(OTSU) Space Science Earth Observation / (COS) Operations and Design of Space Systems - M2 SOAC Climate Dynamics

Master Project

Waves in the ionosphere detected using the Polar Research Ionospheric Doppler Experiment (PRIDE)

Author :
Ms Cécily NOAILLAC

Supervisors :
Assoc. Prof. Lisa BADDELEY
Dr. Mikko SYRJÄSUO

Version 0.1 of
February 22, 2023

Contents

List of Figures

List of initials, acronyms and abbreviations

PRIDE	*Polar Research Ionospheric Doppler Experiment*
AGWs	*Atmospheric Gravity Waves*
ULF	*Ultra Low Frequency*
MHD	*Magneto Hydro Dynamic*
HF	*High Frequency*
FFT	*Fast Fourier Transform*
STFT	*Short Time Fourier Transform*
AWGN	*Additive White Gaussian Noise*
KHO	*Kjell Henriksen Observatory*
FIR	*Finite Impulse Response (FIR) filter is a filter whose impulse response is of finite duration, because it settles to zero in finite time*

Acknowledgement

I would like to thank my two supervisors, Lisa Baddeley (Main Supervisor) and Mikko Syrjäsuo (Head Engineer), for their help and precious advices during the last six months.

I would like to thank Raphaël, who has been the most amazing support during this internship, as a perfect hiking, skiing, sailing partner.

And thank you, Svalbard, for all the amazing opportunities you are offering me every single day. Living up there is a pleasure, more than ever.

Chapter 1

Doppler Sounding and instrumentation

1.1 The Polar Ionosphere

The ionosphere is a shell of electrons and electrically charged atoms and molecules that surrounds the Earth. Along with the neutral upper atmosphere, the ionosphere forms the boundary between Earth's lower atmosphere and the vacuum of space. Particles in the Earth's atmosphere are ionized by both Solar radiation and the solar wind (a stream of charged particles released from the upper atmosphere of the Sun). Due to this, the ionosphere changes from Earth's day side to night side. It expands from 60km to 800km above sea level.
The ionosphere has an approximately layered structures with three main regions, called the D layer (from 60 to 90km), the E layer (from 90 to 120 km), and the F layer (from 120 to 800km). Unlike the other regions in the atmosphere (like troposphere and mesosphere), these layers do not have a sharp boundaries, and the altitudes at which they occur vary. Figure 1.1 highlights the position of each different layer.

1.2 Doppler sounding

1.2.1 General principle

The Doppler effect, which is the change in frequency of a wave in relation to an observer who is moving relative to the wave source, is used in some types of radar to proceed their measurements. As for a regular speed control radar, the Doppler sounding technique focuses at a change in frequency in the received signal. Doppler sounding is therefore a relatively simple method to investigate dynamics of the ionosphere on short time scales with a time resolution of several seconds. It is based on the transmission of a stable sine

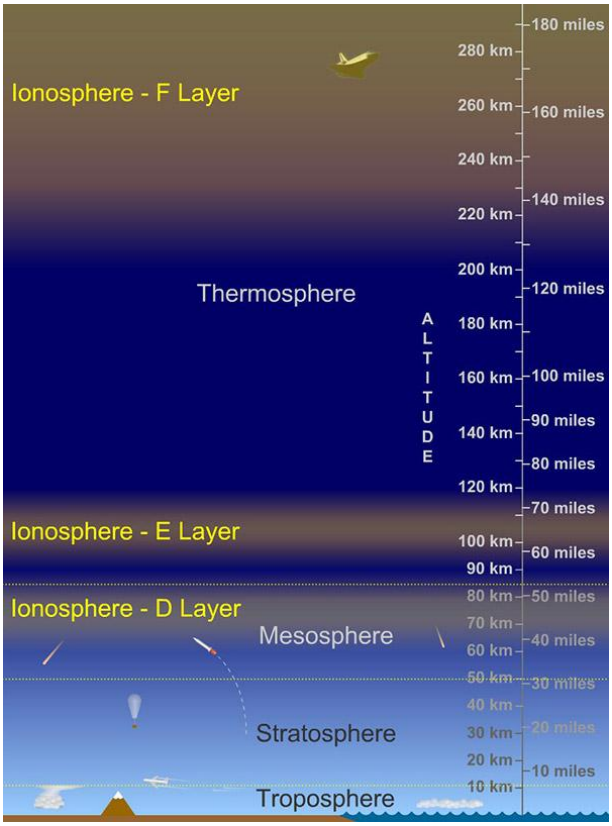

Figure 1.1: Layers in the ionosphere and their position. Credits: UCAR/Randy Russell

wave at a frequency of several MHz that reflects from the ionosphere and goes back toward the ground, where a receiver is placed.
The motion of the ionospheric layers changes the altitude of reflection. Indeed, if the reflection region moves up and down, or if there are plasma density changes along the sounding radio wave, the length of the phase path of the radio wave varies and the frequency of the radio wave shifts owing to the Doppler effect. This shift can then be measured once the signal is received. The frequency variation of the radio wave at the receiving point, Δf , is given by the following equation (2):

$$\Delta f = -\frac{f}{c}\frac{d}{dt}\int_0^l n \, \mathrm{d}l \tag{1.1}$$

from Davies et al. 1962 (5) and Jacobs and Watanabe 1966 (6)
where :

- f is the frequency of the transmitted radio wave
- c is the speed of light
- n is the refractive index

- l is the distance between the transmitting and receiving points

From this equation, it can be stressed that two factors can modify the frequency shift:

- the temporal variation of l, which represents the vertical motion of the reflection point
- the temporal variation of n, which represents the change of the refractive index in the propagation path of the radio wave.

Temporal Variation of l

Considering that $\Delta h << h$, (where h is the altitude of the reflection point) and with h at the midpoint between the transmitter and receiver and assuming that the refractive index does not vary along the propagation path, the variation of the phase path of the sent signal corresponds to the variation of the propagation path. It can be deduced that the variation of the phase path, Δn, is

$$\Delta n = 2\Delta h cos(\theta) \tag{1.2}$$

with θ the incident angle of the radio wave to the ionosphere (5).
Then, the Doppler frequency Δf appears as:

$$\Delta f = -2\frac{f cos(\theta)}{c}\frac{dh}{dt} = -2\frac{f cos(\theta)}{c}v_h \tag{1.3}$$

and the vertical motion of ionospheric plasma, v_h, can therefore be expressed.

Temporal Variation of n

Assuming that the earth's magnetic field and the collision of particles can be neglected, the reflective index n can be expressed as the following equation:

$$n = \sqrt{1 - \frac{f_p^2}{f^2}} = \sqrt{1 - \frac{e^2 N}{4\pi^2 \epsilon_0 m f^2}} \tag{1.4}$$

where:

- $f_p = \sqrt{\frac{e^2 N}{4\pi^2 \epsilon_0 m}}$ is the plasma frequency
- ϵ_0 is the permittivity of the vacuum
- e is the elementary charge
- N is the electron density

- m is the mass of an electron

In this case, the variation of the electron density N introduces a variation of the refractive index, and changes the frequency variation of the radiowave i.e. the Doppler frequency. In this case, the variation of the phase path is given by :

$$\frac{d}{dt}\int_0^l n\,\mathrm{d}l = -\frac{e^2}{4\pi^2\epsilon_0 m f^2}\int_0^h \frac{1}{n}\frac{\partial N}{\partial t}\,\mathrm{d}h \tag{1.5}$$

Finally, the frequency variation of the radio wave is expressed as (5):

$$\Delta f = \frac{e^2}{4\pi^2\epsilon_0 m c f}\int_0^h \frac{1}{n}\frac{\partial N}{\partial t}\,\mathrm{d}h \tag{1.6}$$

In this thesis, we follow the work of Wright et al. and assume that the changes in Doppler frequency are due to changes in the reflection height and thus the path length, l.

There are multiple causes for these variations, from compression waves to radial motion of the reflection region (advective movement). The advection term dominates whether the plasma motion is caused by atmospheric gravity waves (GWs) or by ExB drift, associated with, e.g., magneto-hydrodynamic waves (7)). This two different types of waves will be introduced and developed in Section 1.2.2 below.

1.2.2 Waves in the ionosphere

The PRIDE experiment aims to focus on two major processes:

Atmospheric Gravity Waves (AGWs)

A gravity wave is a vertical wave that can occur between two stable layers of fluids of different density. When the fluid boundary is disturbed, buoyancy forces try to restore the equilibrium. The fluid returns to its original shape. Therefore, overshoots and oscillations set in and propagate as waves.
Some examples of AGWs propagation can be find in figures 1.2 and 1.3.

AGWs transport energy and momentum from high to low latitude, and from the lower atmosphere to the upper atmosphere. A broad spectrum of AGWs can be generated by numerous lower atmospheric processes. For instance, they may originate from disturbances that takes place in the troposphere, or from wind flow over mountain ranges and violent thunderstorms. An initial small amplitude at the tropopause increases with height until the waves break in the mesosphere and lower thermosphere. The momentum of the air imparted by the trigger mechanism forces the parcel of air to rise and the atmospheric

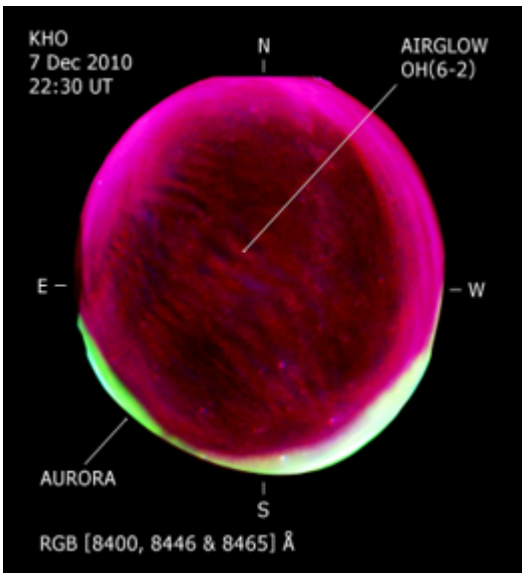

Figure 1.2: ASC image of gravity waves observed in airglow emissions at the KHO, Svalbard

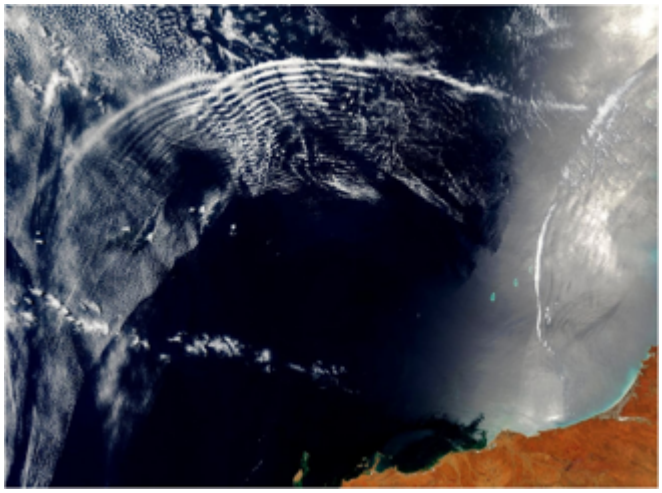

Figure 1.3: Emergence of atmospheric gravity waves over the Indian Ocean via the Moderate Resolution Imaging Spectroradiometer housed in NASA's Terra satellite—emerge. Image credit: Jacques Descloitres, MODIS Rapid Response Team, NASA/GSFC.

stability will induce it to sink again, creating the vertical oscillating motion. At high latitudes, a primary source of AGW might be the auroral processes (3). Figure 1.4 explains the origin and dynamics of planetary and gravity waves:

It can be highlighted from this figure that a portion of the AGW spectrum can break in the mesosphere, depositing its energy and momentum in that region, and leading to the formation of the cold summer mesopause. AGWs can be obserevd in airglow emissions in the mesosphere, figure 1.2. Other parts of the wave spectrum can propagate into the thermosphere where they play important roles in thermospheric and ionospheric dynamics (3). These AGWs are believed to serve as sources for the initial electron density perturbations required by mid latitude and low latitude plasma instabilities producing spread F events (3). Their wavelengths can range up to thousands of kilometers. Their periods range from a few minutes to days.

Ultra Low Frequency (ULF) Magneto hydrodynamic (MHD) waves

Ultra Low Frequency (ULF) waves are the lowest frequency plasma waves in the Earth's magnetosphere, with frequencies from 0.001 to 5 Hz (11). At the lower end of the ULF band, waves are often well described using a magnetohydrodynamic (MHD) approximation (10). The set of equations that describe MHD are a combination of the Navier–Stokes

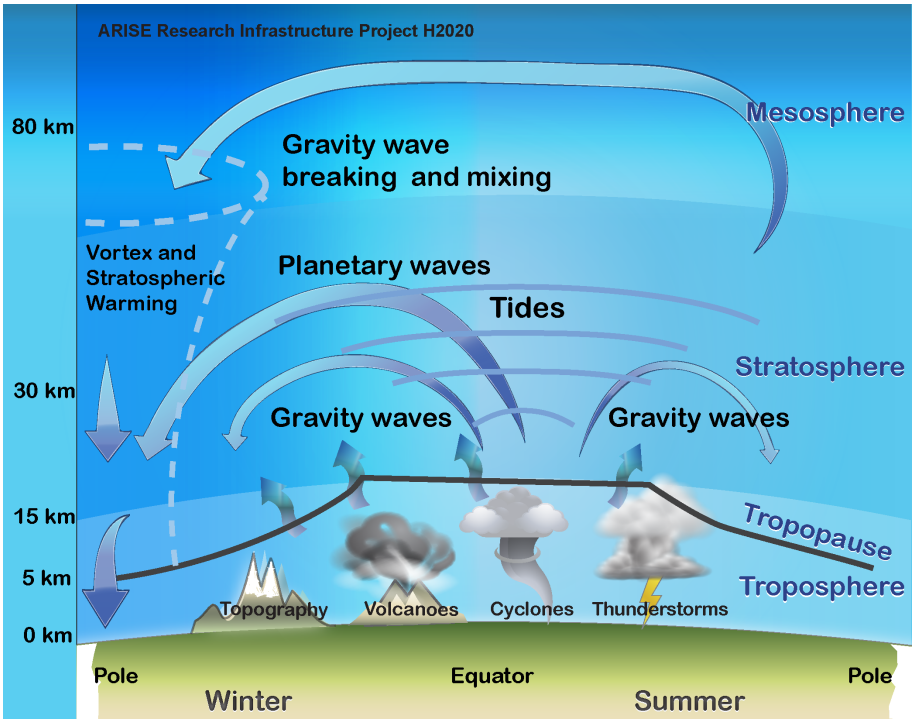

Figure 1.4: Dynamics of the troposphere-stratosphere-mesosphere exchanges including contribution of gravity waves and planetary waves. *ARISE: Atmospheric Dynamic Research InfraStructure in Europe*

equations of fluid dynamics and Maxwell's equations of electromagnetism. The wave modes derived using MHD plasma theory are called magnetohydrodynamic waves or MHD waves. MHD is thus considered the marriage of hydrodynamics to electromagnetism. The first recorded use of the word magnetohydrodynamics is by Hannes Alfvén in 1942:
"At last some remarks are made about the transfer of momentum from the Sun to the planets, which is fundamental to the theory (§11). The importance of the Magnetohydrodynamic waves in this respect are pointed out."
The pulsation frequency is considered to be "ultra" low when it is lower than the natural frequencies of the plasma, like plasma frequency and the ion gyrofrequency. ULF waves play important roles in magnetosphere-ionosphere coupling, ring current/radiation belt dynamics, geomagnetically induced currents, substorms, and other areas relevant to space weather prediction (10). ULF waves can carry significant energy to the ionosphere and play important roles in the magnetosphere-ionosphere coupling. They can cause modulation and enhancement of several ionospheric parameters and provide ion frictional heating in the ionosphere-thermosphere system. ULF waves modulate the Earth's magnetic field, and this modulation also modulates the ExB drift in the ionsophere. Then, this shift in the ExB drift modulates the reflection point of the signal.
ULF waves can be classified according to the period of the pulsation (Jacobs et al., 1964).

Table 1.5 represents the pulsation classes for the ULF MHD waves. Recent studies have

Pulsation classes							
	Continuous pulsations					Irregular pulsations	
	Pc 1	Pc 2	Pc 3	Pc 4	Pc 5	Pi 1	Pi 2
T [s]	0.2-5	5-10	10-45	45-150	150-600	1-40	40-150
f	0.2-5 Hz	0.1-0.2 Hz	22-100 mHz	7-22 mHz	2-7 mHz	0.025-1 Hz	2-25 mHz

Figure 1.5: Typical classification scheme for the ULF waves, according to the period of the pulsation (Jacobs et al., 1964)

shown that ULF wave-related precipitation of energetic electrons can affect ionospheric conductivities and modulate Hall and Pedersen conductances. These large conductivity modulations affect magnetosphere-ionosphere coupling processes. It has been reported that ionosphere-thermosphere heating rates reported total electron content variations related to Pc4 and Pc5-6 ULF waves (10).

1.3 Doppler Radar Experiments

1.3.1 Mid - and Auroral Latitude Experiments

Doppler sounder measurements have been successfully conducted in different countries at lower latitude and at different longitudes.

