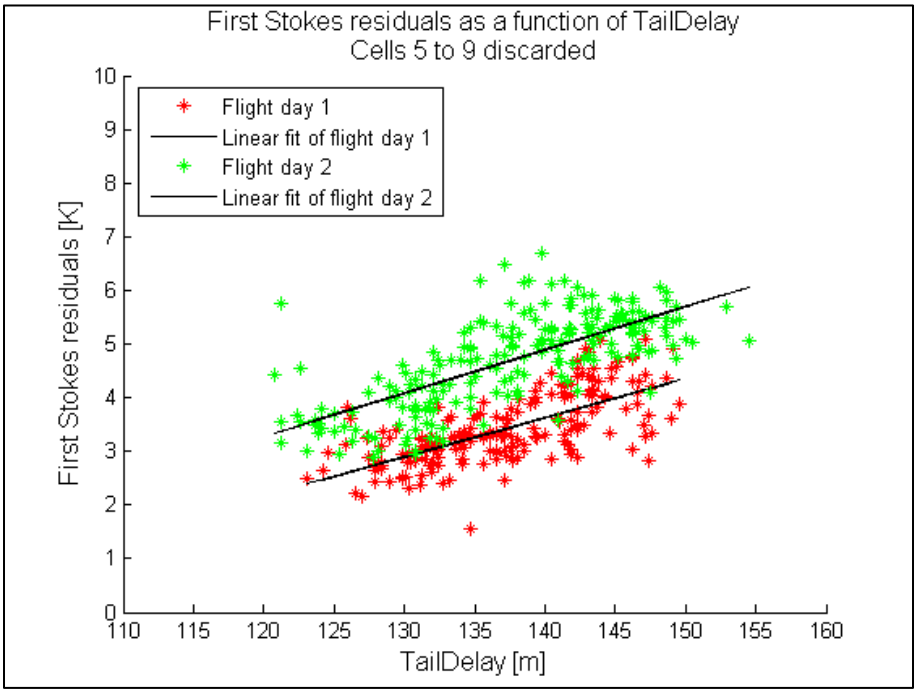


Figure 51: First Stokes residuals as a function of $TailDelay$

Table 18: Linear fitting information relative to Figure 51

	Linear fit equation: $y_{day} = \alpha_{day} \cdot TailDelay + \beta_{day}$	**Correlation coefficient r_{day}**
Day 1	$y_{day1} = 7 \cdot 10^{-2} \cdot TailDelay - 6$	0.7010
Day 2	$y_{day2} = 8 \cdot 10^{-2} \cdot TailDelay - 6$	0.6544

9.1.1.7. *Area under the waveform, or WavArea*

As it was observed in [15], [16] and [17], high constituency among computed waveform areas is observed from one PRN code to another. As a result, no collocation of the data is made according to the radio link: data for all PRN codes, and picked up by both antennae, are treated as a whole.

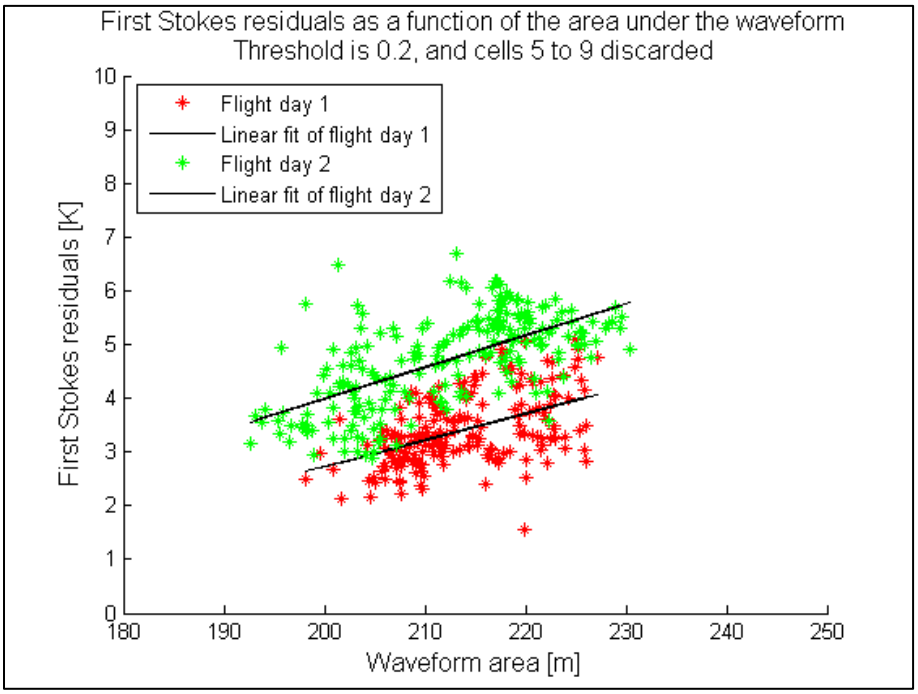


Figure 52: First Stokes residuals as a function of $WavArea$

Table 19: Linear fitting information relative to Figure 52

	Linear fit equation: $y_{day} = \alpha_{day} \cdot WavArea + \beta_{day}$	**Correlation coefficient** r_{day}
Day 1	$y_{day1} = 4 \cdot 10^{-2} \cdot WavArea - 6$	0.4970
Day 2	$y_{day2} = 5 \cdot 10^{-2} \cdot WavArea - 7$	0.5360

Taking into account the factor one-half from the First Stokes to the brightness temperature, results presented on Figure 52 agree relatively well with results of [15]: variation of half the first Stokes residuals is about 0.9K in the present work (first day), while it is about 1K in [15]. Moreover, correlation is also established on the second day, which is a new result.

9.1.2. Doppler Delay map parameters

As mentioned in section 6.2.3, DDMs have been constructed for the Doppler spread [-100Hz; +100Hz], with steps of 50Hz. The resulting maps do not clearly show the glistening zone, as can be seen from Figure 53. Computing the area or the volume of the DDM under a certain threshold makes sense if the contribution of the whole glistening area is included in the latter, and this fails in the present situation.

Additionally, IEEC informed Aalto that some DDMs were generated for Campaign A with a broader Doppler frequency range. It turned out that the Doppler spread of the glistening area is very little (above 100Hz, but much below 1KHz). This means that very little information about the roughness would be obtained from the DDM of this campaign.

.