Tromsø, Norway

The DOPE (DOppler Pulsation Experiment) HF Sounder, located near Tromsø, Norway (Corrected GeoMagnetic (CGM) lat: approximately 66°N, lon: approximately 105°E was operational from May 1995 until the early 2000s and is designed to make measurements of the Doppler ionospheric signatures of ULF waves at high latitudes . DOPE has three azimuthally-separated propagation paths and has been used to provide the first statistical examination of small scale-sized waves in the ionosphere (4). During this study, the signal received from DOPE was compared to co-located ground magnetometer data. This was done in order to see if the signals had a correlating signature in the ground magnetometer data, i.e. if the Doppler signal is large enough in term of horizontal scale size to be observed by ground magnetometers or not. Some small scale ULF waves were observed, as figure A.2 in appendix stresses, and this event has been considered "uncorrelated": The event is so small in horizontal scale size that it is 'invisible' to ground magnetometers (the X and Y components from 3 ground magnetometer stations located beneath the ionospheric

reflection point of the doppler trace are shown in the bottom 6 panels).
However, some Large Scale ULF Waves events have been found to have a ground magnetometer signature (so called 'correlated' events), as depicted in figure A.2 in appendix: An FFT of the Doppler and magnetometer traces are shown to the right - showing the wave had a periodicity of approximately 8mHz (125 seconds).

Chesapeake Bay, Florida, USA

A Doppler Sounder has been successfully deployed in Chesapeake Bay, Florida, in order to study the AGW in the ionosphere. Figure A.4 in appendix is an example of an AGW trace from this instrument. In this case there are 3 Doppler paths (shown by the different colour traces) and the received signal has been split into an O-mode and X-mode trace. The AGWs have much longer periods, of the order of 30 minutes, as can be seen from the FFT analysis shown in the bottom plot, figure A.5 in appendix

Multiple locations, Japan

Since 2003, an HF Doppler sounding experiment has been conducted in Japan to study atmospheric and ionospheric processes. The sounder uses a new digital signal processing technique. The data and processing code for the system are available online `http://gwave.cei.uec.ac.jp/%7Ehfd/index.html`. The processing code and raw data from the Japanese experiment has been used as a baseline with which to test the code developed for PRIDE. The data from the Japanese experiment has been processed using the code provided from the website and also the code developed for PRIDE.
The language of their code is Python, and in order to create a 24 hours Doppler shift plot, the algorithm requires three days of data: the day of interest, the day before and the day after. This allows a correct windowing at any time of the day of interest (used for the first and last samples of the day), but creates a relatively slow algorithm, as three time the amount of data is required. The figure 1.6 below represents a one day Spectrogram and Doppler shift processed with the Japanese code on python.

1.3.2 Observations in the high latitude regions

The cusp is the region where the field lines from the Earth's magnetic field become open to those with the Interplanetary magnetic Field through magnetic reconnection. This is an important site of interaction between the solar wind and the Earth's upper atmosphere. The location of the cusp is fixed with respect to the sun, and is therefore located on the dayside of the Earth, high above the arctic circle. The Earth rotates beneath the cusp location and Svalbard is placed under the cusp region for few hours per day. It is therefore a strategical place to study upper polar atmosphere physics and electromagnetic waves and

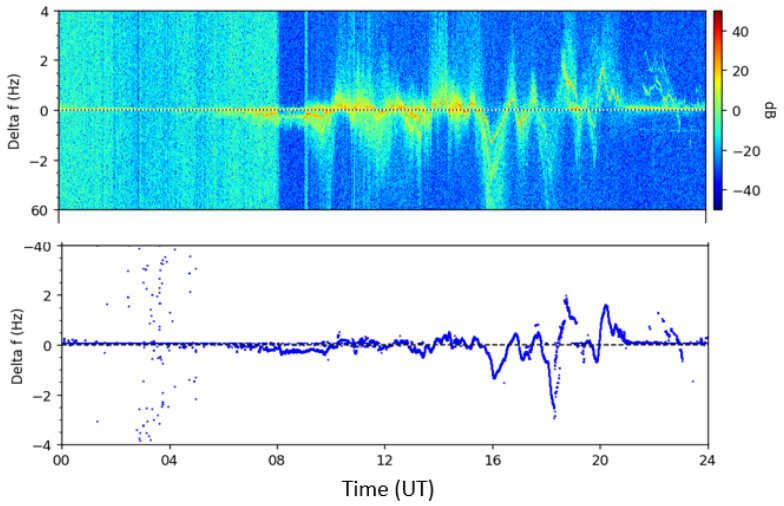

Figure 1.6: One day Spectrogram on 14-04-2022, processed with Python using the Japanese code written by Pr. Keisuke Hosokawa.

structures in the ionosphere.
While the high latitude location of this facility is particularly aimed at studies of the cusp and polar cap region, a Doppler radar of this kind has not been successfully deployed on Svalbard before. The dataset will therefore be entirely new and probe a new region of the polar atmosphere.

The topology of the magnetic lines around Svalbard may lead to difficulties in detecting small scale ULF MHD waves with a Doppler Radar. Figure A.1 in appendix illustrates the magnetic field, the currents governing the exchanges of energy in the ionosphere (Pederson and Hall), and a drift V_{ExB} component. This V_{ExB} drift has two components: a, smaller, vertical and a, larger, horizontal one. The ULF waves modulate this ExB drift and the Doppler sounder can detect the vertical component. But the higher in latitude, the more vertical the Earth's magnetic field lines are. Figure 1.7 illustrates the very high inclination of the lines on Svalbard. This can also be seen looking at the EISCAT radars in Breinosa (Longyearbyen). One of the two dishes that composes the radar (see figure 1.8 is a 42-meter fixed parabolic antenna aligned along the direction of the local geomagnetic field, and the angle between its inclination and the vertical is small.

The drift is perpendicular to the magnetic lines, and the more orthogonal to the ground the B line is, the smaller the V_{ExB}. As the magnetic lines in the Arctic are almost orthogonal to the ground, this component might be small and difficult (or even impossible) to detect with the PRIDE radar.

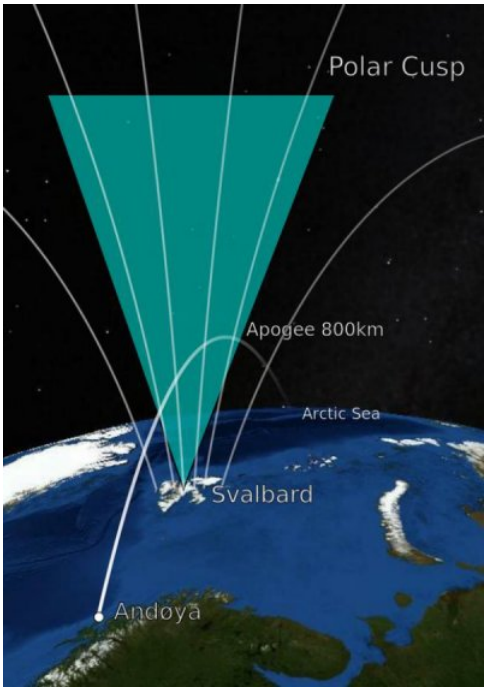

Figure 1.7: Illustration of the magnetic field lines in Svalbard, EDGAR project, SIOS, `https://sios-svalbard.org/EDGAR_2019`

Figure 1.8: EISCAT Svalbard radar, 42 meters fixed parabolic antenna aligned along the direction of the local geomagnetic field (red line). The blue line represent the local vertical. Credits: Katie Herlingshaw

1.4 PRIDE

A new instrument was installed in Svalbard during the summer 2020 to study the wave structures in the ionosphere: the Polar Research Ionospheric Doppler Experiment (PRIDE). The subsection below 1.4.1 provides general characteristics of the system, while the subsection 1.4.2 describes the components of the receiver.

1.4.1 General characteristics

A low power single frequency high frequency (HF) transmitter, located at the Polish Research Base in Hornsund, transmits a signal into the ionosphere, where it is reflected before being received at the Kjell Henriksen Observatory in Longyearbyen.
Figure 1.9 represents the archipelago of Svalbard, and the red and blue points the locations of both the receiver and the transmitter, both situated on the main island, Spitzbergen. Their coordinates are the following ones:

- Transmitter: Polish Polar Station - Hornsund : 77.00°N 15.33°E 10 m above sea level (red dot)

- Receiver: Kjell Henriksen Observatory - Longyearbyen: 78.148°N, 16.043°E , at an altitude of 520 m (blue dot)

Figure 1.9: Locations of the transmitter and the receiver of PRIDE in Svalbard

Figure 1.10 represents the path of the wave, from Hornsund to Longyearbyen. The PRIDE

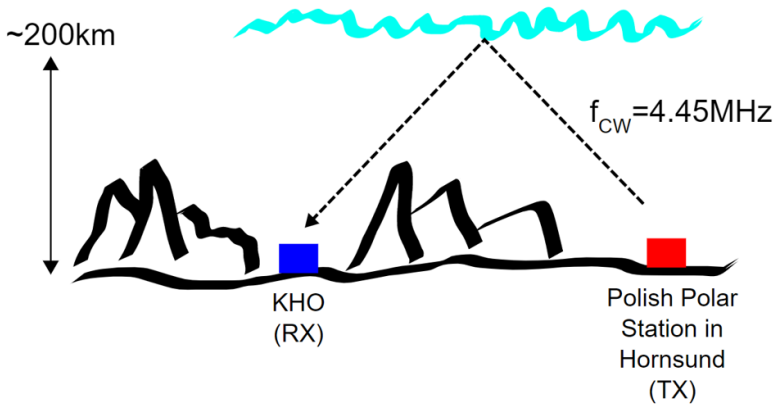

Figure 1.10: A wave is transmitted from Hornsund, reflects on the ionosphere back toward the ground, and is received at the KHO

instrument has the following operation parameters:

- Power of the transmitter: 20W
- Frequency of the transmitter: $F_{cw} = 4.45$ MHz
- Transmitted signal: continuous carrier wave at 4.45MHz

The antenna is a half wave dipole and requires no maintenance from personnel at Hornsund. The transmitter unit is stored in box at the base of the antenna does not require additional input once switched on (other than occasional maintenance, if fault develops, from personnel at Hornsund) (13).

1.4.2 Receiver description

The received uses an active loop antenna connected via a lowpass filter to a software-defined radio called USRP: Universal Software Radio Peripheral. It is designed and sold by Ettus Research and most USRPs connect to a host computer through a high-speed link, which the host-based software uses to control the USRP hardware and transmit/receive data. Figure 1.11 represents the USRP model used for PRID and its architecture can be found in the appendix A.6. The Ettus company describes the general architecture as followed:

"While some characteristics and specifications vary from model to model, all USRP devices use the same general architecture. In many cases, the RF frontend, the mixers, filters, oscillators and amplifiers required to translate a signal from the RF domain and the complex baseband or IF signals. The baseband of IF signals are sampled by ADCs, and the digital samples are clocked into an FPGA. The stock FPGA image provides digital down-conversion, functionality, which includes fine-frequency tuning and several filters for decimation. After decimation, raw samples or other data are streamed to a host computer through the host interface. The reverse process applies to the transmit chain."

Figure 1.11: N200/N210

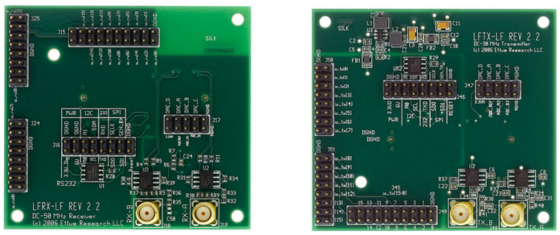

Figure 1.12: Dughterboard LFTX

The model used for PRIDE is the N200/N210, with the specifications `https://kb.ettus.com/N200/N210`.
A modular front-end, called a daughterboard, is used for analog operations such as up/down-conversion, filtering, and other signal conditioning. The LFTX daughterboard utilizes two high-speed operational amplifiers to allow transmission from 0-30 MHz and the outputs of the LFTX is processed as a single I/Q pair (complex signal). A picture of the LFTX daughterboard can be found in figure 1.12.
The receiver uses a GPS-Disciplined Oscillator (GPSDO) to provide a stable 10MHz reference signal to the receiver. A GPSDO is a combination of a GPS receiver and a high-quality, stable oscillator such as a quartz or rubidium oscillator whose output frequency is adjust to agree with the signals broadcast by GPS or other GNSS satellites. GPSDOs work well as a source of timing because the satellite time signals must be accurate in order to provide positional accuracy for GPS in navigation. These signals are accurate to nanoseconds and provide a good reference for timing applications.The receiver local Oscillator (LO) is locked

to this reference signal and the baseband IQ-samples are provided at 200kHz sampling frequency. The signal is decimated to a 100-Hz sampling rate and IQ-data are stored in files comprising all data from one hour.
The transmitter sends a sine wave at 4.45MHz. The receiver is tuned to a frequency that is 25Hz below that frequency to avoid any DC-component introduced by the analog mixer. When using a mixer, one of the input signals is typically the signal to be processed while the other one comes from a Local Oscillator (LO). Changing the frequency of the LO tunes the radio to a different frequency range. Here, the mixer is used to downconvert the received signal to the desired baseband. A summary of all the operations on the signal is depicted on the block diagram A.8. The expected frequency shifts are in the range $\pm$ 5..15Hz.

Chapter 2

Signal Processing

2.1 Terminology

In order to be as precise as possible regarding the technical terms, a small dictionary containing the more used terminologies is found below:

Time Domain dictionary

- Sampling frequency (Fs): The sampling frequency is the number of data samples acquired per second. For instance, a sampling rate of 100 samples/second means that 100 discrete data points are acquired every second. This can be referred to as 100 Hertz sample frequency. The choice of the sampling frequency is crucial to represent the correct shape and amplitude of the signal. To achieve that, the criterion of Nyquist Shannon (see section 2.2) has to be respected.

- Sampling interval (Δt): The sampling interval, also called δt, is the inverse of the sampling frequency described above. It is the amount of time between data samples collected in the time domain.

- Block size (N): The block size is the total number of time data points that are captured to perform a Fourier transform. In this study, block size of 800 samples are used to produce the Doppler shift.

- Frame Duration (T): The frame duration is the total time (T) to acquire one block of data.

All of these terms are linked by formulas: $Fs = \frac{1}{\Delta t}$ and $T = \frac{N}{Fs} = N\Delta t$ and are represented in figure 2.1.

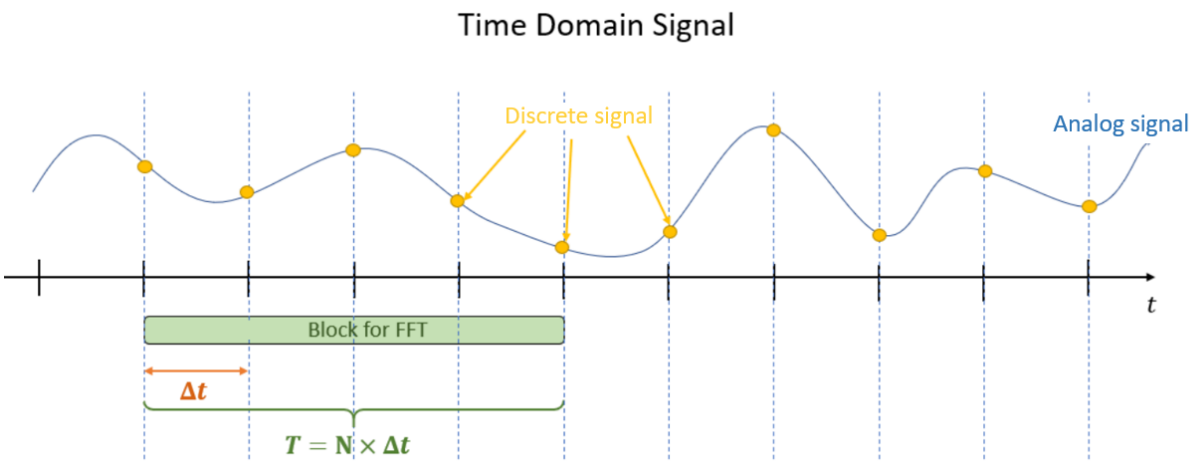

Figure 2.1: Time Domain representation, highlighting the block size, the frame size, and the time resolution

Transition from time domain to frequency domain

Spectral analysis refers to methods of estimating the spectral density function, or spectrum, of a given time series. Traditional spectral analysis is a modification of Fourier analysis,which is concerned with approximating a function by a sum of sine and cosine terms, called the Fourier series representation (17).
We are sampling from a continuous time signal and that estimation needs to be done with a Discrete Fourier Transform (DFT). This algorithm uses The Fast Fourier Transform (FFT) as DFT.

- Windowing: The process of windowing a signal involves multiplying the time record by window. The resulting signal will start and end with a zero (or almost a zero), and the time window to be analysed appears periodic. There are different shapes of windows possibles. Their amplitude is centered and varies gradually towards zero at the edges of the window. As multiplication in the time domain is equivalent to convolution in the frequency domain, the spectrum of the windowed signal is a convolution of the spectrum of the original signal with the spectrum of the chosen window.
  The most common are the Hann and Hamming windows, which are raised cosine window (because the zero-phase version is one lobe of an elevated cosine function). Hosokawa's code (Japan) (2) (`http://gwave.cei.uec.ac.jp/%7Ehfd/dat.html`) for on Python is using a Hamming window to produce the FFTs. This raised cosine window has the form :

$$w(n) = \alpha + (1.0 - \alpha) * cos((\frac{2 * \pi}{N} * n) \tag{2.1}$$

  with $0 \leq n \leq N$ with N a positive integer. The shape and response of the Hamming

window is represented in the figure 2.2. As specified in the Matlab documentation

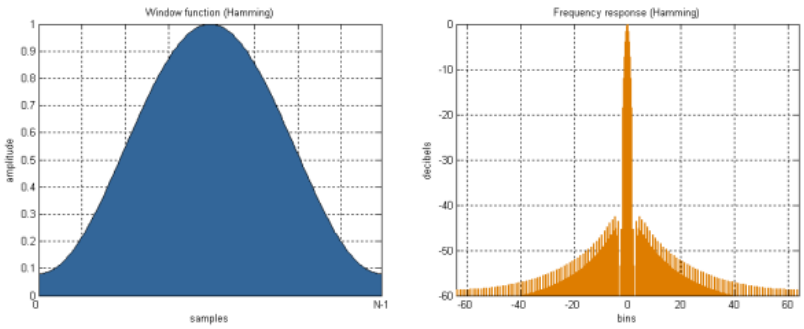

Figure 2.2: Hamming window, a0 = 0.53836 and a1 = 0.46164. The original Hamming window would have a0 = 0.54 and a1 = 0.46.

of the power spectrum function (20), the window used by default by the function is the Kaiser window. It is using the Bessel functions in order to approximate a DPSS (discrete prolate spheroidal sequence), a group of functions that are maximizing the energy concentration in the main lobe. The main lobe ends at a frequency bin given by the parameter α. The Kaiser window can be described by the function:

$$w(n) = \frac{I_0\pi\alpha\sqrt{1-(\frac{2n}{N}-1)^2}}{I_0\pi\alpha} \tag{2.2}$$

with $0 \leq n \leq N$

$$w_0(n) = \frac{I_0\pi\alpha\sqrt{1-(\frac{2n}{N})^2}}{I_0\pi\alpha} \tag{2.3}$$

with $-\frac{N}{2} \leq n\frac{N}{2}$

Figure 2.3 is a representation of the Kaiser window, using $\alpha = 2$:

- Overlap: An overlap occurs when two neighboring observation time blocks are using some of the same samples. It is the amount of samples used by two adjacent FFT, expressed as a percentage of the observation time. Overlapping can be interesting for two main reasons.
  - It prevents missing important frequency contents in the processed results: more FFT are produced, and therefore additional details are found.
  - A zero percent overlap can create some "dots" in the spectrogram, because of the window, and therefore hide some trends in the final plot. Applying overlap can help overcome these window effects. It can also make the plots more smooth and make trends in the data more observable (19).

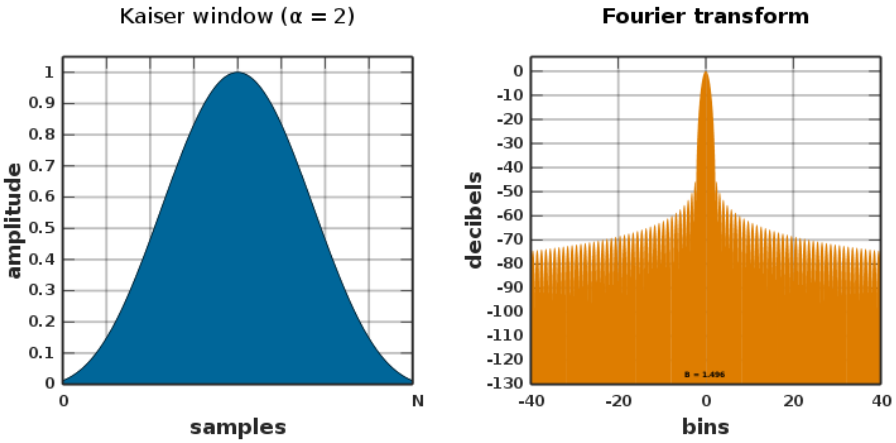

Figure 2.3: Kaiser window, using $\alpha = 2$, https://en.wikipedia.org/wiki/List_of_window_functions#Hann_and_Hamming_windows

The principle of overlap in data block is illustrated in Figure 2.4.

The algorithm used to go from the time domain to frequency domain (21) is the following one and is illustrated in figure 2.5:

- 1) Define analysis window (ex: 40seconds wideband)
- 2) Define the amount of overlap between the windows (ex: 75 percent)
- 3) Define a windowing funciton (ex: Hamming, Kaiser...)
- 4) Generate windowed segments (multiply signal by windowing function)
- 5) Apply the FFT to each segment

Frequency domain dictionary

As for the time domain, the frequency domain has its own keywords, which represent the characteristics of the process:

- Frequency point spacing (or frequency resolution) (Δf): it is the spacing (in Hertz) between data points of the DFT. This frequency depends on two parameters described below.
- Bandwidth (BW): The bandwidth is the width of the spectral window: if a signal is composed of several frequencies, with the maximum frequency $f+$ and the minimal frquency $f-$, then the bandwidth is $f+ - f-$.

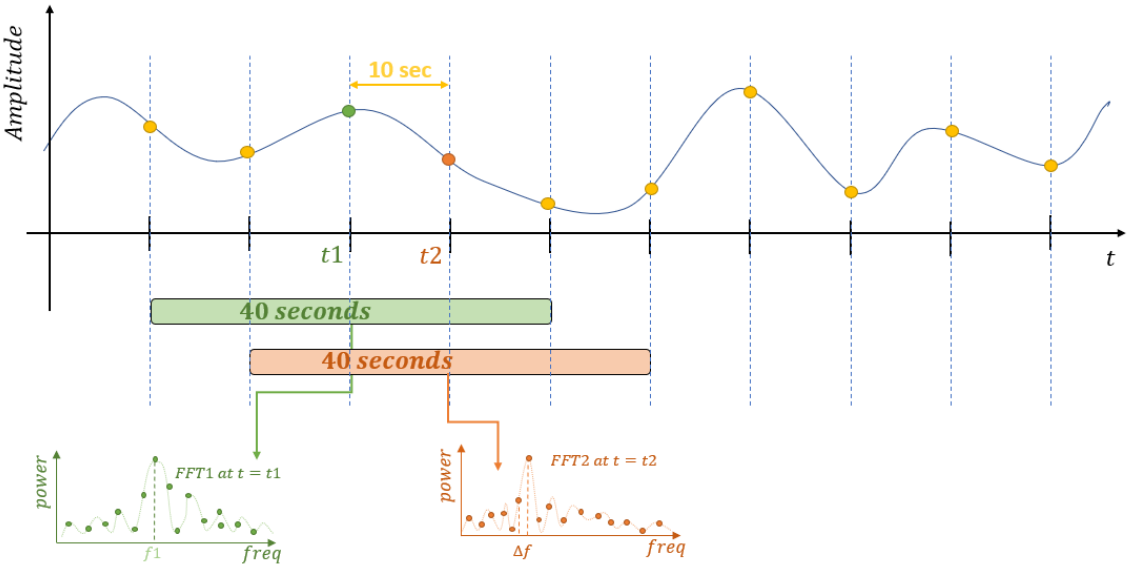

Figure 2.4: Sliding window of 40 seconds, with 75 percent overlap

- Spectral lines (SL): After performing a Fourier transform, the spectral lines (SL) are the total number of frequency domain data points. This is analogous to N in the time domain.

These three parameters are linked by the equation:

$$\Delta f = \frac{Bandwidth}{Spectrallines} = \frac{BW}{SL} \tag{2.4}$$

and can be shown schematically on figure 2.6:

Transition from FFT to spectrogram

It is then possible to create a visual representation of the frequencies spectrum as it varies with time. This representation is a spectrogram. Each portion of the signal that has been taken to produce the FFT then corresponds to a vertical line in the image. An overviex of the process is illustrated in figure 2.7: The spectrogram is a way of having multiple FFTs on one plot. Each time stamps is showing a frequency versus power a that time.
As for the others plots, the spectrogram has its own specific parameters:

- Time Resolution: The time resolution of spectrogram is expressed in seconds, and this argument controls the block size needed to compute the short-time power spectra that form spectrogram estimates. Therefore, the temporal resolution is the duration of a pixel of the spectrogram. When using Matlab routines, the 'TimeResolution' can not be specified simultaneously with 'FrequencyResolution'

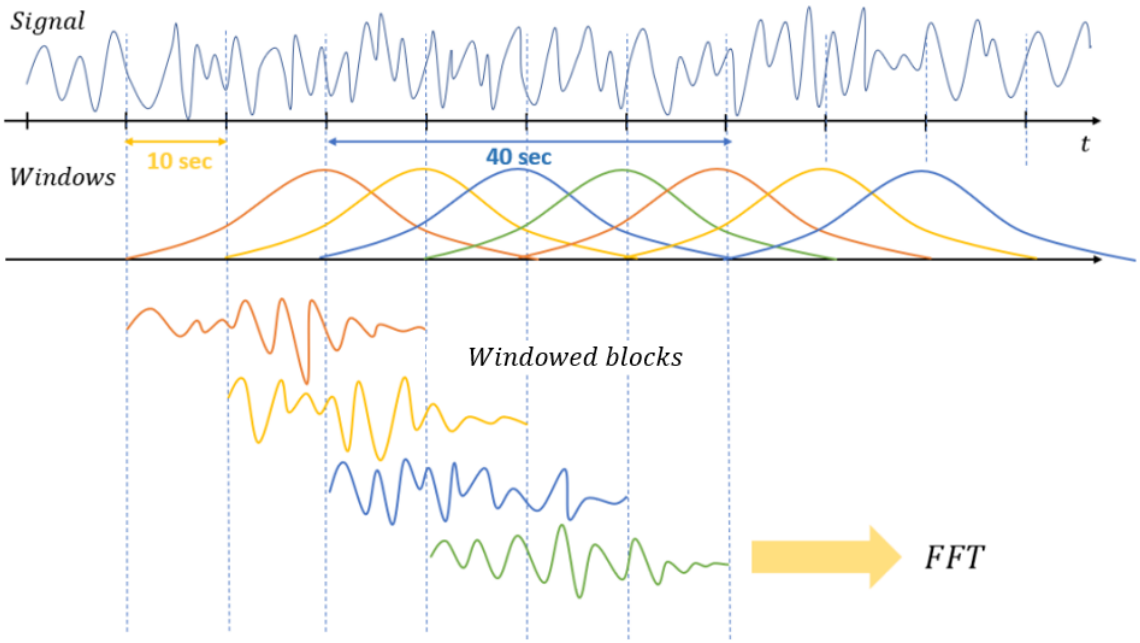

Figure 2.5: Creating the FFT from the time domain signal, using overlapping windows using a 40-seconds windowing with a 75-percent overlap

2.2 Nyquist Shannon Theorem

The Nyquist Shannon theorem for a real signal (ie. no complex components): It states that the minimum sampling frequency of a continuous signal that will not distort its underlying information should be double the frequency of its highest frequency component. If f_S is the sampling frequency, then the critical frequency (or Nyquist limit) f_N is defined

$$f_N = \frac{f_{max}}{2} \tag{2.5}$$

If a real signal is sampled at a frequency F_S, it is possible to unambiguously represent the frequency content in the region $[0; F_S/2]$. No additional information can be held in the other half of the spectrum when the samples are real, because of the conjugate symmetry exhibited by real signals in the frequency domain.

However, this symmetry does not apply when it comes to complex signals, so a complex signal sampled at rate fs can unambiguously contain content from 0 to f_s. The total bandwidth (BW) of the signal is therefore f_s. The Nyquist Shannon criterion for a complex signal becomes therefore: If F_S is the sampling frequency, then the critical frequency (or Nyquist limit):

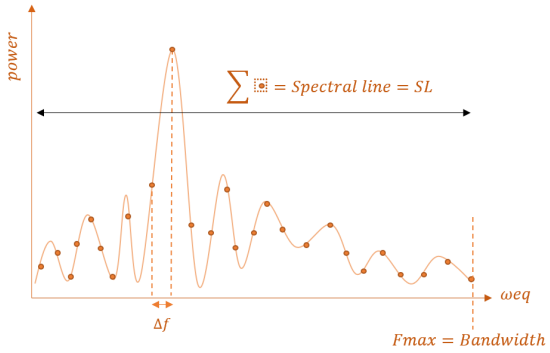

Figure 2.6: FFT with the frequency resolution highlighted

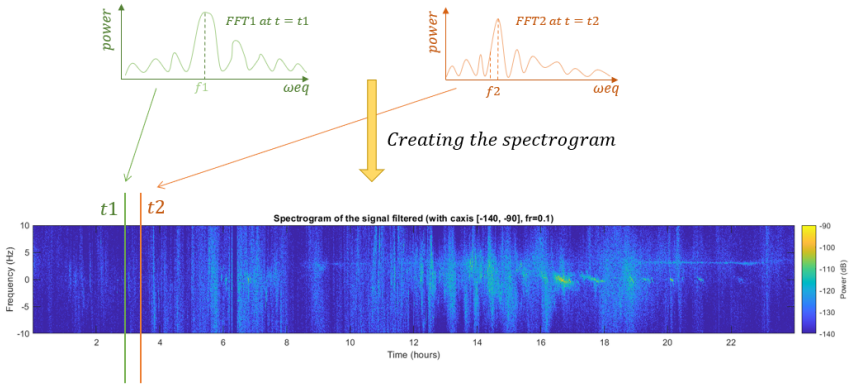

Figure 2.7: Creating the spectrogram from the FFT of the signal

$$f_N = f_S \tag{2.6}$$

2.3 Signal processing operations

The algorithm to process the data is using basic signal processing operations in order to make the signal as clear as possible, within a reasonable file size. During the process, the signal is going through different steps:

Downsampling

In digital signal processing, downsampling can be seen as a compression, as we are getting rid of the parts of the signal that are of no interest. The point is to reduce the amount of data to be processed by an interger factor M. It produces an approximation of the sequence

that would have been obtained by sampling the signal at a lower rate. The new signal y obtained by down-sampling the signal x by a factor of M can be seen as the following equation:

$$y(n) = x(Mn), n \in \mathbb{N}$$

Decimation

The decimation is a combination of a lowpass filtering and a downsampling. It accomplishes a reduction of the sampling rate by a factor M of a given $x(n)$ signal after this signal passes through an anti-aliasing filter h(n). Anti-aliasing filters are always analog filters as they process the signal before it is sampled. In our case, we are using a low-pass filter with a cut-off frequency chosen in order to respect the Shannon criterion so the replicated spectra of the sampled signal do not overlap each other.Figure A.7 in appendix represents the decimation technique conducted on a signal, step by step. A filter is applied to the digital signal, and a down-sampling operation is then conducted (22).

The intermediate signal $x'(n)$ and the final signal $y(n)$ can be expressed as:

$$x' = \begin{cases} x(n) & n = 0, \pm M, \pm 2M \\ 0 & else \end{cases}$$

$$y(n) = x'(Mn) = x'Mn)$$

This operation also has an influence in the frequency domain. When a discrete signal is compressed, events in the signal happen over a fewer number of samples. The more the signal is compressed (and will therefore get close to an impulse event), the more the it will appear like a constant in the frequency domain. If the signal is compressed in time domain by a factor of M, it will be dilated by this same factor in the frequency domain: The interval $[0 : \pi/M]$ becomes the interval $[0 : \pi]$. The change of spectra in the frequency domain is illustrated in figure 2.8, with an example of M=4 (22).

Lowpass filter

Lowpass filters are used at several steps in the processing code in order to remove the high-frequencies components.
After the signal is received at 4.45MHz, a lowpass filter is installed with a cut-off frequency of 30MHz.
The received signal, after the pre-processing steps, has a DC component at f=0Hz, and the signal of interest is at approximately +22Hz. The signal is mixed with a -25Hz signal to bring the signal of interest at 0Hz and the DC component at -25Hz, and then a decimation process that includes a lowpass filter is applied to remove of the DC component and the high frequency perturbations.

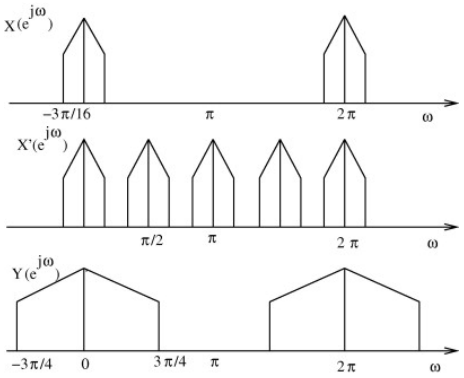

Figure 2.8: Frequency spectra of downsampled signals for M=4. Credits: Anke Meyer-Baese, Volker Schmid, in Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition), 2014 (22)

2.4 Block Diagram

The diagram A.8 in the Appendix provides an overview of the signal processing.

Chapter 3

Treatment of missing samples

3.1 Identification of the problem

The PRIDE instrument is producing one hour data files. If we were to receive everything correctly, we should have exactly 360000 samples per hour (60 minutes x 60 seconds/minutes x 100 samples/second). However, an overlook at the previously collected data highlights that this is not the case: Full files are rare, and most of them always have 5-7 samples missing. The tables below 3.1 3.2 and 3.3 indicates the maximum amount of missing samples on a 1 hour file per month, and the time it represents.