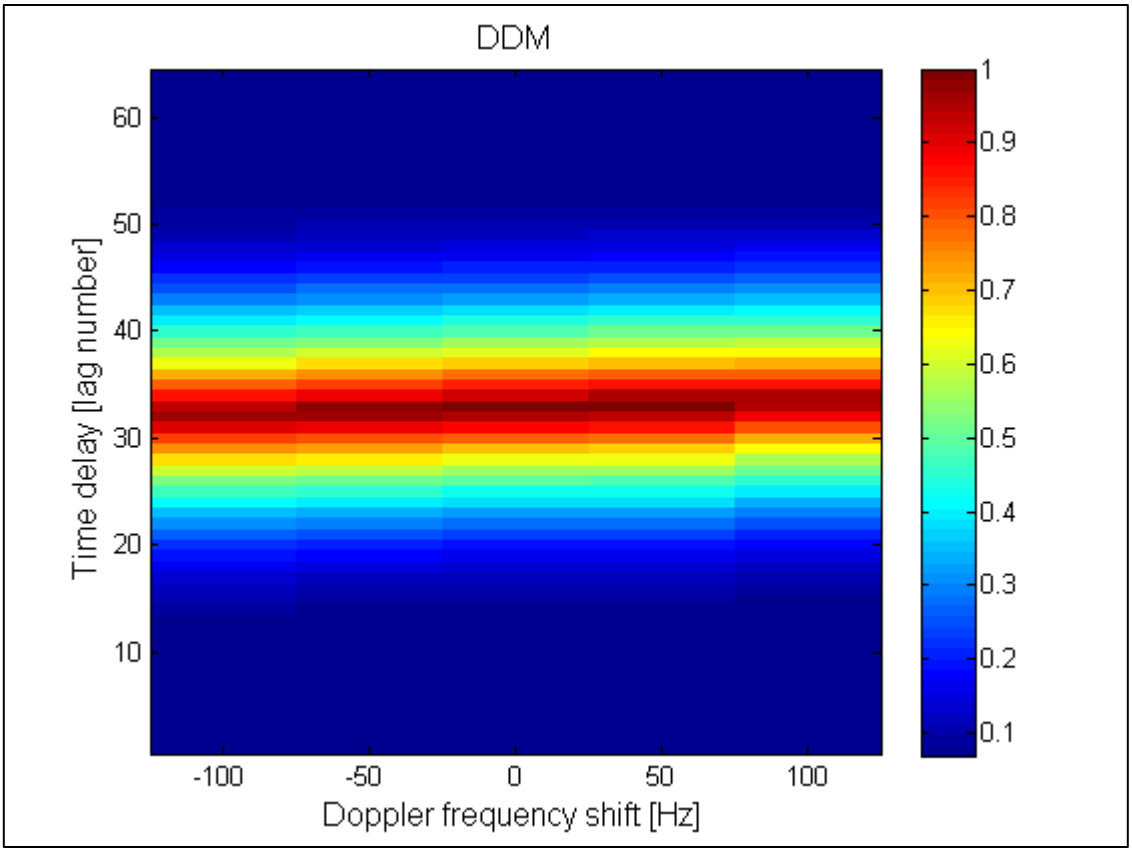


Figure 53: Example of computed Doppler Delay Map

9.2.Campaign B

Because the flight route is not a straight line in Campaign B, data were allocated differently compared to Campaign A. A 10-second incoherent averaging of both EMIRAD and GOLD-RTR data is performed, assuming that the coordinate points at which those measurements were made fall in a rather square ground patch of

width around 2km (which can be called a cell as well). This assumption is valid and justified by the aircraft's ground speed (about 100m.s^{-1}), the diameter of EMIRAD footprint (roughly 1500 meters), and its deviations from the route (within 1000m). As for MSS in particular, available data were already 10-second averaged, so for each MSS data point, EMIRAD data collected between five seconds before and five seconds after were averaged to form the counterpart to MSS.

Measurements for cells either over land, close to the shore line, or showing obvious abnormalities (first Stokes residuals higher than the expected value by more than 10K), were discarded.

9.2.1. Delay map parameters

9.2.1.1. *MSS estimates*

Figure 54 presents the first Stokes residuals for Campaign B, with color gradient as an indicator of time evolution (the bluer, the later). Five different clouds of points can be identified. With the eyes only, one can observe that those clouds seem to fit linear functions of MSS, with similar slopes (although different offsets).

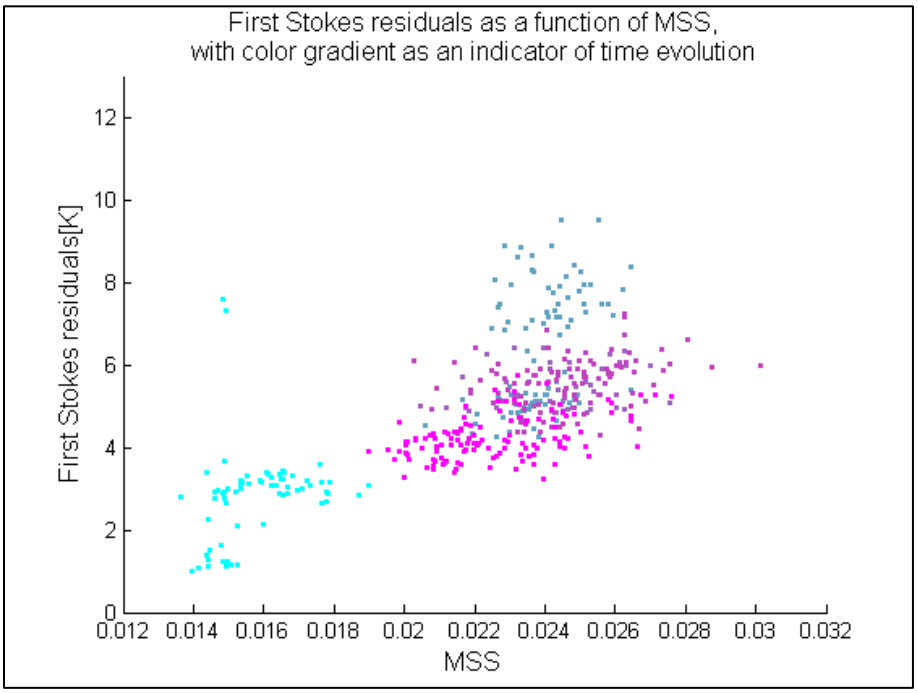


Figure 54: First Stokes residuals as a function of MSS for Campaign B

If the five clouds are individually isolated and fitted with a linear function, they indeed show linear correlations. Identification of each cloud appears to match distinct parts of the flight, where either the heading is different, or the elapsed time "since the previous cloud" is important. Figure 55 and Table 20 present the fitting results, while Figure 56 presents the different parts of the route to which each cloud corresponds. Save

Cloud 5, clouds show similar slopes, all around 110, which bears out the correlation established in Campaign A (see section 9.1.1.1, slope around 105).