Month	A*	S*	O	N	D
Nb Gap	298655	8	8	8	8
Missed time (s)	2986.55	0.08	0.08	0.08	0.08
Total per month (gap/month)	324379	4149	3989	4144	4141

Table 3.1: Maximum missing samples on a 1 hour file, per month, in 2020

Month	J	F	M	A	M	J*	J	A*	S	O	N	D
Nb Gap	8	8	8	8	8	8	8	27	8	8	8	8
Missed time (s)	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.27	0.08	0.08	0.08	0.08
Total month (gap/month)	4456	3991	4453	4298	4452	3415	3578	2638	3532	3988	4146	4143

Table 3.2: Maximum missing samples on a 1 hour file, per month, in 2021

The * that stands after some months means that not all the data files are available for these months.

Month	J	F*	M	A*
Nb Gap	8	8	26	27
Missed time (s)	0.08	0.08	0.26	0.27
Total per month (gap/month)	4480	3556	9306	11180

Table 3.3: Maximum missing samples on a 1 hour file, per month, in 2022

Looking at these tables, two different types of gaps can be highlighted:

- Files that contains short amount of missing samples (about 5-7 samples). They are the most common and almost systematic ones. This can be seen as a problem in the receiver software/system (related to the decimation from 250kHz sampling frequency to 100Hz). It must be said that we are only receiving one signal, while the other scientific group working on similar radars (for instance in Japan (2))are receiving many signals simultaneously. This detail could partly explain why samples are missing.

- Longer chain of missing samples. They are quite unusual, but a power cut or a maintenance operations on the system might cause several minutes of missing samples. This is for instance the case on August 2020. On this day, the head engineer at KHO has been working on the receiver PC, which explains the large amount of missing samples during this day. Indeed, on August 24th, the 13UT file is missing 22055 samples (7% of the data), and the 14UT file is missing 298655 samples (80% of the data). Even if none of this problem has been reported since, it is an issue that might happen and needs to be tackled for future data collection.

The work presented in this sections will focus on how to handle the missing samples, and decide on a criteria for which method to use: removing the whole data file, completing the gaps with "0", or using an interpolation function. In other words, we are looking for a routine to deal with incomplete data sets, depending of the number of missing points.

3.2 Test signal

In order to choose how to deal with gaps in our data, test signals will be used. Using Matlab, a synthetic and representative signal is created and a group of data are removed in order to simulate missing samples. Different methods of interpolation will be tested to find the most robust and relevant one. Starting with clear sinusoidal signal,a Gaussian noise will then be added make the test signal more realistic.

3.2.1 Single sinusoidal test signal

The signal has been created with a typical ULF wave frequency of $12.5mHz$ (4), as the ULF waves have a shorter period than the Gravity waves. Therefore, using the wave with the highest frequency leads us to a "worst case" study. The signal is sampled at 100 Hz, with 360000 points per hour, as in the real data PRIDE. In order to observe how the algorithm reacts nearby the missing samples, we are creating a subsignal with a length that can be change in the algorithm. The aim is to observe the signal on a few periods, exactly as it will be observed in the real Doppler data (Accoridng to (9) paper, where signals containing 9, 4 and 5 periods were observed). On figure 3.1 and 3.2 are gaps of 6% and 25% in the signal, out of a 80000 points (which represents around 13.20 minutes of subsignal in total):

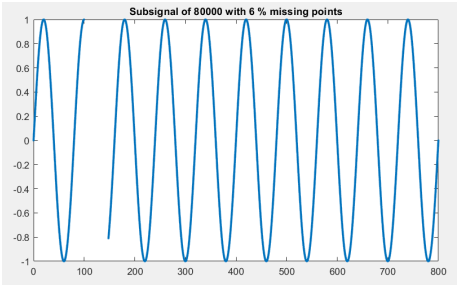

Figure 3.1: Clear synthetic ULF signal with a frequency of 12.5 mHz and a gap of 6%

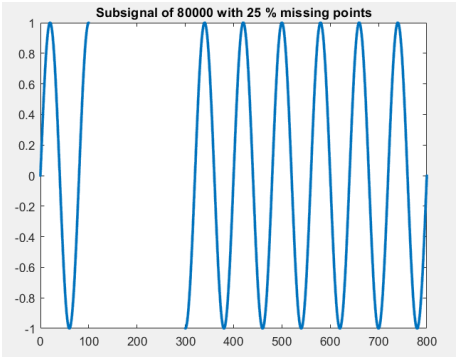

Figure 3.2: Clear synthetic ULF signal with a frequency of 12.5 mHz and a gap of 25%

Interpolation techniques from Matlab

After being created, the subsignal is then processed using the Matlab function interp1(). Matlab provides different interpolation techniques, and a table that summarizes all of them can be found in the appendix A.9.

From this table, some observations can be done:

- The techniques “nearest”, “next” and “previous” might not be relevant as they will not follow the curves. The interpolation will be made by using horizontal straight lines.
- The “Linear” method will only join the two nearest points, regardless of the curves or period.

- Assuming that memory and computation are not one of our major issues, all 4 techniques, 'pchip', 'cubic', 'makima' and 'spline' can be used without prioritizing one of them. However, if this criteria becomes important later, it must be observed that the ranking is:
  computation time $_{pchip}$ <computation time $_{makima}$ < computation time $_{spline}$.
- The method 'cubic' requires regularly spaced data, which is not our case, as we have data- gap for x points – data again. This method can therefore not be use.
- The technique "spline" requires a C^2 condition to work properly.
- "pchip" is flattening aggressively.

Looking at this first consideration, 4 techniques will be tested: "linear", "makima", "pchip" and "spline". We are looking for a technique that preserves the waving movement and the period of the wave. The amplitude comes in second consideration.

Results

Two subsignals of 80000 points with different % of missing samples have been tested. Figure A.10 in appendix contains a 6% missing samples (less than one full period) and figure A.11 in appendix contains a 25% missing samples, which includes several missing periods. For all graphs, the resulting signal is plot in blue, and the initial signal is represented with dotted red points. It cna be observed that:

- None of the waving structure is represented with "Linear" method, as expected.
- Both methods "Makima" and "pchip" are reacting the same way. A waving movement is depicted but it modifies the period of the wave by not taking in account the lowest part of the signal.
- For the moment, "spline" is the most accurate one, as it modifies the less the signal's period.

None of these functions are fairly representing the signal if many periods are missing. The spline one is joining the border points with a more curvy line than the others, but the initial period is not conserved. This phenomenon can modify the interpretation of the scientist and should be avoided.

3.2.2 Synthetic signal with Gaussian Noise

A Gaussian noise is added to the previous signal using the function awgn(), that allows the Signal to noise ratio (SNR) to be changed. Stating from the previous results, the conclusion

that the interpolation should not be conducted on an amount of missing samples longer than one minute arises. The amount of missing samples will therefore be fixed to that to that limit while the SNR can be changed. Our subsignal contains 80000 points, and the maximum gap corresponds to 6000 points, which represents 7.5%. A SNR ratio of 20 is chosen, to represent the real signal.
The time representation of the subsignal is reprensented in figure 3.3.

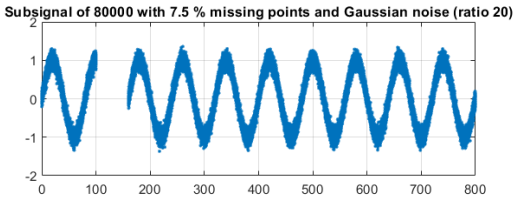

Figure 3.3: Noisy subsignal, SNR 20

Figure 3.5 represents the interpolation with the 4 methods. An other Gaussian noise is added to the signal after interpolation, re-create the same background as the initial signal.

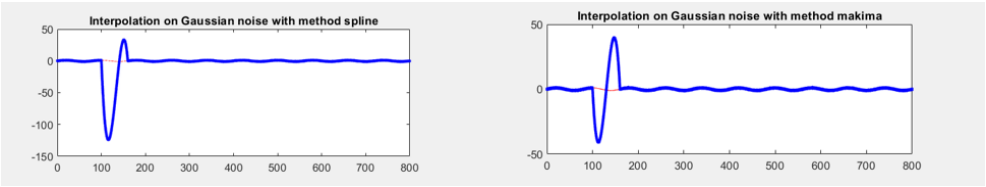

Figure 3.4

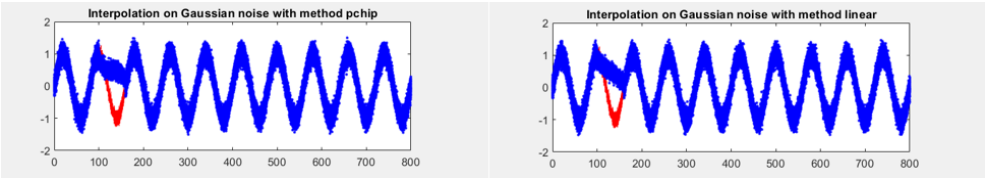

Figure 3.5: Interpolation of noisy subsignal, SNR 20

Looking at these plots, we can see how different the results are when we are adding Gaussian noise (ie getting closer to real dataset). While the "linear" and "pchip" methods are providing acceptable result in a sense that the solution does not diverge, this is not the case for the "makima" and "spline" methods, where the curves are getting out of control. This solutions does not make sense are are not robust for not continuous nor derivable functions. But as the behaviour of these functions can change aggressively, the importance to test them with real data is a necessity.

3.3 Test on real Data

We are using the data from 13th February 2021 14UT, as they provides a clear and changing signal for several minutes. The data are loaded, and the equivalent of one minute of data at 14:30UT is removed. Figure 3.6 represents the regular spectrogram, and the one with the missing samples: On the second plot, the missing minute is not shown, and the spectrogram "stick" the samples right before and after the removed minute.

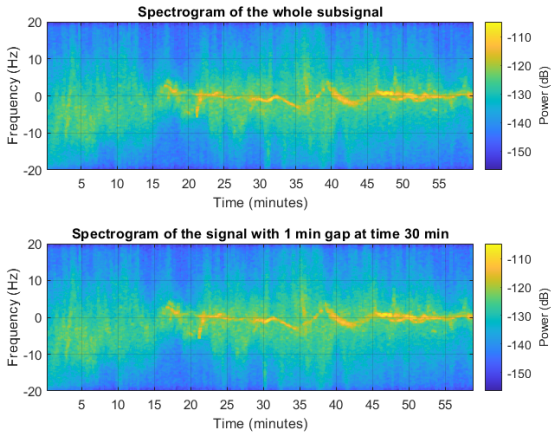

Figure 3.6: Spectrograms of the subsignal and the truncated subsignal with a 1min gap

The four previous methods of interpolation are tested on this real non continuous dataset. Figure 3.7 and figure 3.8 represents these results. A zoom has been made to observe how the newly built data interact with the initial ones and the background:

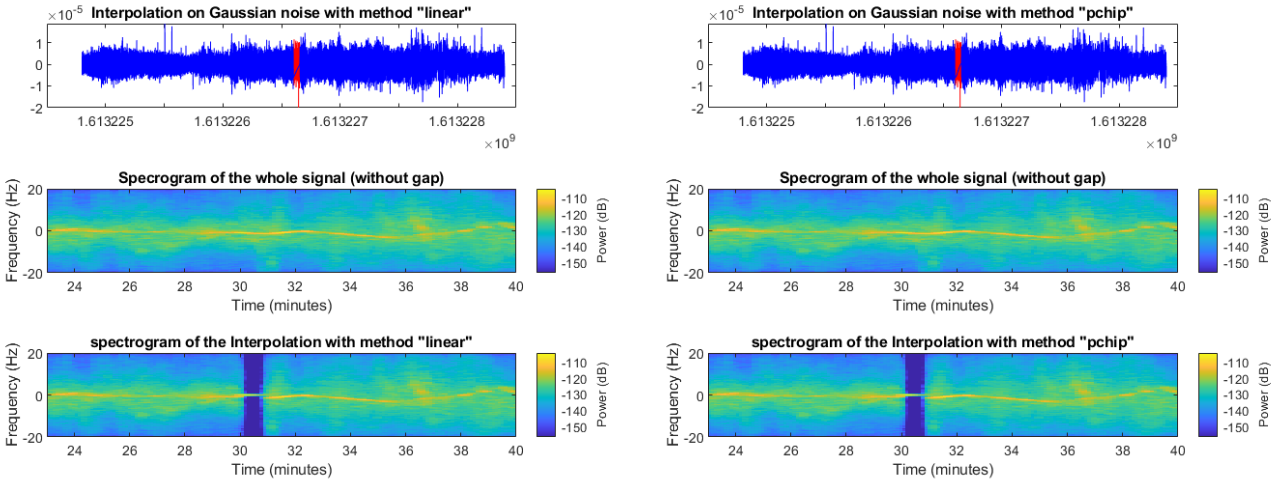

Figure 3.7: Interpolation using method "linear" (left) and "pchip" (right)

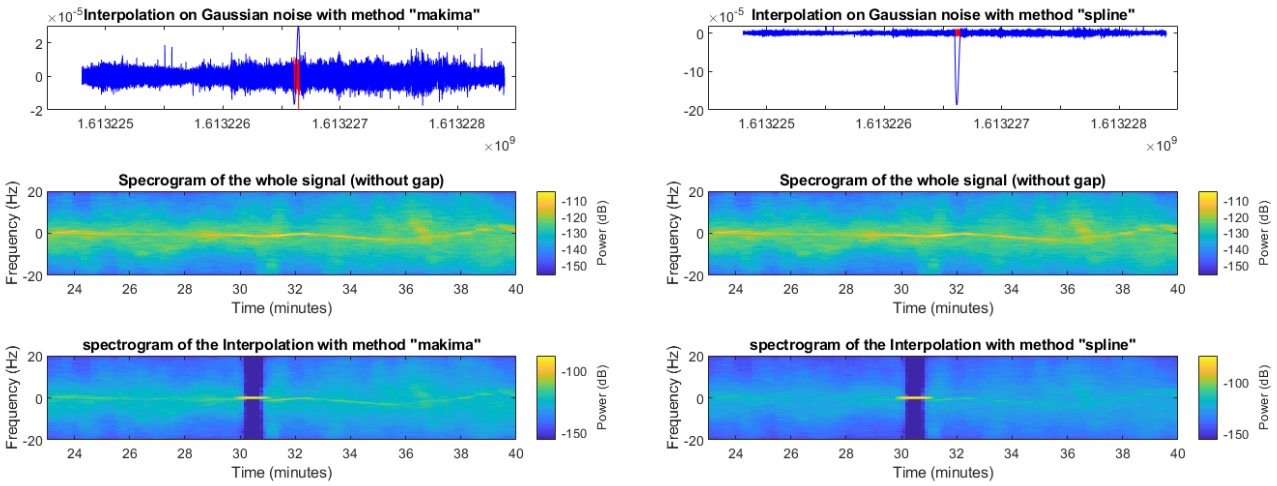

Figure 3.8: Interpolation using method "makima" (left) and "spline" (right)

The results are globally the same as for the noisy synthetic signal. It can be seen on the time representation that both methods "spline" and "makima" diverge. on the opposite, "pchip" and "linear" are less ambitious, as they do not reproduce the oscillations of the signal, but they do not diverge.

3.4 Choice of the method and implementation of the algorithm

3.4.1 Method of decision

Taking in account the theorical use of each function i.e. for which signal it has been made for, how each method react with the signals, and properties like memory requirement, a matrix of decision can be dressed in order to decide which method will be used. This function might be useful for files with less than one minute of missing samples. For longer amount of missing samples, it has been decided to follow the procedure used by scientists (in Tromsø) who are using a similar instrument, and apply "0" to fill up the missing samples.
In order to sort the interpolation methods in a relevant way, 5 criteria have been chosen. Based on these criteria, we try to find a mark for every methods we have. The mark was from 0 to 5. We finally decided to apply a weight for all these criteria :

- Theoretical use of the method **Weight=2**
  - 5 = The method is perfectly adapted for the (waving) data we have

  - 0 = not at all adapted to our dataset

- Memory and computation time **Weight=1**
  - 5 = The method has a very low complexity and computation time
  - 0 = High complexity of the method

- Reaction to synthetic signals **Weight=1**
  - 5 = The method works very well for interpolating synthetic signals, and respect the waving behaviour
  - 0 = The method does not interpolate fairly the data

- Reaction to real data **Weight=3**
  - 5 = The method works very well for interpolating our real data
  - 0 = The method does not interpolate nicely the data

- Robustness **Weight=3**
  - 5 = The method is robust and we do not fear possible divergence
  - 0 = Not robust at all

The method was therefore the following one. We try to evaluate as precise as possible the mark for every interpolation function for all these criteria. We then sum of all them to give a global mark to each objective. With our method the goal was therefore to have the best mark.

3.4.2 Method of decision

The global mark, taking into the weight associated to the criteria. This is given in the table 3.4.

With the respect to this table, we can observe that a group of methods have a higher mark than the other. This is mostly due to the fact that "spline" and "makima" diverge and are therefore not robust for our signals. The choice between "pchip" and "linear" is more challenging, but we decided to choose "pchip" as an interpolation function, according to the matrix. At this time of the thesis, the relevance of an interpolation function has been questioned, and it has been chosen to follow the other studies method and to complete the missing samples with zeros. An implementation of an interpolation function might be conducted later.