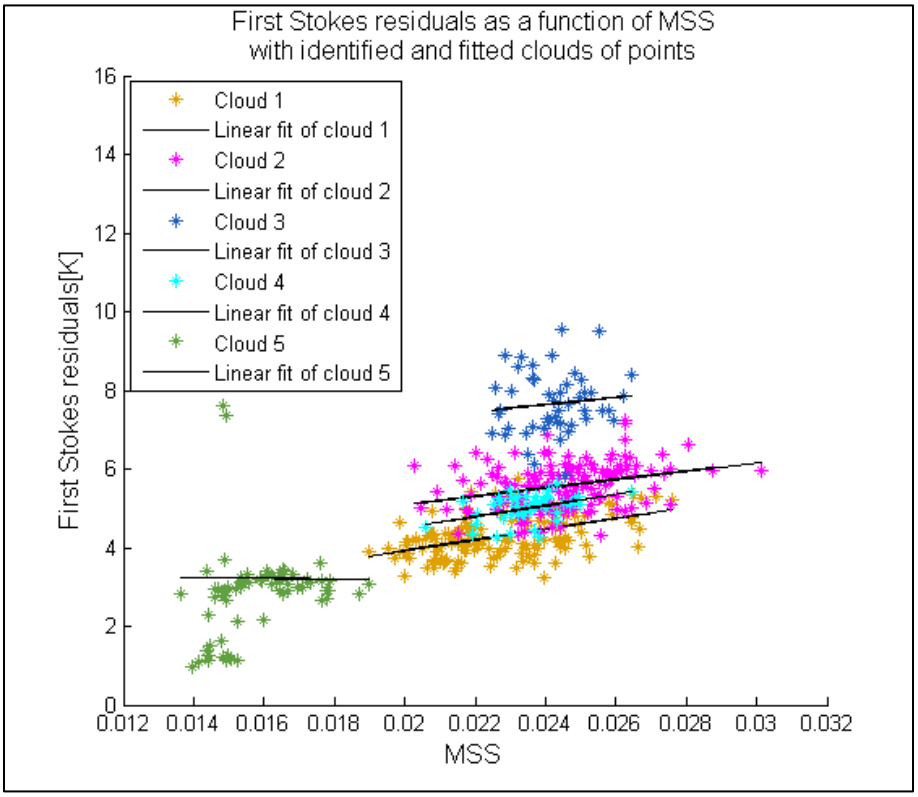


Figure 55: First Stokes residuals as a function of MSS, fitted for each cloud of points

Table 20: Linear fitting information relative to Figure 55

	Linear fit equation: $y_{cloud} = \alpha_{cloud} \cdot MSS + \beta_{cloud}$	**Correlation coefficient r_{cloud}**
Cloud 1	$y_{cloud\ 1} = 135 \cdot MSS + 1.2$	0.4677
Cloud 2	$y_{cloud\ 2} = 106 \cdot MSS + 3.0$	0.3063
Cloud 3	$y_{cloud\ 3} = 90 \cdot MSS + 5.5$	0.1127
Cloud 4	$y_{cloud\ 4} = 138 \cdot MSS + 1.8$	0.4198
Cloud 5	$y_{cloud\ 5} = -13 \cdot MSS + 3.4$	-0.0056

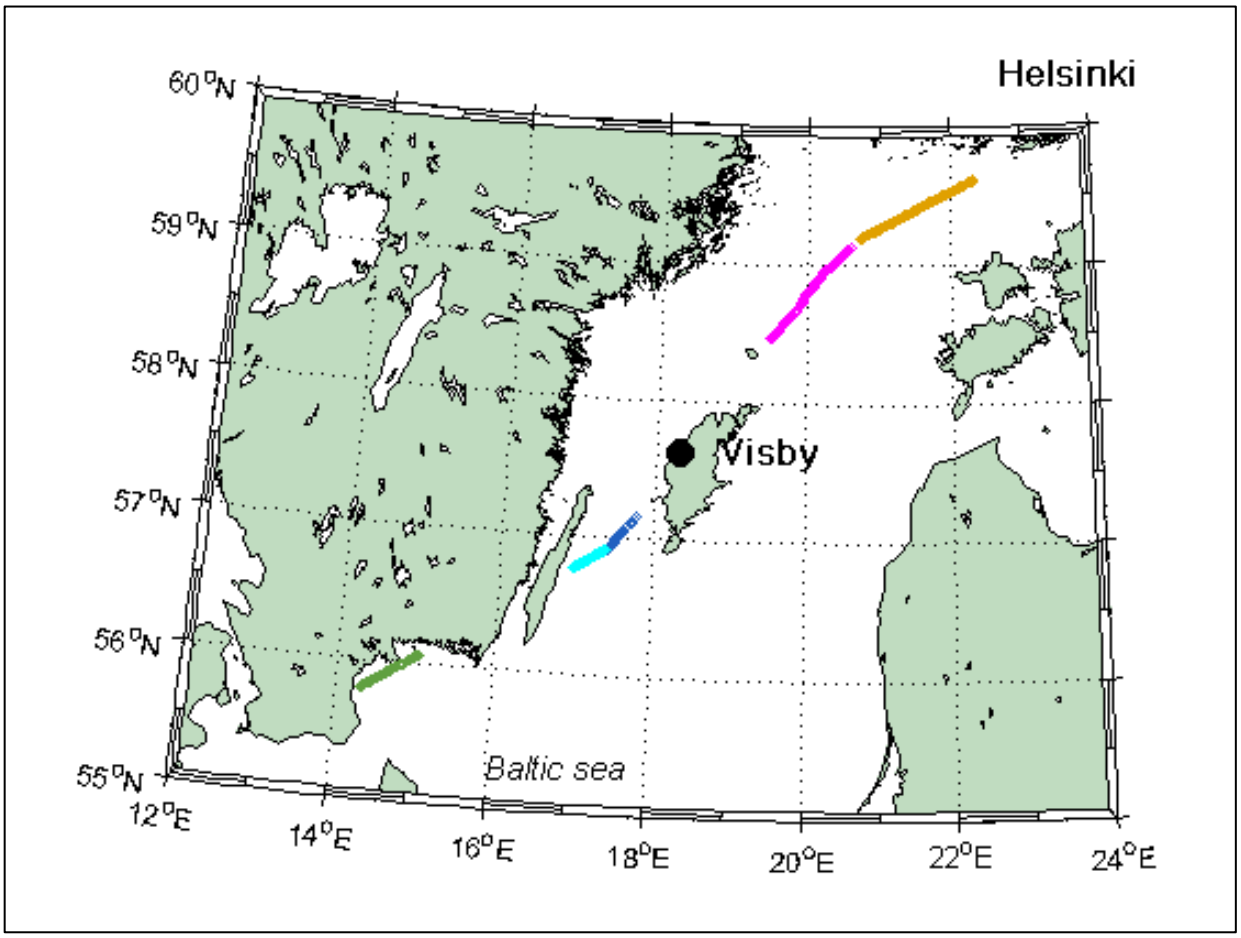


Figure 56: Identification and location of the four clouds of points (the colors do match those of Figure 55)

9.2.1.2. *Delay of the waveform's peak, or PeakDelay*

Given a GPS satellite of certain PRN code, each "cloud" of points of Campaign B does not necessarily contain GOLD-RTR data made with this satellite: the duration of Campaign B is so long (around 4 hours just for the valuable part) that, indeed, no satellite is at high elevation during the whole flight experiment. That is why results from the whole experiment are not presented for one unique PRN code as done in Campaign A: this would correspond to too many graphs (23 to be exact). Graphs were however computed, and it turned out that, given a PRN code, the first Stokes residuals do not always increase with $PeakDelay$ (although they should) for each cloud of points. Rather than selecting the proper PRN codes for each cloud of point (as done in Campaign A), the whole dataset is considered instead, knowing that such a processing of the data brings much more noise and discrepancies in the final results.