	Theory	Computation	Synthetic signal	Real data	Robustness	Total
Weight	2	1	2	3	3	
Linear	2	4	2	4	5	**39**
Makima	4	3	2	2	2	**27**
Pchip	3	3	4	4	4	**41**
Spline	4	1	1	2	2	**23**

Table 3.4: Matrix of decision

3.4.3 Implementation of the algorithm

The interpolation routine will then be implemented in MATLAB, and will take place before decimating the signal with a factor of 2500. The advantage to complete the missing samples with zeros at 250kHz sampling rate is that it should result in no visible missing samples at 100Hz data.
The final code will be tested with a noisy test signal sampled at 250kHz where samples will be removed, and the code will recreate all the path from the raw data to the Doppler shift.

Code description

The code described below is available in the appendix **??**. An "if" test is implemented inside a for loop to spot a possible missing sample by checking the difference between two consecutive time elements (theorical value is 4E-6s, and an empirical ϵ is chosen as a margin). At the end of the loop, two vectors are created. The first one "missing_points", contains the positions of the first missing sample, and the second one "amount_of_missing_points" contains the number of missing samples that followed the first one.
The next step contains another loop and add "Zeros" to completing the missing samples. A verification test is made after the principal loop to ensure that all the missing points have been detected. The test is a comparison between the length of the signal and the theoretical one. If samples are missing, they are then added at the end of the signal.

Code test

The signal from section 3.2.2but with a sampling frequency at 250kHz is tested. This represents 900000000 samples. A subsignal of 250000 samples is extracted to reduce the amount of time needed to process the file. Inside this signal, different gaps are created and will all be filled up with zeros.

- One missing sample

- A group of 5 missing samples.
- A larger group of 1000 missing samples.

In total, the test signal contains 248994 samples, and 1006 points will be added to the signal. At the end of the algorithm, the test signal contains 250000 samples. It is then relevant to see how this algorithm is changing the analysis of the signal. Figure 3.9 is a zoom around the missing samples to highlights the difference between the two signals in the time and the frequency domain. Small changes can be seen on the PSD. The differences between the values of both signals PSD is represented in figure 3.10. The maximal difference is $\Delta = 5.69 * 10^{-3}$ kHz.

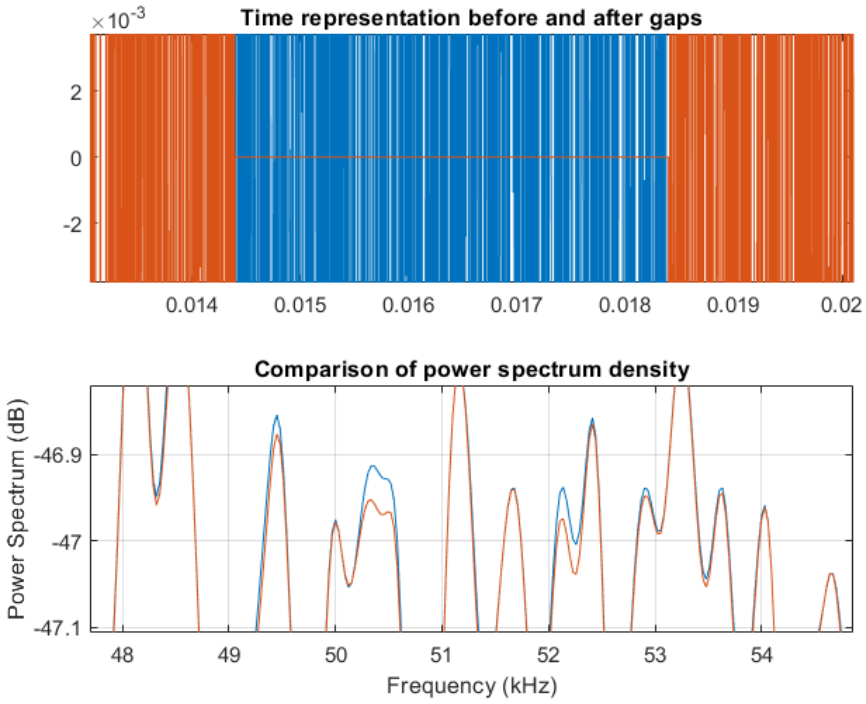

Figure 3.9: Zoom of the precedent figure: Time (figure up) and PSD (figure down) representation of both the pure (blue) and of the processed (orange) signals

Decimation process

This first test gives a first overview of how the signal will be affected by completing the missing samples with zeros. But this large amount of data is time consuming and a first decimation process has to be conduct. Here appears a choice between a large decimation (which will allow a quick time of calculus but the missing points will be more visible in the final signal) and a small decimation factor (longer processing time but smoother signal).

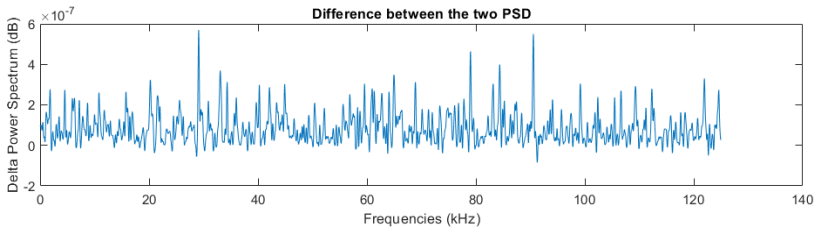

Figure 3.10: Time (figure up) and PSD (figure down) representation of both the pure (blue) and of the processed (orange) signals

We are starting from a sampling frequency of 250kHz to reach a sampling frequency of 100Hz, and the Nyquist-Shannon criterion 2.2 has to be respected. After using a lowpass filter to remove the components that are above 100Hz, a first decimation is conduct, missing samples are completed and the signal is decimate again to reach F_s= 100Hz.
Previous research stated rules to conduct the decimation process with the smallest cost and computation time (23). Here are the main ones:

- As previously written, all down-sampling factors should be chosen with respect to the Nyquist Shanon criteria for complex signals.
- The computational and memory requirements of the filters can usually be reduced by using multiple stages.
- It is possible to decimate in multiple stages as long as the decimation factor M is not a prime number.
- Using two or three stages is usually optimal or near-optimal.
- Decimation should be conducted in order from the largest to smallest factor. In other words, the largest factor should be used at the highest sampling rate.

As a summary, figure 3.11 explains how the signal will be processed by the algorithm.

Validation of the process

Once the code has been properly implemented, tests have been made with several groups of missing points. The algorithm is able to recognize the place where and how many samples are missing. They are replaced by zeros, and the timestamps are completed.
In order to do the validation of the system and to the test signal from section 3.2.2 is used and two periods are removed from the signal. The signal is processed and figure 3.12 represents the comparison in time and frequency domain between this signal and the

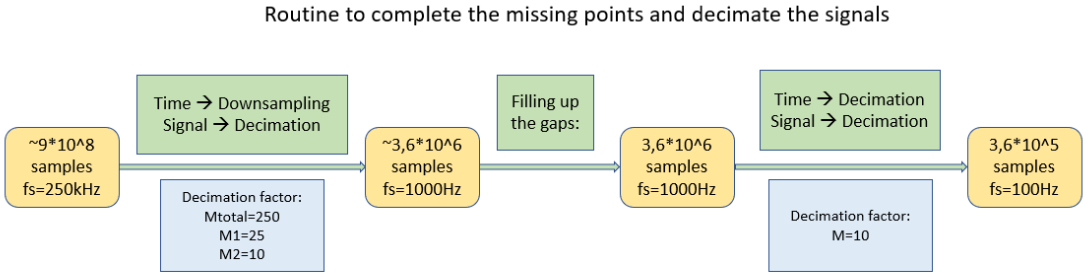

Figure 3.11: Routine to decimate the signal and fill up the missing samples

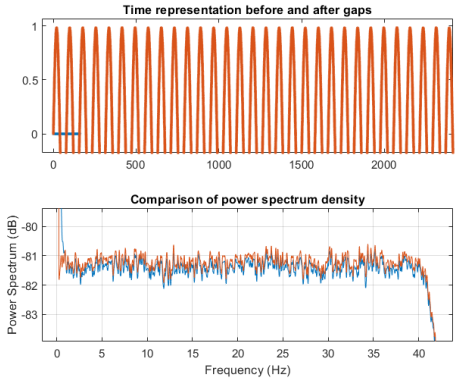

Figure 3.12: Comparison between the original and the signal with zeros added, both decimated at 100Hz. First plot is the time representation, and the second one is the PSD of both signals.

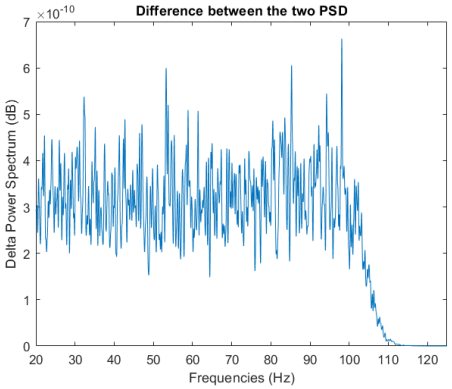

Figure 3.13: Difference in dB of the two PSD of the signal: the original and the signal with zeros added, both decimated at 100Hz.

testimony one (without missing samples). The figure 3.12 represents the difference between these two PSD.

The scale of the changes are in 10 exp −10. Knowing that the scales is in decibels, we can consider the impact of adding zeros on a two-periods missing samples almost insignificant. The limit that will decide where the acceptable maximal amount of missing samples should be placed can therefore be at least as big as two periods.

3.5 Conclusion: Maximum amount of missing points allowed

The aim of this section was to find how many samples can be missing without being noticed in the final signal. We decided to answer this question by looking at it from both a scientific and an engineer point of view.

Looking at the issue on a scientific point of view raises new questions. The aim of this thesis is to provide a clear and reliable dataset to the scientific community. This includes providing data that are clear enough to not be misunderstood. More precisely, we want the period to be clearly identifiable and avoid a period misinterpretation due to an interpolation or a group of missing samples. A worst case study can be imagined: According to the paper (9), between 4 and 9 periods are usually observed, and the smallest ULF wave has a period of 80 seconds (4). To be able to identify clearly the period of the wave, we consider that at least two periods are required. This leads to an allowed missing periods of two with a 80 seconds period wave. Sampled at 100Hz frequency, this represents 8000 points in the 100Hz file. , that has been decimated by a factor of 2500 from the raw file. This means that the maximum amount of missing points allowed on the raw file (before the decimation at M=2500) is $4 * 10^7$.

As the section 3.4.3 highlights, the differences between the PSD of the original signal and of the one that has been modified are small, even for a two period missing samples of an ULF waves. As a power cut or a longtime breakdown of the received signal is unusual (only two time over three years of data), we will consider that the engineering criterion is not the one that will decide how small the amount of missing samples should be. Group of missing samples can be handled and completed with zeros, without any strong impact on the final plots if conducted before a decimation. This is the technique used by other scientific groups that are working with the same type of instrument. Here, a decimation by a factor of 5 will be conduced after the filling with zeros. However, the missing samples are lost forever and can not be brought back in the datafile.
But if at some point ones think that an interpolation would really make a difference and would represent the missing points in a more accurate way while bringing a significant change to the plots of data with missing points, the section 3.4.1 could indicate him or her which Matlab functions are the most precise ones. At this point of the project, we agreed that this is a work that might take place later.

Chapter 4

Observations of Waves in the Ionosphere

4.1 24 hours plots and parameters

In order to identify some possible waves in the data files, 24 hours spectrograms will be plotted with our routine. We are using the filtering routine, the code to plot all the 24 hours files and the windowing from **??**, and the "completing the missing samples" techniques from 3.
Ground magnetometer data was used to identify times of potential ionospheric wave activity. Both the X and the Y components are examined. The PRIDE dataset was then examined for corresponding wave signatures.

File conversion

The data from PRIDE are provided in the format .npz, that can not be analyzed by Matlab. These files need to be converted into a .mat format in order to be readable on MATLAB. However, the .npz format is a good alternative for use with Python. The data files contain two parameters: the time and the complex voltage (I/Q) values sampled at 100Hz (iq). The process has been automated in order to create the one hour file data sets in .mat for a chosen day. The code is detailed in appendix A.1 using the third of February 2021 as an example.

Data plots

As detailed in the introduction of this chapter, the algorithm is taking the 100Hz sampled data files and is creating full data files by assigning '0' to the missing points, before processing the data. As detailed in section 2.4, filtering operation is required on the 100Hz

data, either by using a regular lowpass-filter or by implementing a decimation with a factor of five. Both techniques will be tested on all the days of interest, in order to determine which method is the most reliable and relevant one. After filtering/decimation a dynamic DFT is performed on the data using a 40 second window with a 10 second overlap (as discussed in section 2). This is done for 24 hours worth of data and produces a standard spectrogram with spectral power as a function of time (on the x-axis) and frequency (on the y-axis). For each FFT, the maximum spectral power, f_{max}, is determined and plotted out as a function of time to produce a Doppler plot.

This represents the Doppler shift, and 8640 points will therefore be plotted as a function of time.
As a summary of the algorithm routine, the schematic 4.1 explains how the 24 hours data files will be processed to be visualized.

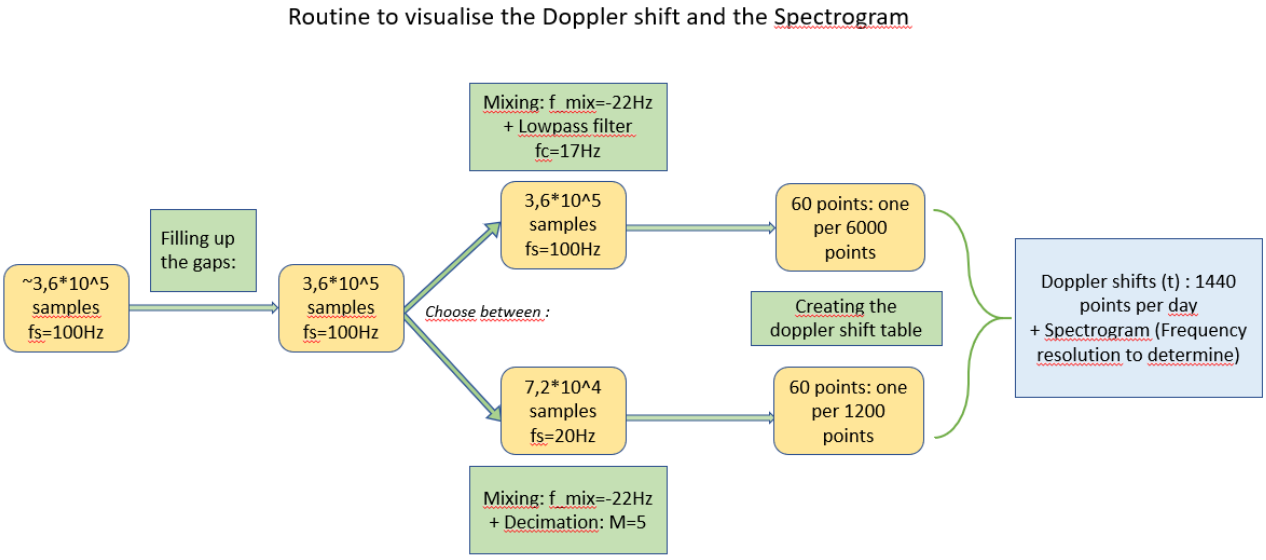

Figure 4.1: Two possible routines to visualize the data. The next section will determine which filtering technique is the most relevant one.

In order to decide which technique is the best one, ground magnetometer data from the Longyearbyen and Ny-Alesund stations from the IMAGE (International Monitor for Auroral Geomagnetic Effects) magnetometer chain are used. The IMAGE chain consists of 42 magnetometer stations maintained by 9 institutes from Finland, Germany, Norway, Poland, Russia, Sweden and Denmark. `https://space.fmi.fi/image/www/index.php?`. Nine days of interest, where some ULF wave activity was observed in the magnetometer data, have been chosen, as well as Three other random days to see how the representation was affecting "normal" days.
The summary of these experiments can be found in table A.1 in appendix. It contains the

empirical optimized parameters for both techniques (lowpass filtering or decimation) and indications about which technique gives the best representation.
Both the lowpass filtering and the decimation process solutions are providing good results, and the data can be observed precisely with both methods. However, the decimation allows us to work with less points while keeping a good representation. We will therefore choose this technique and all the data will be processed with a decimating factor of five, and the final sampling rate will be 20Hz.

In order to highlight the results presented in this table and to justify the choices of the parameters, two examples can be highlighted below.
Figure 4.2 shows data from the 10/02/2021. As written in the table, a clear, coherent signal can be observed. The figure 4.2 represents the signal of this day on a 24hours plot and compares the choice of parameters on decimated signals. The plot on the top has been processed with a 'FrequencyResolution' of 0.1 Hz, and a scale of color on and set at 'caxis'=[-140, -90] dB. This dB range is most suitable for the vast majority of datasets, highlighting the Doppler signal while reducing the background noise. The plot in the middle represents the same signal, but without any further signal processing and allowing Matlab to chose the dB scale limits and frequency resolution most suitable for this dataset. The plot at the bottom represents the Doppler shift. From this plot, it can be observed

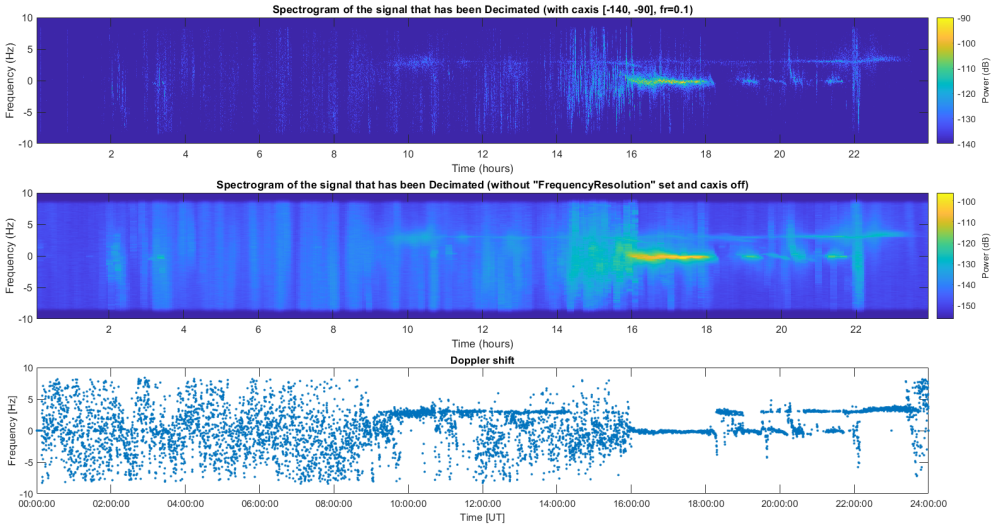

Figure 4.2: 24 hours plot from the 10/02/2021. Fig a (top) is with parameters set, fig b (middle) is without any parameters set, and fig c (bottom) represents the Doppler shift

that setting the Frequency Resolution at 0.1 Hz helps visualize the signal while making it more precise. The small scale variations can be better observed and its variations are more precised. Moreover, the colorbar on helps reducing the background noise and is set in order to highlight the Doppler signal that we want to observe. Figure 4.3 shows the

same dataset but only from 10.30 - 20.00UT. On this plot, it can clearly be seen that a setting of the Resolution Frequency is necessary in order to highlight the small variations in the signal, which is exactly the data of interest. Therefore, this first example demon-

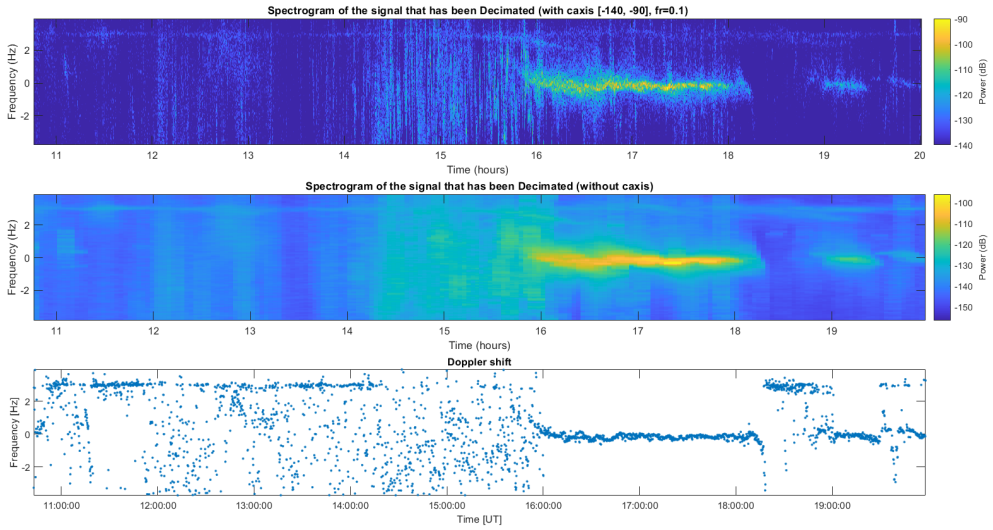

Figure 4.3: A zoom of the 24 hours plot on the 10/02/2021. Fig a (top) is with parameters set, fig b (middle) is without any parameters set, and fig c (bottom) represents the Doppler shift

strated the necessity of a choice of a resolution frequency that allows the waves to be seen, as well as a colorbar on in order to reduce the background noise and to highlight the signal. This colorbar has to be the same for every plot in order to create a uniform data set.

The second example has data from 10/03/2021. On this day, a very clear and long signal is observed, as well as strong background noise and possibly another signal during the afternoon. All in all, over the whole day, three different types of signal can be highlighted: the background noise, the clear signal, and the superposition of the Doppler signal, the background, and an other signal. The plots are shown in figure 4.4. As for the previous plot, the subplot on top represents the signal with all our chosen parameters on, and the second plot has the 'FrequencyResolution' on at 0.1Hz, but the 'caxis' function is off. The bottom plot represents the Doppler shift. Once again, the colorbar set at [-140, -90] dB highlights the signal while reducing the background noise. The other parasite signal can be easily spotted on the spectrograms, but as it is less powerful that our waves, it does not appear strongly at the beginning of the Doppler shift plot. When our signal gets overwhelmed by the background, both the spectrograms and the Doppler shift gets confused, and it is difficult to identify things, despite the Frequency Resolution parameters and the color scale. However, the figure 4.5 is a zoom on the clear signal from this day and coherent

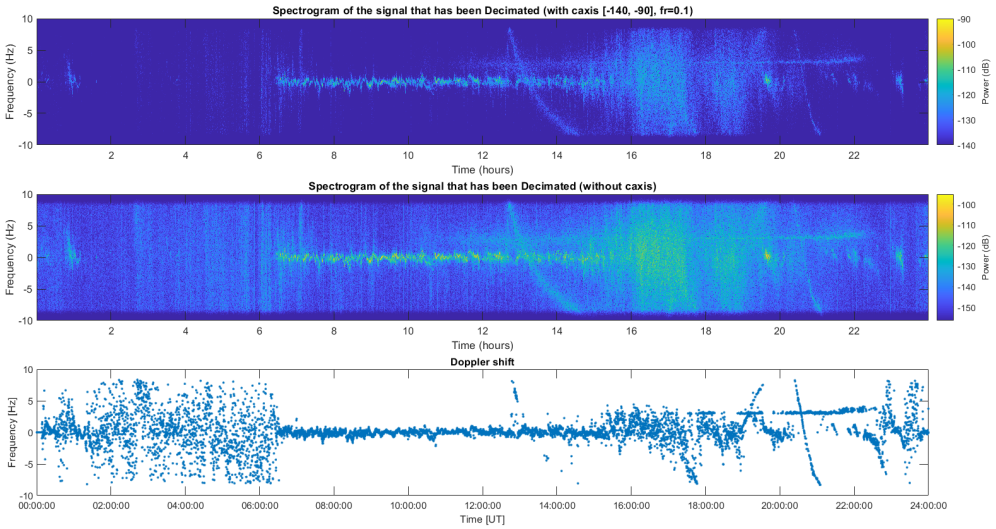

Figure 4.4: A zoom of the 24 hours plots on the 10/03/2021. Fig a (top) is with parameters set, fig b (middle) is without any parameters set, and fig c (bottom) represents the Doppler shift

signal can be spotted over few hours. In order to test the data plotting tools for a 'quiet' day, data from 13th April 2021 was plotted out. On this day, the magnetometers are quiet and no strong signal can be observed in the data. The results are shown in figure 4.6. As for the previous figures, the top panel is with the caxis on ([-140, -90]dB), the middle on is without caxis, and the Doppler shift is shown at the bottom. There are no coherent Doppler signal on this day, and the background is the only 'strong' signal.

Conclusion: final choices of the parameters

After looking at different data sets for diverse conditions (calm day, clear signal, clear signal + background noise, strong background noise), the parameters implemented into the code in order to best visualize the data are:

- Decimation: Even if the difference between the decimation process and the lowpass filtering are small, the decimation allows us to work with less amount of data while obtaining a similar result, and fits with the Resolution Frequency chosen below.
- 'ylim': The y axis limits on the spectrogram will be set to [-10, 10] Hz, in order to highlight the relevant part of the spectral analysis results. This parameters fits for all datasets and ensures a pertinent scale for data visualization.
- 'FrequencyResolution': The frequency resolution had to be set in order to visualize

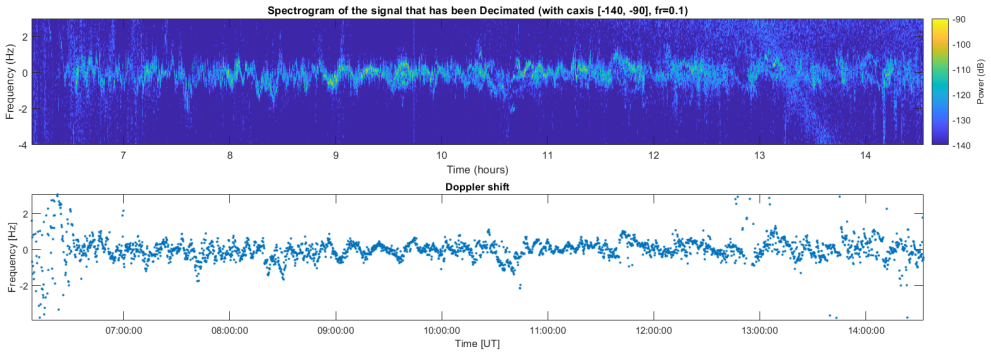

Figure 4.5: Plots from the 10/03/2021, from 6:00 to 14:30 UT. Fig a (top) is with parameters set, fig b (second) is without any parameters set, and fig c (bottom) represents the Doppler shift

the waves and their variations as precisely as possible while respecting the Nyquist Shannon theorem. Different frequencies have been tried, from 0.01 to 0.5 Hz, and the frequency that provides the best results for most of the plots is 'FrequencyResolution'=0.1Hz.

- 'caxis': To allow quick vizualization between data from different time periods, the colorbar scale will remain fixed at caxis=[-140, -90]dB. For most of the plots, the signal of interest is around -90dB, and the background is below -140dB.

To allow a quick visual inspection of the data, 4 common parameter ranges and techniques have been chosen. However, if further inspection is needed, for one particular day, then the scales (e.g. colorscale) and resolution ('FrequencyResolution') would need to be amended. The final routine for waves observations is depicted in schematic 4.7.

4.2 Automization of the algorithm

As the aim of the project is to make a database which will be available to the wider scientific community, the algorithm to visualize the data has to be automatized.

4.2.1 Matlab general code

The code implemented for a single day of data, as described in 4.7, (producing both the spectrogram and the Doppler shift, see section above), is placed into a programming loop to analyse and plot data for a whole month. The code can be found in appendix A.1.

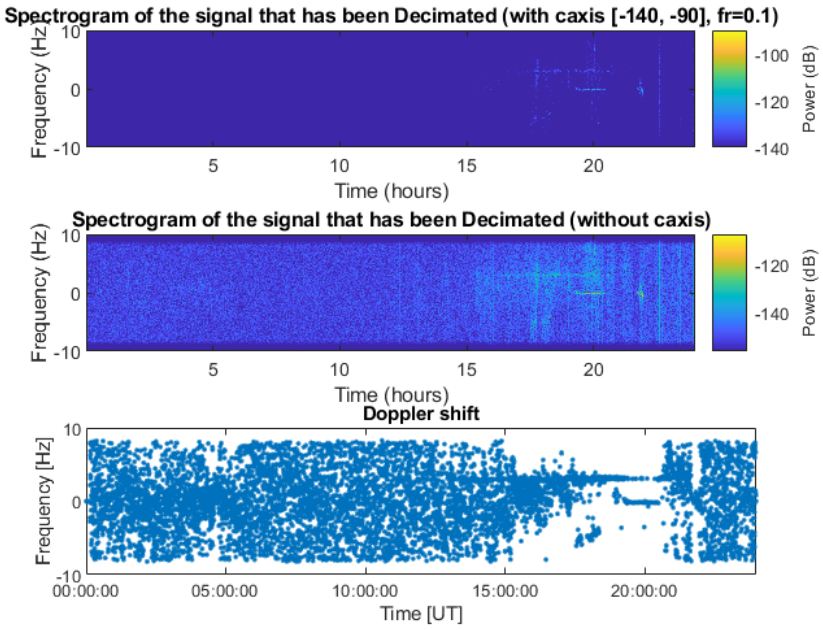

Figure 4.6: 24 hours plots on the 13/04/2021. Fig a (top) is with parameters set, fig b (middle) is without any parameters set, and fig c (bottom) represents the Doppler shift

The signal is mixed with a complex exponential in order to have the component of interest centered, and the strong DC component ready to be filtered. Then, this mixed signal is decimated by a factor of 5 using the decimation function from Matlab. The signal is then lowpass filtered using the filter proposed by Matlab, which is a Chebyshev Type I infinite impulse response (IIR) filter of order 8. (15). At the end of this process, the sampling frequency is of 20Hz. Then, the frequency of the maximal power is extracted to create the Doppler Shift. A temporary file is created and contains 40 seconds of data (which represents 800 points). This process will create the 'window' that will be used to make the PSD. The created tables, 'FreqMax' and 'dBMax' will then be used in order to plot the Doppler shift.
As it can be seen, 4 zeros are added to these two tables: two at the beginning, and two at the end. This comes from the fact that the sliding window requires 40 seconds of data to provide the PSD, and this process can not be applied to the first twenty seconds or the twenty last seconds of each hourly file. As a result, only 356 points are created each hour, as oposed to the required 360 points. A choice has been made to complete these missing points with zeros, as this would not change the dynamic of the final Doppler shift.
An other way to deal with this issue is to include both the file from the hour before, and from the hour after. This is the method chosen by the Japanese team during their own data processing: In order to plot a full day spectrogram, both the day after and before the day of interest have to be loaded. This process induces a much longer time of computation

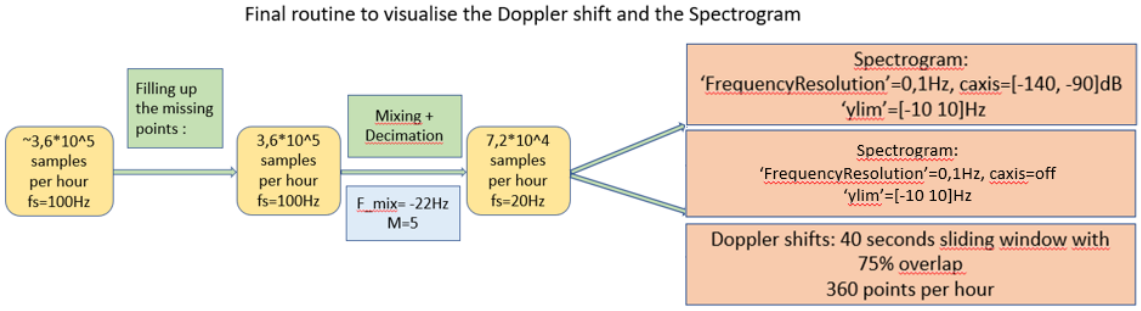

Figure 4.7: Schematic representing the routine to process the data. After dealing with any missing data points, the signal is decimated by a factor of 5 and both the spectrogram and the Doppler shift are plotted using the above, chosen parameters

and has therefore not be chosen in our code at the moment. The figure is then created with both the spectrogram of the 24 hours plot (with and without caxis), and the Doppler shift. Finally, the figure is saved into two formats:

- .jpg : The .jpg (Joint Photographic Experts Group) format will provide a good and simple overview of the dayfile with the parameters chosen above. Plots will be open in a regular picture opener software, and it will be fast and convenient to go across the days. This format will be interesting for a good and fast overview, if no zooming or change in the parameters are needed. The average memory requirement for this format is about 85kb (for comparison, .png requires about 3 times more memory and .mat about 1500 times more).

- .fig: The .fig format is the one representing the Matlab format for figures. It can then be opened again using Matlab, and the figure can be display by using the line below:

```
1                openfig("C:\Users\SFF\Documents\PFE\
  PRIDE_DATA\PRIDE\2021\03\03\doppler_lyr_20210303.fig",'new','
  visible')
2
```

  Once this figure is open again, parameters can be modified and the zoom function can be used.

Moreover, two different 'try'/'catch' test have been implemented:

- The first prevents the code from stopping after a "file not found" error. If the file can not be found, then a "blank" file is created using the regular timestamps for the time axis, and a table of zeros for the signal. Zeros have been chosen istead of NaNs values

because the pspectrum function and the "spectrogram" functions from Matlab can not handle NaNs values. The PSD is therefore processed with the zero table, and this "not found file" full of zeros is added to the others to complete the daily file. On the final plot, this file will be represented by the lowest power allowed by the "caxis" bar, and the points will be removed from the Doppler shift. This method does not have an impact on the rest of the day and allow an easy spot of the missing file.

- The second one is handling the case of an error during the saving process of the file: If the plots can not be saved, the algorithm will move toward the day after while plotting a message explaining that the saving failed, as in can be seen in the code above.

The figure 4.8 provides the final representation of the 10th of March, 2021. One hour of data has been removed in order to show how the algorithm responds to a missing file. As detailed above, the three panels represent the spectrogram, with "caxis" on for the top one and without "caxis" on for the middle one, and the Doppler shift at the bottom. Information about the day,the frequency of the transmitter in Hornsund and the frequency and time resolution of the spectrogram has been added.