Still, in order to illustrate how the first Stokes residuals vary with $PeakDelay$ when considering only one PRN code, Figure 57 presents the first Stokes residuals as a function of $PeakDelay$, where the considered PRN code for each cloud is the most selected one by GOLD-RTR (most measurement points).

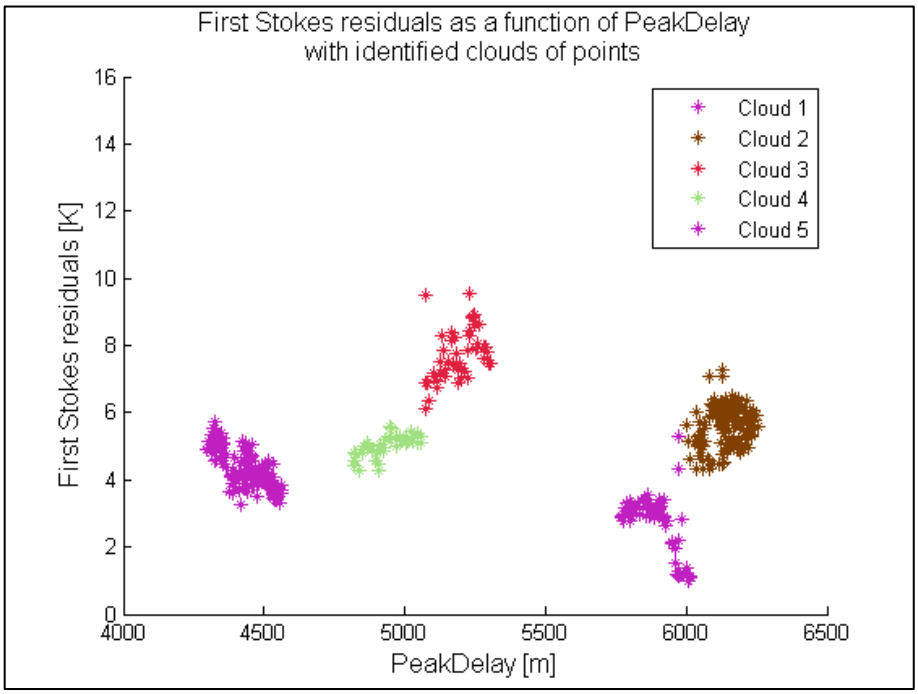


Figure 57: First Stokes residuals as a function of $PeakDelay$

9.2.1.3. *SpecDelay*

Similarly to $PeakDelay$, $SpecDelay$ data do not correlate with the first Stokes residuals. Results of Figure 58 are obtained like $PeakDelay$, that is: measurements are filtered according to the most selected PRN code. The fact that each cloud of points is itself divided into two different clusters is u expected. Moreover, the fact that those clusters are all separated by a similar offset value (around 500m) is very peculiar. It is obvious that the computation of $SpecDelay$ was not properly done during the whole campaign, but those results still remain surprising and unexplained.

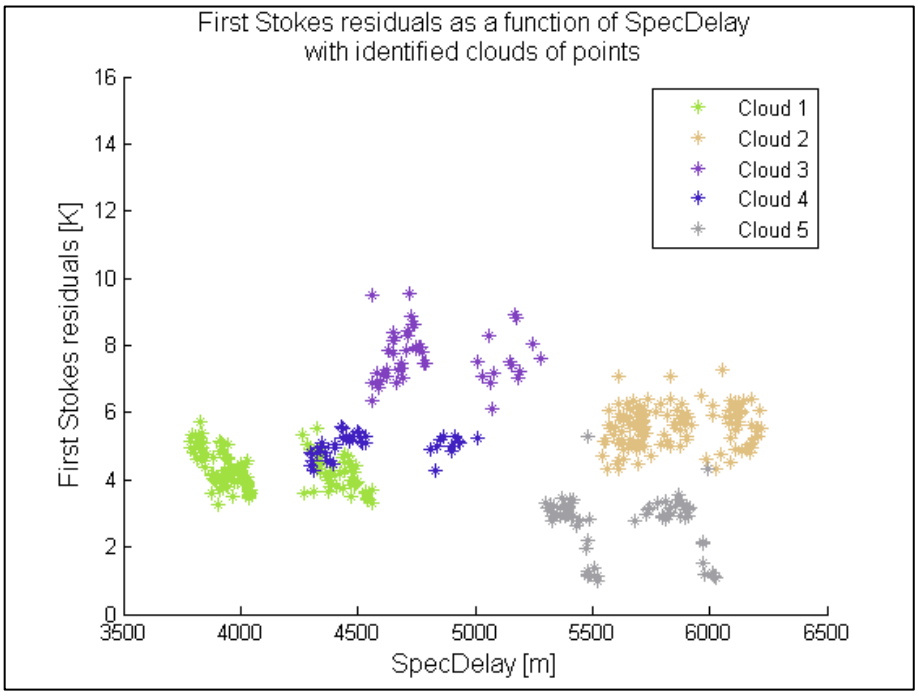


Figure 58: First Stokes residuals as a function of $SpecDelay$

9.2.1.4. *MaxWavDelay*

As in Campaign A, correlation between the first Stokes residuals and $MaxWavDelay$ can be observed in Campaign B, although the linearity is less evident. Correlation coefficients are quite low, but as explained in section 8.2, it is enough to observe with the eyes that correlation is present. What's more is the reliability of the first and fourth clouds of points, for which correlation coefficients are relatively satisfactory (as opposed to the other clouds). Those two clouds also showed the highest coefficients in the fitting of MSS estimates (see Table 20).