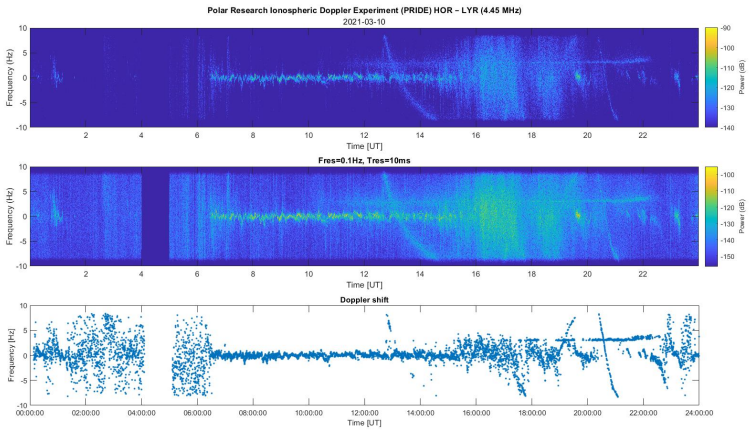

Figure 4.8: Final representation of the data acquired during the 10th of March 2021

Chapter 5

Data exportation

The final part of the project consist of the creation of a website in order to publish the data and to make them available for the scientific community.

5.1 Data and metadata exportation

5.1.1 Example Database - The Nansen Legacy

When it comes to data sharing, the goal is to provide a clear and user-friendly interface for the scientist to download of visualize them. Some instructions are provided by organisations which aim is to promote new knowledge database, such as the Nansen Legacy.
The Nansen Legacy is an arctic research project providing integrated scientific knowledge on the rapidly changing marine climate and ecosystem. This project gathers about 280 researchers, students, and technicians from ten Norwegian research institutions, providing a unique collaborative community and supervision across institutions and disciplines. This gives rise to a new generation of holistic thinking Arctic research leaders (25).

The Nansen Legacy uses a distributed data management system where all datasets are documented with standardized discovery metadata and use metadata (exceptions may occur for some data), governance of data within mandated data centers, but discoverable and accessible through a central hub (25).
In order to help other people to promote their data in the most efficient way, instruction videos are available the a dedicated Youtube channel "Nansen Legacy Data Management" `https://www.youtube.com/channel/UCxzSj9B0UZoL0T7QoybH-LQ/videos`. The Centre follows the FAIR principle for data management, which is detailed below.

FAIR Principle

Acording to the Nansen Legacy, the data publishing should follow the FAIR principle (26):

- Findable: Data should be findable by humans and computers and assigned a persistent identifier.
- Accessible: Data can be freely and openly accessed.
- Interpolable: Data should use standardised terms (common vocabularies) and file structures allowing use in different applications or workflows with minimal manual intervention.
- Reusable: Data are richly and clearly described for humans and computers to understand and have a clear data usage license.

5.1.2 Files available

The files and codes provided to the scientific community are summarized in figure 5.1 in Appendix. "Almost raw" data will be available in addition to summary plots and codes. These data will be in .hdf5 format, and the missing samples will already be processed. Instead of "zeros", "NaNs" values are added. This operation is not conducted for the .mat files as the 'pspectrum' function does not handle "NaNs" values, and requires "zeros" ones.

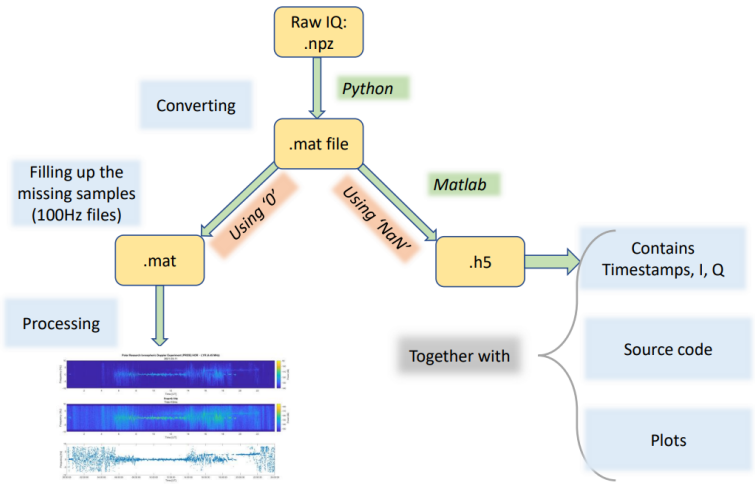

Figure 5.1: Different operations conducted on the file with the format associated. The files provided to the scientific community, after being processed, stand on the right side

5.1.3 Metadata

Metadata can be used to identify, locate and describe a dataset and assist users in finding, organizing, and using information. For this project, the following metadata have been added to the .hdf5 file:

- Name: Longyearbyen
- Institute: University Centre in Svalbard
- Location: Kjell Henriksen Observatory
- Receiver: 78.14798N 16.04235E
- RX Frequency: 4450000-25
- Transmitter: 77.00145N 15.54021E
- TX Frequency: 4.45 MHz

5.2 Github and source code

5.2.1 Github

In order to make the source code available for everyone, for working with both the raw and processed data, Github has been used. The codes from the projects are available from the repository 'PRIDE', from the Github profile created using the pseudonym 'CecilyNoaillac'. Github is a repository hosting service for Git that also has a web-based graphical interface. The main advantages for using Github is that it is a well-known and powerful repository for a project. The service includes access controls as well as a number of collaboration features that might be useful for the next steps of the project. Once the data will be available, all scientists will be able to find the code on Github.
The repository from the project can be accessed from the link:

`https://github.com/CecilyNoaillac/PRIDE`.

Inside this repository, the codes available are the following ones:

- *figure-opening.m* : The aim of this code is to open the .fig file (figure that has already been created but with the visibility on "off"). This code will be useful for someone that has had a look at the summary plot available, and that would like to get deeper into one special day: The Matlab figure has already been created, but if it is loaded using Matlab, it will be possible to have a zoom on it while keeping a good resolution on the picture.

- *Waves-representation-one-day-from-mat.m* : This second code is used to create the 24 hours summary plots. It is takes all the one hour files in .mat format and plots the spectrograms and the Doppler shift. This program can be used if the one hour file plots are available and if the user wants to re-create the plots, having the opportunity to change some parameters.

License

Public repositories on GitHub are often used to share open source software. For a repository to truly be open source, a license is required. The licence Apache 2.0 is a weak software license that has terms to prevent contributors and distributors from suing for patent infringement. Apache 2.0 is therefore recommended for a weak license on a short program. The characteristics of this license can be seen on figure 5.2. This license

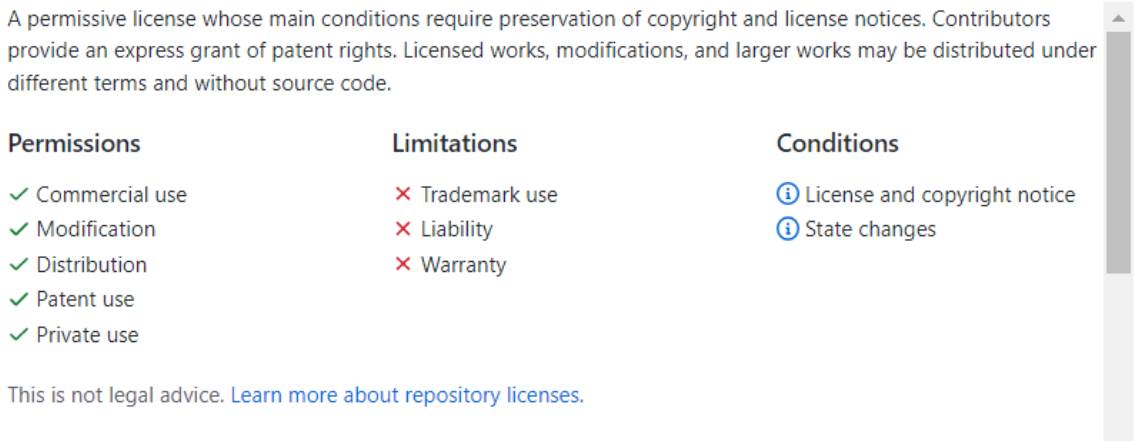

Figure 5.2: Screenshot from Github at the decision of the license. The figure desribes what Apache 2.0 allows and the security it provides

has been successfuly implemented on the whole PRIDE repository on Github on 11th of October 2022, and all the license terms can be find on the following URL: `https://www.apache.org/licenses/LICENSE-2.0`.

5.2.2 KHO website

The summary plots (in .png format) produced as part of this project have been successfully added to the Kjell Henriksen Observatory (KHO) website. They are included in the keyogram overview showing quick look plots from multiple instruments at the KHO. These plots can be found here: `http://kho.unis.no/Keograms/keograms.php?year=2022&month=1`. Figure A.12 in appendix is a screenshot of the Website where the PRIDE data have been added:

Chapter 6

Conclusions

This work investigated the first data acquired during the three first years of the PRIDE HF Doppler system. Using the previous studies conducted at middle and auroral latitudes on similar instruments, a code has been developed in order to prcoess the raw data and create data plots for each day file. Using ground magnetometer data to select a relevant data sets, default parameters have been chosen to create files with two spectrograms and one Doppler shift plot. Issues such as the treatment of missing samples have been tackled and an interpolation method has been investigated for later implementation.
The result of this thesis is the publishing of all the data on the Keogram website from the KHO. The codes are also available on Github and will allow the scientists to get a closer look at the data they are interested in, in order to understand better the ionospheric stuctures in polar regions.

Bibliography

[1] (D. M. Wright, T. K. Yeoman, P. J. Chapman) High-latitude HF Doppler observations of ULF waves. 1. Waves with large spatial scale sizes, *Ann. Geophysicae 15, 1548±1556 (1997)*

[2] (Hiroyuki Nakata, Kenro Nozaki, Yuhei Oki, Keisuke Hosokawa, Kumiko K. Hashimoto, Takashi Kikuchi, Jun Sakai, Ichiro Tomizawa and Satoko Saita) Software-defned radio-based HF doppler receiving system, *Nakata et al. Earth, Planets and Space (2021) 73:209 https://doi.org/10.1186/s40623-021-01547-5*

[3] Crowley, G., and F. S. Rodrigues (2012), Characteristics of traveling ionospheric disturbances observed by the TIDDBIT sounder, Radio Sci., 47, RS0L22, doi:10.1029/2011RS004959.

[4] (L. J. Baddeley, T. K. Yeoman, and D. M. Wright) HF doppler sounder measurements of the ionospheric signatures of small scale ULF waves *Annales Geophysicae, 23, 1807–1820, 2005 SRef-ID: 1432-0576/ag/2005-23-1807*

[5] Davies K, Watts J, Zacharisen D (1962) A study of F2-layer effects as observed with a Doppler technique. J Geophys Res 67:2. https://doi.org/10.1029/JZ067i002p00601

[6] Jacobs JA, Watanabe T (1966) Doppler frequency changes in radio waves propagating through a moving ionosphere. Radio Sci 1(3):257–264

[7] Chum, J., Urbář, J., Laštovička, J. et al. Continuous Doppler sounding of the ionosphere during solar flares. Earth Planets Space 70, 198 (2018). https://doi.org/10.1186/s40623-018-0976-4

[8] Liu, J. Y., Chiu, C. S., and Lin, C. H. (1996), The solar flare radiation responsible for sudden frequency deviation and geomagnetic fluctuation, J. Geophys. Res., 101(A5), 10855– 10862, doi:10.1029/95JA03676.

[9] (D. M. Wright, T. K. Yeoman, P. J. Chapman) High-latitude HF Doppler observations of ULF waves. 1. Waves with large spatial scale sizes *Ann. Geophysicae 15, 1548±1556 (1997) Ó EGS ± Springer-Verlag 1997*

[10] (Hartinger Michael D., Takahashi Kazue, Drozdov Alexander Y., Shi Xueling, Usanova Maria E., Kress Brian) ULF Wave Modeling, Effects, and Applications: Accomplishments, Recent Advances, and Future , Frontiers in Astronomy and Space Sciences, 2022 https://www.frontiersin.org/articles/10.3389/fspas.2022.867394 DOI=10.3389/fspas.2022.867394 ISSN=2296-987X

[11] J.A. Jacobs, K.O. Westphal, Geomagnetic micropulsations, Physics and Chemistry of the Earth, Volume 5, 1964, Pages 157-224,ISSN 0079-1946, https://doi.org/10.1016/S0079-1946(64)80005-7. (https://www.sciencedirect.com/science/article/pii/S0079194664800057)

[12] David Anderson,Adela Anghel,Kiyohumi Yumoto,Mutsumi Ishitsuka,Erhan Kudeki, Estimating daytime vertical ExB drift velocities in the equatorial F-region using ground-based magnetometer observations https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2001GL014562

[13] Research in Svalbard Portal, RiS ID 11522, Polar Research Ionospheric Doppler Experiment (PRIDE) https://www.researchinsvalbard.no/project/20000000-0000-0000-0000-000000009571/project-info

[14] Mikko Syrjäsuo, Lisa Baddeley, Radar hardware - mixer and oscillator, AGF 304 Fieldwork (Master of Geophysics), UNIS Svalbard

[15] Matlab Help Center about the decimation function: *https://fr.mathworks.com/help/signal/ref/decimate.html*

[16] Matlab Help Center about the downsampling function: *https://fr.mathworks.com/help/signal/ref/downsample.html*

[17] (Chris Chatfield) The Analysis of Time Series, An Introduction. Sixth Edition. Chapter 7, "spectral Analysis", pages 121-155. *Chapman Hall/CRC, Texts in Statistical Science Series*

[18] Physik-Institut, Universität Zurich: *https://www.physik.uzh.ch/local/teaching/SPI301/LV-2015-Help/lvanlsconcepts.chm*

[19] ON-DEMAND WEBINAR: Fundamentals of Digital Signal Processing, Siemens *https://webinars.sw.siemens.com/en-US/digital-signal-processing/*

[20] Matlab Help Center about the pspectrum function: *https://fr.mathworks.com/help/signal/ref/pspectrum.html*

[21] Ricardo Guitierrez-Osuna: Introduction to speech processing. L6: Short time Fourier Analysis and synthesis *https://www.yumpu.com/en/document/read/19303059/l6-short-time-fourier-analysis-and-synthesis*

[22] Anke Meyer-Baese, Volker Schmid, Chapter 3 - Subband Coding and Wavelet Transform, Editor(s): Anke Meyer-Baese, Volker Schmid, Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition), Academic Press, 2014, Pages 71-111, ISBN 9780124095458

[23] (S. W. Smith) The Scientist and Engineer's and Guide to Digital Signal Processing *California Technical Publishing, 1997. ISBN 0-9660176-3-3.* ,`https://dspguru.com/dsp/faqs/multirate/decimation/`

[24] International Monitor for Auroral Geomagnetic Effects, *https://space.fmi.fi/image/www/index.php?page=home*

[25] The Nansen Legacy: Arctic research project providing integrated scientific knowledge on the rapidly changing marine climate and ecosystem. *https://arvenetternansen.com/about-us/*

[26] How to publish FAIR Nansen Legacy datasets, A complete step by step guide. The Nansen Legacy, Luke Marsden. *https://drive.google.com/file/d/1skC5oNYFyv5jkfG4lVlVOGANbstiSTIe/view*

[27] Choose an open source license, from Github: `https://choosealicense.com/`

[28] How to Choose a License for Your Own Work, GNU Operating System: `https://www.gnu.org/licenses/license-recommendations.en.html`

Appendix A

Appendix

	Description	Lowpass filtering	Decimation
04/01/2021	The background is noisy and contains a power line in the back. Signal can be highlighted thanks to the caxis	'FreqResolution'=0.1 'caxis' =[-140, -90]	'FreqResolution'=0.1 'caxis' =[-140, -90]
09/01/2021	The background is noisy and contains a power line in the back. Signal can be and highlighted thanks to the caxis	'FreqResolution'=0.1 'caxis' =[-140, -90]	'FreqResolution'=0.1 'caxis' =[-140, -90]]
03/02/2021	1.5 hours or clear waves at the beggining of the day Can be clearly identified with caxis on and the chosen parameters	'FreqResolution'=0.1 'caxis'=[-140, -90]	'FreqResolution'=0.1 'caxis'=[-140, -90]
06/02/2021	Some shifts but the signal is weak and caxis has to be smaller to highlight the signal. Both methods are providing the same results.	'FreqResolution'=0.2 'caxis' =[-140, -110]	'FreqResolution'=0.1 'caxis' : off or [-140, -110]
10/02/2021	Two hours of bright and clear shift can be seen. Setting parameters helps to improve the visualization. Decimation offers a clearer signal	'FreqResolution'=0.05 'caxis' : off or [-140, -90]	'FreqResolution'=0.05 'caxis' : off pr [-140, -90]
13/02/2021	Waves can be easily spotted and identified (midnight to 2 AM) and are highlighted by the caxis on	'FreqResolution'=0.1 'caxis'=[-140, -90]	'FreqResolution'=0.1 'caxis'=[-140, -90]
03/03/2021	Some shift can be observed, but the signal appears better without "caxis" on	'FreqResolution'=0.05 'caxis': off or [-140, -90]	'FreqResolution'=0.05 'caxis': off or [-140, -9]
10/03/2021	Clear and long signal can be observed, as well as a strong background noise and an long perturbation signal	'FreqResolution'=0.1 'caxis'=[-140, -90]	'FreqResolution'=0.1 'caxis'=[-140, -90]
11/01/2022	Some coherent signals can be observed. The colorbar is highlighting the shifts	'FreqResolution'=0.1 'caxis'=[-140, -100]	'FreqResolution'=0.1 'caxis'=[-140, -100]
16/01/2022	The signal is weak and difficult to see with a strong 'caxis' limit. Impossible to see some waves.	'FreqResolution'=0.05 'caxis'=[-170, -110]	'FreqResolution'=0.1 'caxis'=[-170, -110]
24/03/2021	Calm day, with some signal around 8pm	'FreqResolution'=0.1 'caxis'=[-150, -90]	'FreqResolution'=0.1 'caxis'=[-150, -90]
13/04/2021	Almost calm day	'FreqResolution'=0.1 'caxis'=[-140, -90]	'FreqResolution'=0.1 'caxis'=[-150, -90]