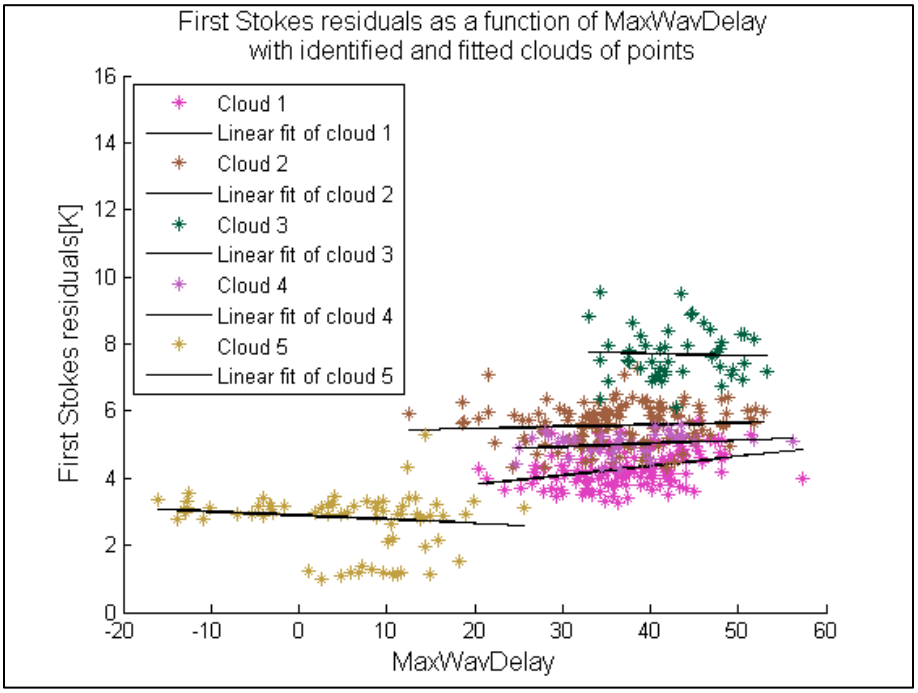


Figure 59: First Stokes residuals as a function of $MaxWavDelay$

Table 21: Linear fitting information relative to Figure 59

	Linear fit equation: $y_{cloud} = \alpha_{cloud} \cdot MaxWavDelay + \beta_{cloud}$	**Correlation coefficient** r_{cloud}
Cloud 1	$y_{day1} = 2.8 \cdot 10^{-2} \cdot MaxWavDelay + 3.2$	0.3342
Cloud 2	$y_{day1} = 0.5 \cdot 10^{-2} \cdot MaxWavDelay + 5.4$	0.0691
Cloud 3	$y_{day1} = -0.5 \cdot 10^{-2} \cdot MaxWavDelay + 7.9$	-0.0336
Cloud 4	$y_{day1} = 1.0 \cdot 10^{-2} \cdot MaxWavDelay + 4.6$	0.2104
Cloud 5	$y_{day1} = -1.1 \cdot 10^{-2} \cdot MaxWavDelay + 2.9$	-0.0819

9.2.1.5. *ScattDelay*

Similarly to Campaign A, several clouds of points are observed in $ScattDelay$ values (one shall pay attention to the fact that, here, the term cloud is not related to the different parts of the flight, but to the range of values that is found for $ScattDelay$): one around 60m, one around 500m, one around 0m with negative values, and a last one around 120m which is clearly noise (scattered and much fewer points). It is

pretty clear that only the first cloud matches realistic values: $ScattDelay$ cannot be as high as 500m, but it cannot be too low (or even negative), and 60m is moreover a very reasonable value. Data were thus filtered so that remaining values fall in the bin [40; 90] meters. The obtained correlation coefficients are surprisingly low, considering the correlation that can be observed with the eyes (see Figure 60), and they are again best for clouds 1 and 4. No correlation is concluded as regards $ScattDelay$ for this campaign because discrepancies accumulate: there is a gap between the average slope found in Campaign A ($22.5 \cdot 10^{-2}$, see section 9.1.1.5) and the present one (around $1 \cdot 10^{-2}$ if only clouds 1 and 4 are considered), the correlation coefficients are too low, despite a heavy filtering of the data.

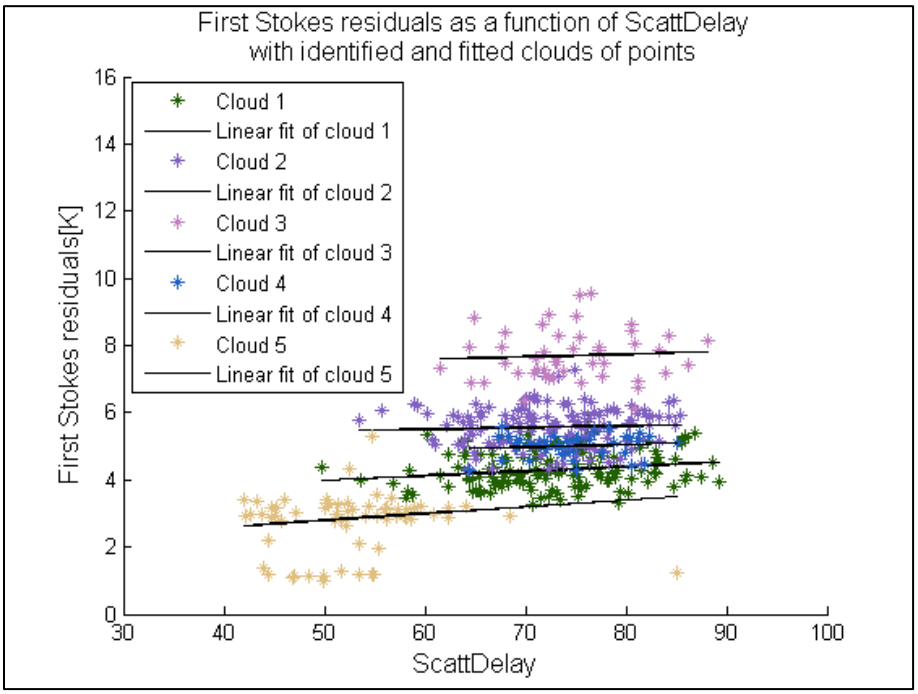


Figure 60: First Stokes residuals as a function of $ScattDelay$

Table 22: Linear fitting information relative to Figure 60

	Linear fit equation: $y_{day} = \alpha_{day} \cdot ScattDelay + \beta_{day}$	**Correlation coefficient r_{cloud}**
Cloud 1	$y_{day1} = 1.4 \cdot 10^{-2} \cdot ScattDelay + 3.3$	0.1933
Cloud 2	$y_{day2} = 0.5 \cdot 10^{-2} \cdot ScattDelay + 5.2$	0.0594
Cloud 3	$y_{day2} = 0.8 \cdot 10^{-2} \cdot ScattDelay + 7.1$	0.0630
Cloud 4	$y_{day2} = 0.7 \cdot 10^{-2} \cdot ScattDelay + 4.5$	0.1085
Cloud 5	$y_{day2} = 2.1 \cdot 10^{-2} \cdot ScattDelay + 1.8$	0.1084

9.2.1.6. *Waveform's tail time length, or TailDelay*

Although correlation coefficients are quite low, a linear dependency of the first Stokes on $WavArea$ seems to be present.

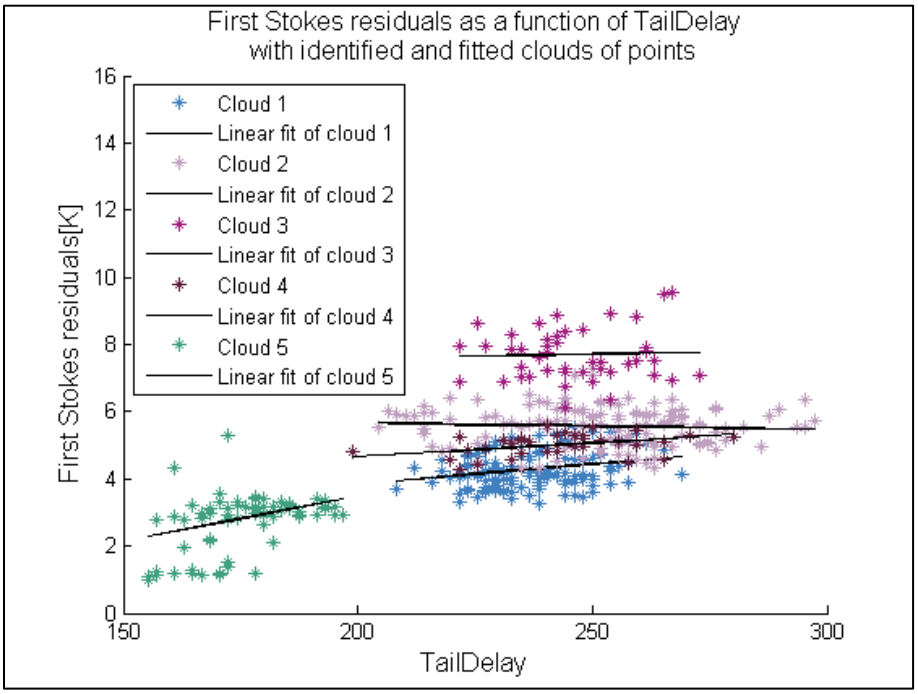


Figure 61: First Stokes residuals as a function of $TailDelay$

Table 23: Linear fitting information relative to Figure 61

	Linear fit equation: $y_{cloud} = \alpha_{cloud} \cdot TailDelay + \beta_{cloud}$	**Correlation coefficient** r_{cloud}
Cloud 1	$y_{day1} = 1.2 \cdot 10^{-2} \cdot TailDelay + 1.5$	0.2514
Cloud 2	$y_{day1} = -0.2 \cdot 10^{-2} \cdot TailDelay + 6.0$	-0.0623
Cloud 3	$y_{day1} = 0.2 \cdot 10^{-2} \cdot TailDelay + 7.1$	0.0384
Cloud 4	$y_{day1} = 0.8 \cdot 10^{-2} \cdot TailDelay + 3.0$	0.3945
Cloud 5	$y_{day1} = 2.8 \cdot 10^{-2} \cdot TailDelay - 2.0$	0.2208

Since no clear correlation could be concluded, data for each cloud were put in bins of width one meter, and the median value of radiometric measurements were considered for each bin. It would be biased and

meaningless to perform a least-square fit of such data, but this method helped to better visualize whether there is a linear dependency or not. As can be seen on Figure 62, the first Stokes residuals seem indeed to depend linearly on $TailDelay$, although variations are so little that it seems like they are constant with $TailDelay$. Nevertheless, residuals are known to vary during the flight experiment, so the very flat profiles of Figure 62 are misleading: there is no reason for the residuals to be constant with $TailDelay$, because there is no reason for them to be constant with any GNSS-R observable. The reason why the sensitivity of the first Stokes to $TailDelay$ is much lower than what was established for Campaign A (see 9.1.1.6) is an open question.

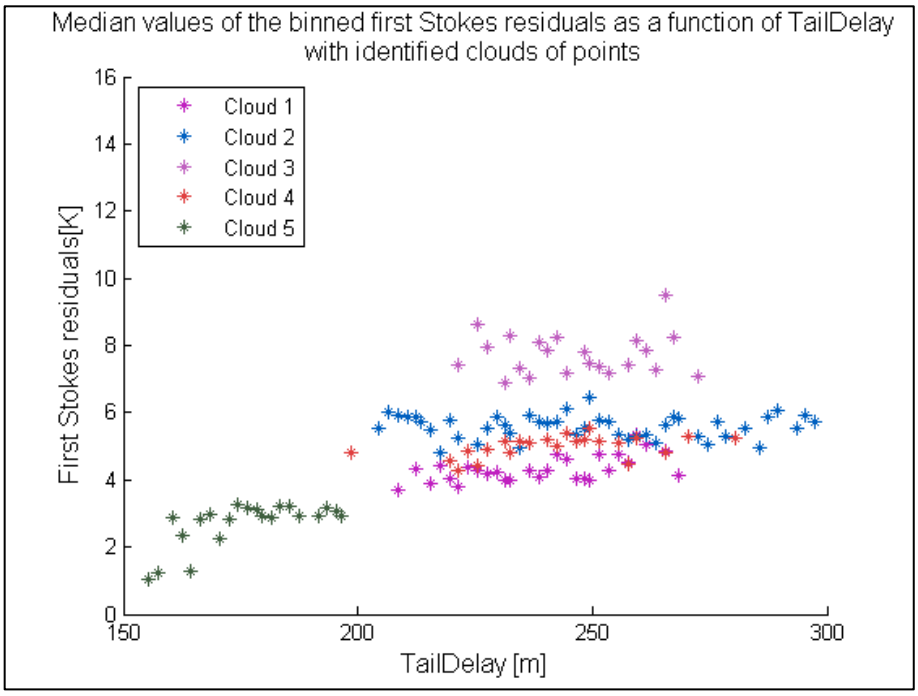


Figure 62: First Stokes residuals (median values of binned data) as a function of $TailDelay$

9.2.1.7. *Area under the waveform, or WavArea*

Similarly to $TailDelay$, a linear dependency of the first Stokes on $WavArea$ seems to be present, although correlation coefficients are quite low.

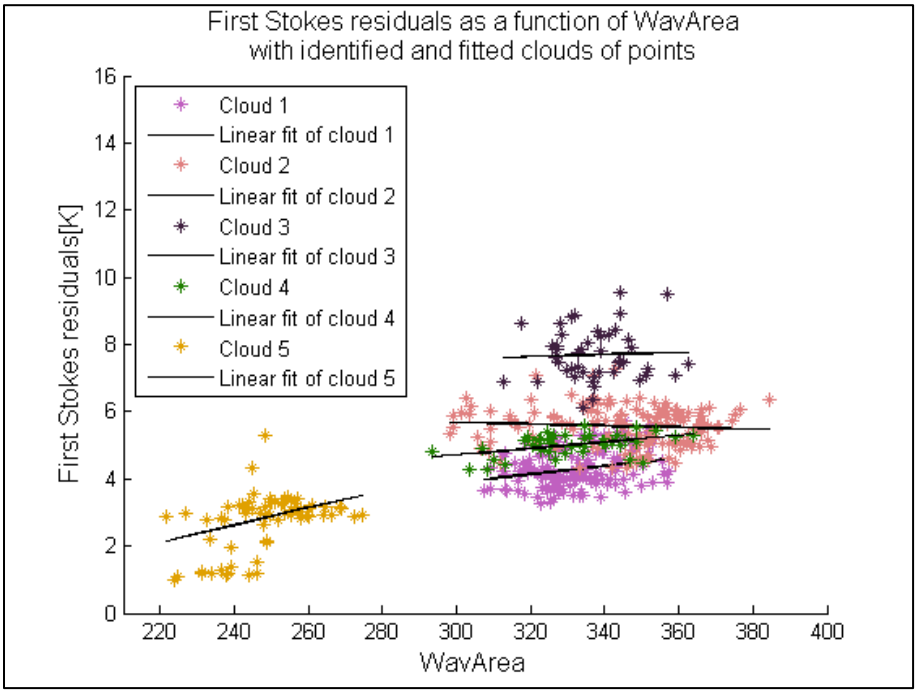


Figure 63: First Stokes residuals as a function of $WavArea$

Table 24: Linear fitting information relative to Figure 51

	Linear fit equation: $y_{cloud} = \alpha_{cloud} \cdot WavArea + \beta_{cloud}$	**Correlation coefficient** r_{cloud}
Cloud 1	$y_{day1} = 1 \cdot 10^{-2} \cdot WavArea + 0.2$	0.2594
Cloud 2	$y_{day1} = -0.2 \cdot 10^{-2} \cdot WavArea + 6.4$	-0.0797
Cloud 3	$y_{day1} = 0.3 \cdot 10^{-2} \cdot WavArea + 6.7$	0.0408
Cloud 4	$y_{day1} = 0.9 \cdot 10^{-2} \cdot WavArea + 1.9$	0.4493
Cloud 5	$y_{day1} = 2.6 \cdot 10^{-2} \cdot WavArea - 3.6$	0.2256

As with $TailDelay$, data were binned, and the median value of the residuals in each bin was considered. Figure 63 presents the corresponding results. It is then easier to ascertain that the two variables are correlated.

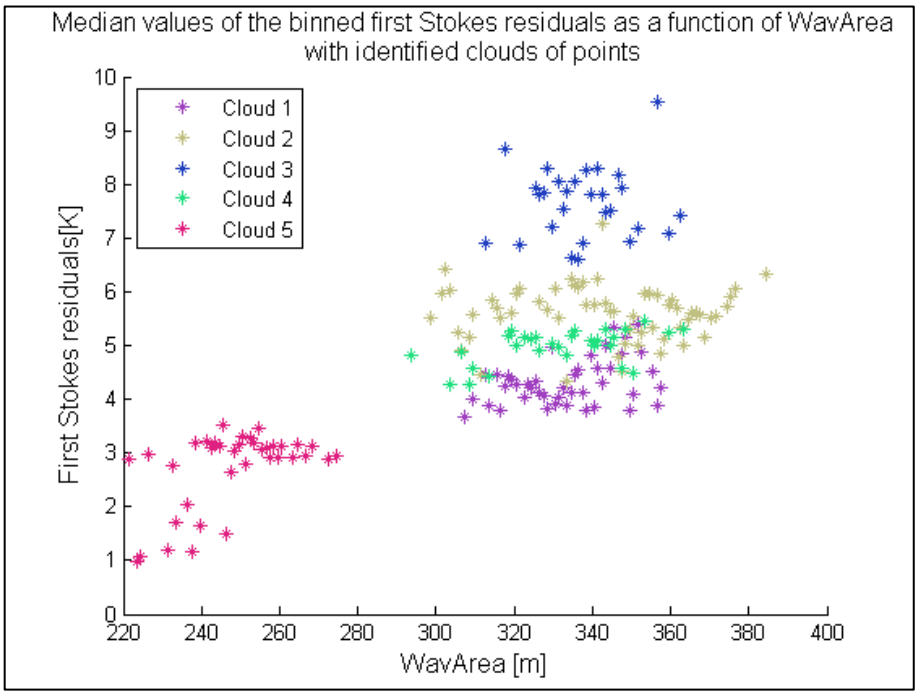


Figure 64: First Stokes residuals (median values of binned data) as a function of $WavArea$

10. Conclusion

Among the seven GPS-R observables that were investigated in the frame of the thesis, five showed linear correlation with radiometric measurements for both datasets considered (Campaign A and Campaign B). Those are MSS estimates $MaxWavDelay$, $ScattDelay$, $TailDelay$, and $WavArea$. Table 25 is a summary of the obtained results: average slopes for Campaign B were computed from the clouds of points number 1 and 4 only, since most correlation coefficients were satisfactory for those only.

$ScattDelay$, $TailDelay$ and $WavArea$ are considered more reliable as compared to $MaxWavDelay$, because they are not defined with respect to the estimated specular delay, computed by GOLD-RTR from a simplified model, but with respect to the actual real specular point. All three are a measure of how much wider the glistening area becomes with increasing roughness: $ScattDelay$ describes how much the peak of the waveform is delayed as compared to the expected time of reception, $TailDelay$ describes how flat the trailing edge of the waveform is, and $WavArea$ describes how large the contribution from the glistening area is.

MSS estimates, on the other hand, are the combination of GPS-R information with a statistical description of the waves. As such, the computation of MSS needs assumptions to be made about the sea state, which is less straightforward and error-proof than considering measured parameters only. Nonetheless, correlation obtained with MSS is the most evident (highest correlation coefficients), and also the most reliable

since slopes of the fitting functions are very consistent. The average slope found for MSS changes by 31 units from Campaign A and B, while it changes by only $6.5 \cdot 10^{-2}$ for $TailDelay$, but MSS and $TailDelay$ do not have the same order of magnitude: MSS values are typically on the order of $2 \cdot 10^{-2}$, while $TailDelay$ is on the order of $2 \cdot 10^{+2}$. Instead, one shall compare relative change of the slope: $\left|\frac{slope_{Campaign\ B} - slope_{Campaign\ A}}{slope_{Campaign\ A}}\right|$. The relative change is about 23% in the case of MSS, while it is about 87% in the case of $TailDelay$.

Table 25: Summary of investigated correlations of the first Stokes with the different GOLD-RTR observables

Campaign A: 2007 Gulf of Finland / B: 2008 Baltic Sea GOLD-RTR observable	Average slope	Average slope
MSS	105	136
PeakDelay		
SpecDelay		
MaxWavDelay	$13.9 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$
ScattDelay	$22.5 \cdot 10^{-2}$	
TailDelay	$7.5 \cdot 10^{-2}$	$1 \cdot 10^{-2}$
WavArea	$4.7 \cdot 10^{-2}$	$1 \cdot 10^{-2}$

What's more is the quality of those roughness indicators. One can indeed notice that the range of MSS estimates is very different from one day to the other in Campaign A, whereas it is not as clearly the case with the other GNSS-observables. However, sea state evolved quite much between the two days: sea surface was rougher and wind was stronger on the second day ([1]), so the range of roughness indicators is expected to be wider as compared to the first day. That is why MSS estimates seem to better describe how rough the sea surface is.

To conclude, this thesis demonstrated that: first, roughness of the sea is better characterized if GNSS-R measurements are combined with statistical information; secondly, utilization of MSS estimates to correct the model of water emissivity at L-band is most likely the best option. Nonetheless, other GNSS-R observables are worth considering and further work about how they are correlated with the thermal emission of the sea needs to be done. Results that were previously presented by the scientific community about MSS data were

reproduced in the thesis, and sometimes even enhanced. Likewise, data relative to the other GOLD-RTR observables shall be the subject of other investigations. Last, an airborne campaign was organized during the thesis, and the collected data shall be looked into. An enhanced version of the GPS reflectometer was used then, and even though it was under test, the precision of its observables is expected to be higher than with the former version of the instrument.

List of references and sources

[1] J. Kainulainen, K. Rautiainen, J. Lemmetyinen, M. Hallikainen, F. Martin-Porqueras, M. Martin-Neira, "Detection of a sea surface salinity gradient using datasets of airborne synthetic aperture radiometer HUT-2-D," *IEEE Trans. Geosci. Remote Sens.*, accepted for inclusion in a future issue of the journal.

[2] J. F. Marchan-Hernandez, N. Rodriguez-Alvarez, A. Camps, X. Bosch-Lluis, I. Ramon-Perez, E. Valencia, "Correction of the Sea State impact in the L-band brightness temperature by means of Doppler-Delay Maps of Global Navigation Satellite Signals reflected over the sea surface," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2914-2923, October, 2008.

[3] J. Rotbøll, S.S. Sjøbærg, N. Skou, "A novel L-band polarimetric radiometer featuring subharmonic sampling, " *IEEE Trans. Geosci. Remote Sens*., vol. 46, no. 3, pp. 621-645, Mar. 2008.

[4] O. Nogués-Correig, E. Cardellach, J. Sanz, and A. Rius, "A GPS-reflections receiver that computes doppler-delay maps in real time," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pp. 156-174, Jan. 2007.

[5] K. Rautiainen, J. Kainulainen, T. Auer, J. Pihlflyckt, J. Kettunen, M. Hallikainen, "Helsinki University of Technology L-band airborne synthetic aperture radiometer, " *IEEE Trans. Geosci. Remote Sens.*, vol. 46, pp. 717 – 726, Mar. 2008.

[6] L. A. Klein and C. T. Swift, "An improved model for the dielectric constant of sea water at microwave frequencies," *IEEE Trans. Antennas Propag.*, vol. AP-25, pp. 104-111, Jan. 1977.

[7] A. Camps, J. Font, M. Vall-Ilossera, C. Cabarro, I. Corbella, N. Duffo, F. Torres, S. Blanch, A. Aguasca, R. Villarino, L. Enrique, J. J. Miranda, J. J. Arenas, A. Julia, J. Etcheto, V. Caselles, A. Weill, J. Boutin, S. Contardo, R. Niclos, R. Rivas, S. C. Reising, P. Wursteisen, M. Berger, M. Martin-Neira, "The WISE 2000 and 2001 field experiments in support of the SMOS mission: Sea surface L-band brightness temperature observations and their application to sea surface salinity retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 804–823, Apr. 2004.

[8] C. Gabarro, J. Font, J. Miller, A. Camps, D. Burrage, J. Wesson, A. R. Piola, "The use of a semi-empirical emissivity model for a rough estimation of sea surface salinity from an airborne microwave radiometer," *Sci. Mar*., vol. 72, no. 2, pp. 329–336, Jun. 2008.

[9] E. Cardellach, "Sea surface determination using GNSS reflected signals," Ph.D. dissertation, Univ. Politecnica Catalunya, Catalonia, Spain, 2001.

[10] E. Valencia, A. Camps, X. Bosch-Lluis, N. Rodriguez-Alvarez, I. Ramos-Perez, and J. F. Marchan-Hernandez, "Brightness temperature correction of the sea state effect using GNSS-R data," in *Proc. MicroRad, Washington, DC, 2010*, pp. 129–133.

[11] A. Rius, J. M. Aparicio, E. Cardellach, M. Martín-Neira, and B. Chapron, “Sea surface state measured using GPS reflected signals,” *Geophys. Res. Lett.*, vol. 29, no. 23, p. 2122, Dec. 2002.

[12] Ferre-Lillo, P., Rodriguez-Alvarez, N., Bosch-Lluis, X., Valencia, E., Marchan-Hernandez, J.F., Camps, I., "Delay-Doppler Maps study over ocean, land and ice from space", *Geosci. Remote Sens. Symposium,2009 IEEE International, IGARSS 2009,* vol. 2, pp. II-722-725, July 2009

[13] E. Valencia , J. F. Marchan-Hernandez , A. Camps , N. Rodriguez-Alvarez , J. M. Tarongi , M. Piles , I. Ramos-Perez , X. Bosch-Lluis , M. Vall-llossera and P. Ferr , "Experimental relationship between the sea brightness temperature and the GNSS-R delay-Doppler maps: Preliminary results of the ALBATROSS field experiments", *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, p.III-741, July2009.

[14] E. Cardellach, A. Rius, “PARIS in CoSMOS-OS 2007: Scatterometric part”, Final report, IEEC, Ref. FR-S-CoSMOS-07, June 2008

[15] E.Valencia, A. Camps, N. Rodriguez-Alvarez, I. Ramos-Perez, X. Bosch-Lluis, H. Park, “Improving the accuracy of sea surface salinity retrieval using GNSS-R data to correct the sea state effect”, *Radio Sci.,* 46, RS0C02, doi:10.1029/2011RS004688, 2011

[16] Marchan-Hernandez J. F., Valencia E., Rodriguez-Alvarez N., Ramos-Perez I., Bosch-Lluis X., Camps A., Eugenio F., Marcello J., "Sea state determination using GNSS-R data", *IEEE Geosci. Remote Sens. Lett.*, vol. 7, 2010

[17] Valencia, E., Camps A., Bosch-Lluis X., Rodriguez-Alvarez N., Ramos-Perez I., Marchan-Hernandez J., “Brightness temperature correction of the sea state effect using GNSS-R data”, Proceedings of the Microwave and Remote Sensing of the Environment, MICRORAD 2010, pp. 129-133, IEEE, Washington D.C.

[18] A. Rius , E. Cardellach and M. Martin-Neira "Altimetric analysis of sea surface GPS reflected signals", *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp.2119 - 2127 , 2010.