Table A.1: Table representing the observations and the parameters for data observation of 12 days of interest

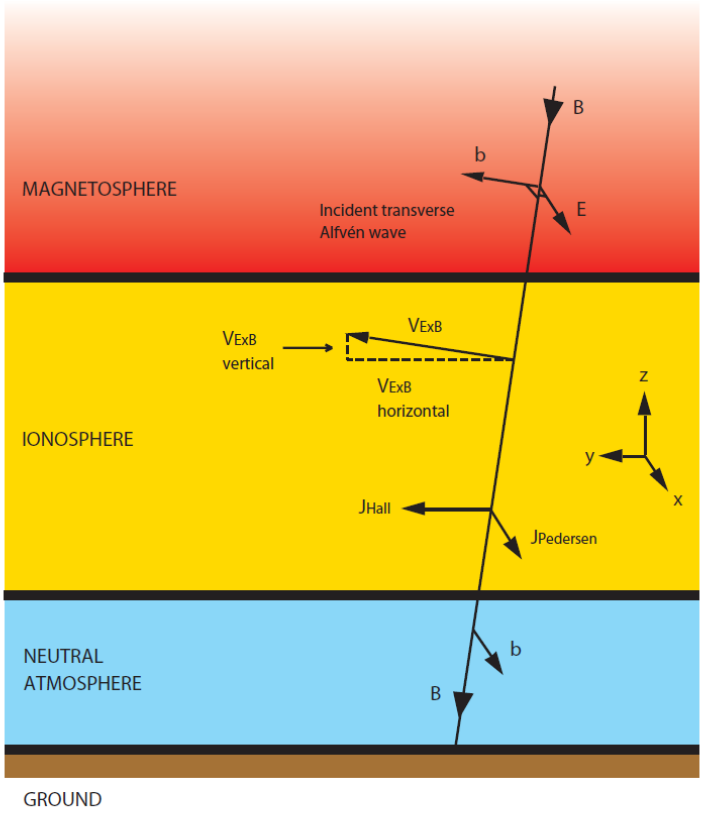

Figure 4.4. Schematic indicating the vertical and horizontal components of the **E**∧**B** drift in the ionosphere which arises due to the fact that the magnetic field lines are not completely orthogonal to the ground

Figure A.1: Representation of the $V_{E\times B}$ drift in the ionosphere

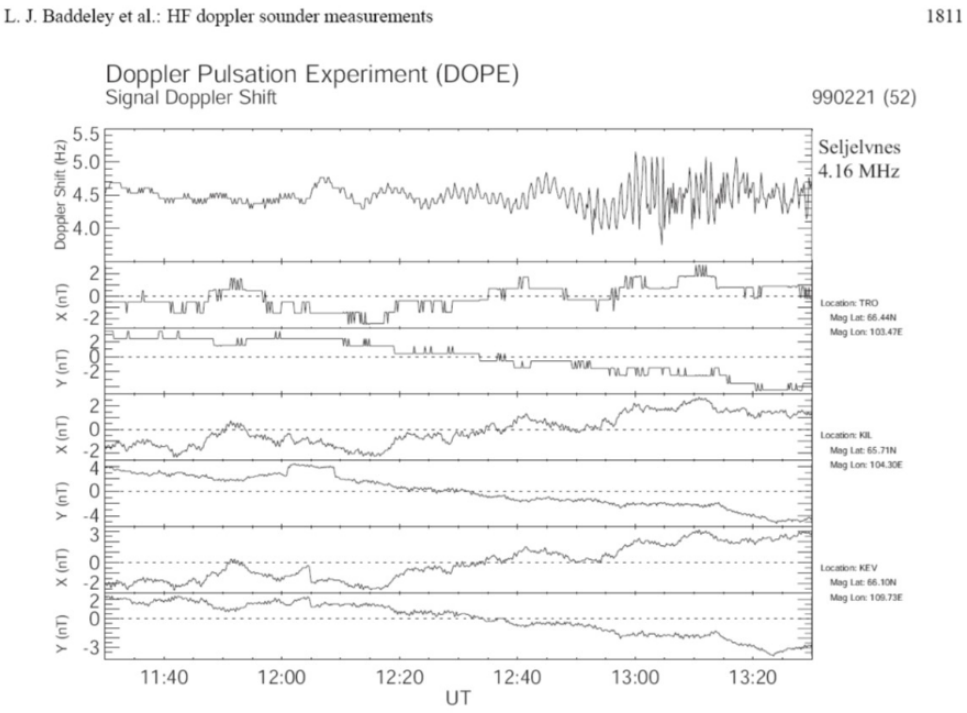

Fig. 3. A DOPE trace (Seljelvnes trace from Fig. 2d) and also magnetometer data from the IMAGE network. This wave event is classified as uncorrelated as there is no magnetic ground signature.

Figure A.2: Uncorrelated event, plots taken from Baddeley, L.J., T. K. Yeoman and D. M. Wright, HF doppler sounds measurements of the ionospheric signatures of small scale ULF waves (2005), Ann. Geophys., 23, 1807-1820, 2005 (4)

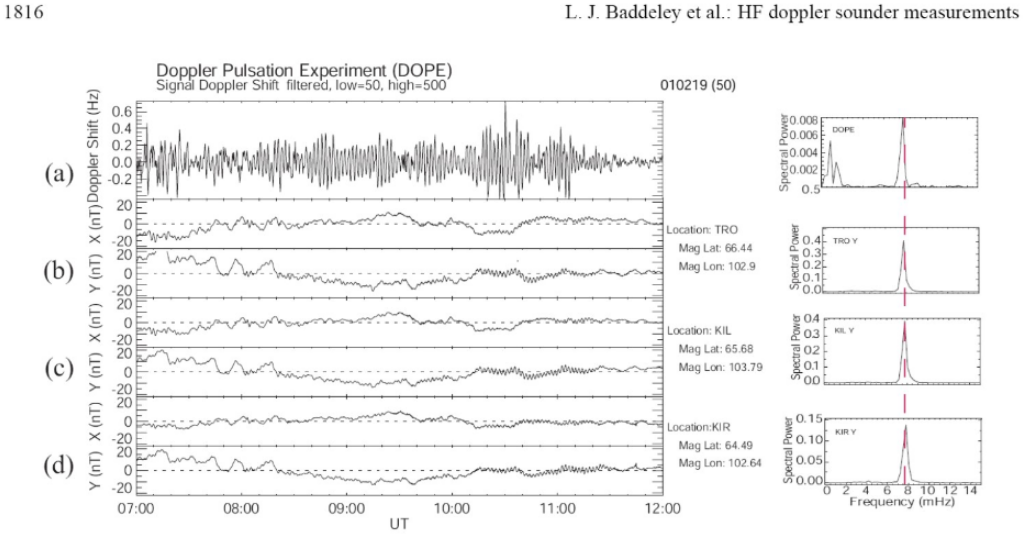

Fig. 8. (**a**) indicates a Doppler trace from the DOPE HF sounder for the event on the 19th February 2001. (**b**), (**c**) and (**d**) indicate magnetometer data from the TRO, KIL and KIR stations respectively. Panel (a) indicates that the wave was observed by DOPE between 08:00 and 11:40 UT (10:00–13:40 MLT) and panels (b)–(d) indicate the wave observed by the magnetometers between 10:20 and 11:40 UT (12:20 to 13:40 MLT). The panels to the right of the magnetometer and DOPE traces show the result of a FFT analysis undertaken of the DOPE data from 08:00–11:40 UT and the Y component of the magnetometer data from 10:20–11:40 UT. The dominant frequency component is at ~8 mHz for all the traces.

Figure A.3: Correlated event, plots taken from Baddeley, L.J., T. K. Yeoman and D. M. Wright, HF doppler sounds measurements of the ionospheric signatures of small scale ULF waves (2005), Ann. Geophys., 23, 1807-1820, 2005 (4)

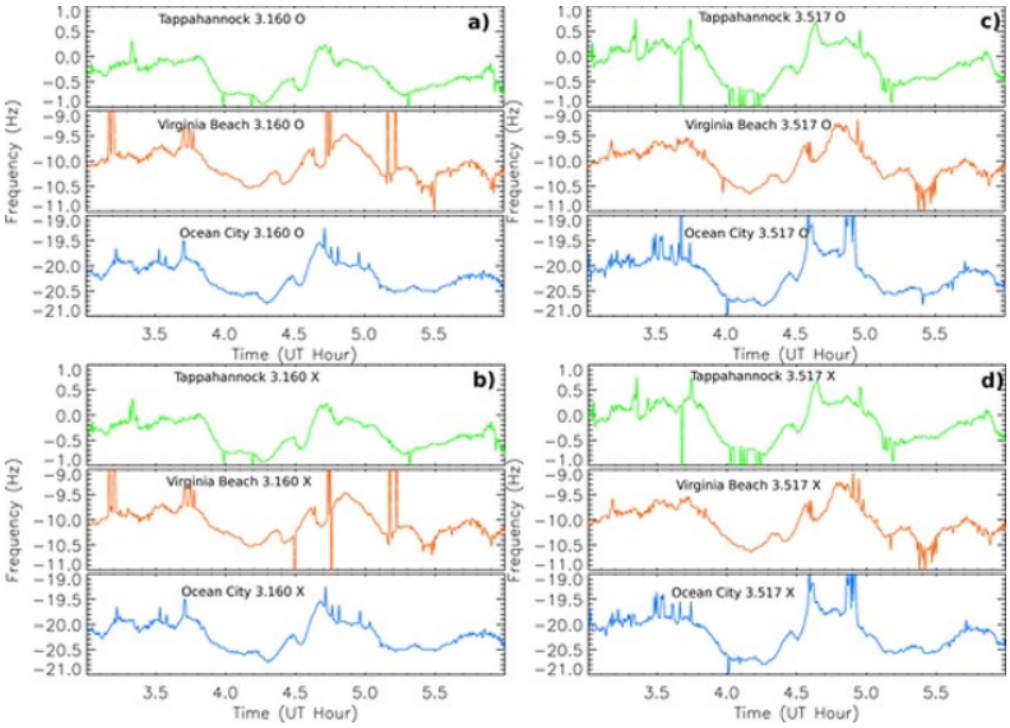

Figure A.4: Doppler time series of X and O mode traces measured at 3 different locations, plots taken from Crowley, G., and F. S. Rodrigues (2012), Characteristics of traveling ionospheric disturbances observed by the TIDDBIT sounder, Radio Sci., 47, RS0L22, doi:10.1029/2011RS004959. (3)

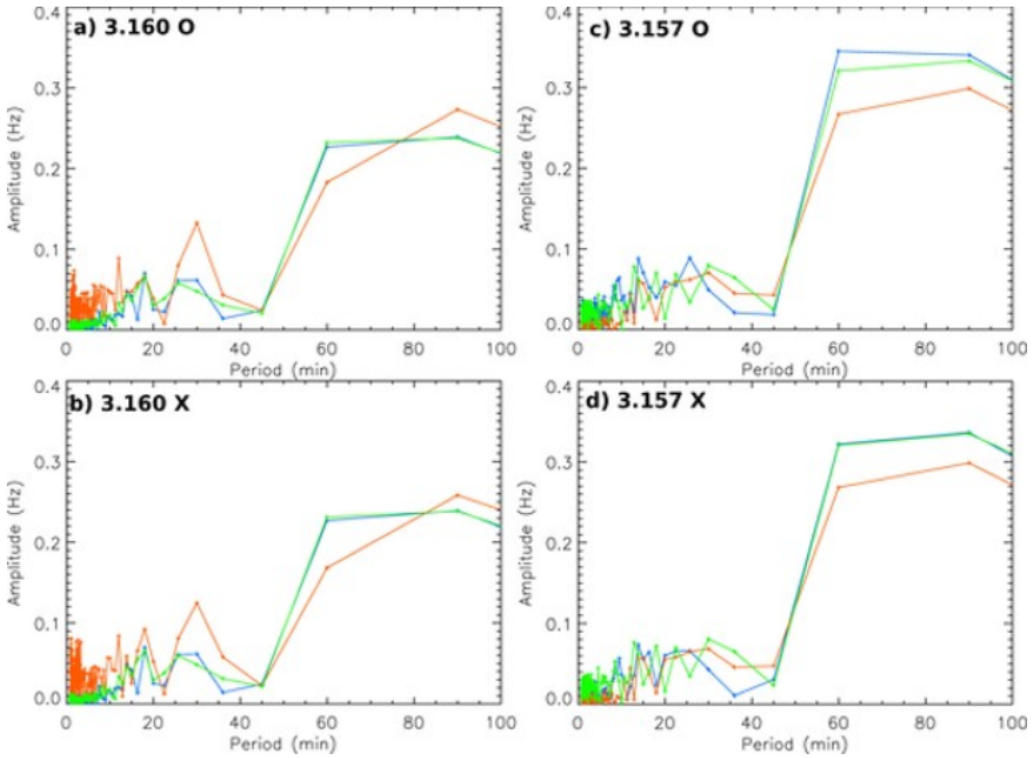

Figure A.5: FFT amplitude vs period for 3-6 UT (The colour codes are the same as for figure above), plots taken from Crowley, G., and F. S. Rodrigues (2012), Characteristics of traveling ionospheric disturbances observed by the TIDDBIT sounder, Radio Sci., 47, RS0L22, doi:10.1029/2011RS004959. (3)

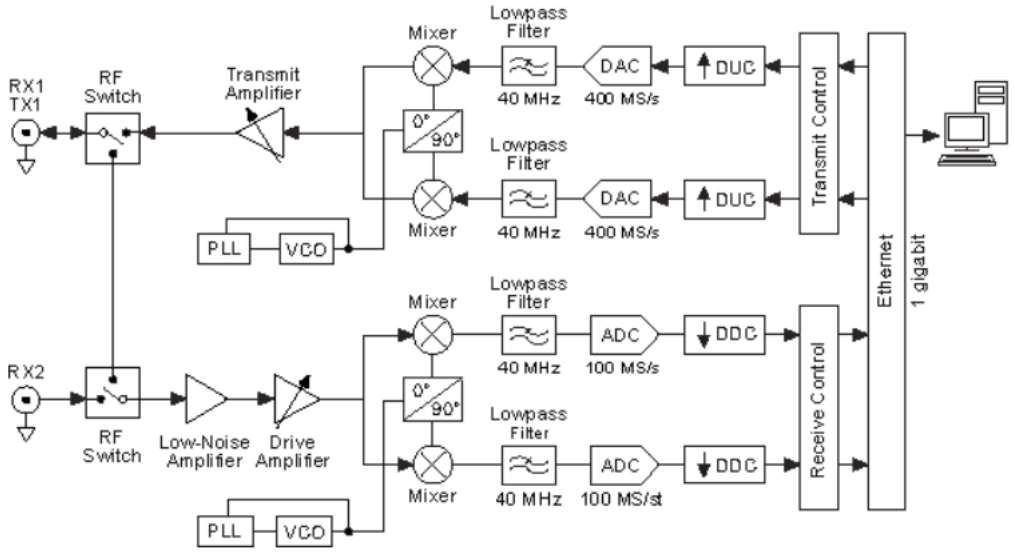

Figure A.6: General USRP Architecture, credits: Ettus

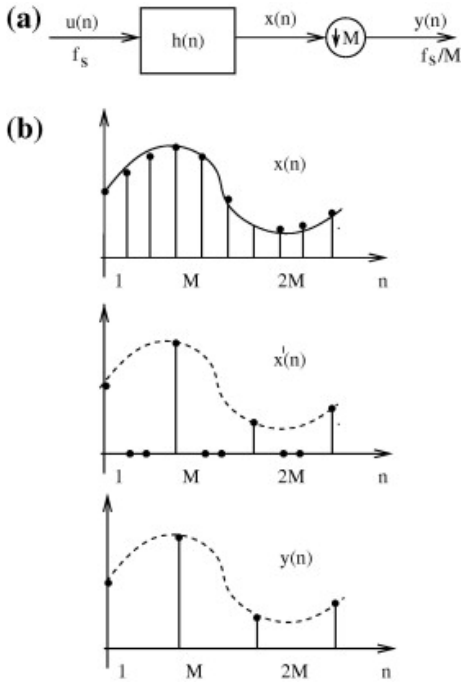

Figure A.7: Representation of decimation technique: (a) filter and downsampler, (b) typical time sequences of the intermediate signals. Credits: Anke Meyer-Baese, Volker Schmid, in Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition), 2014

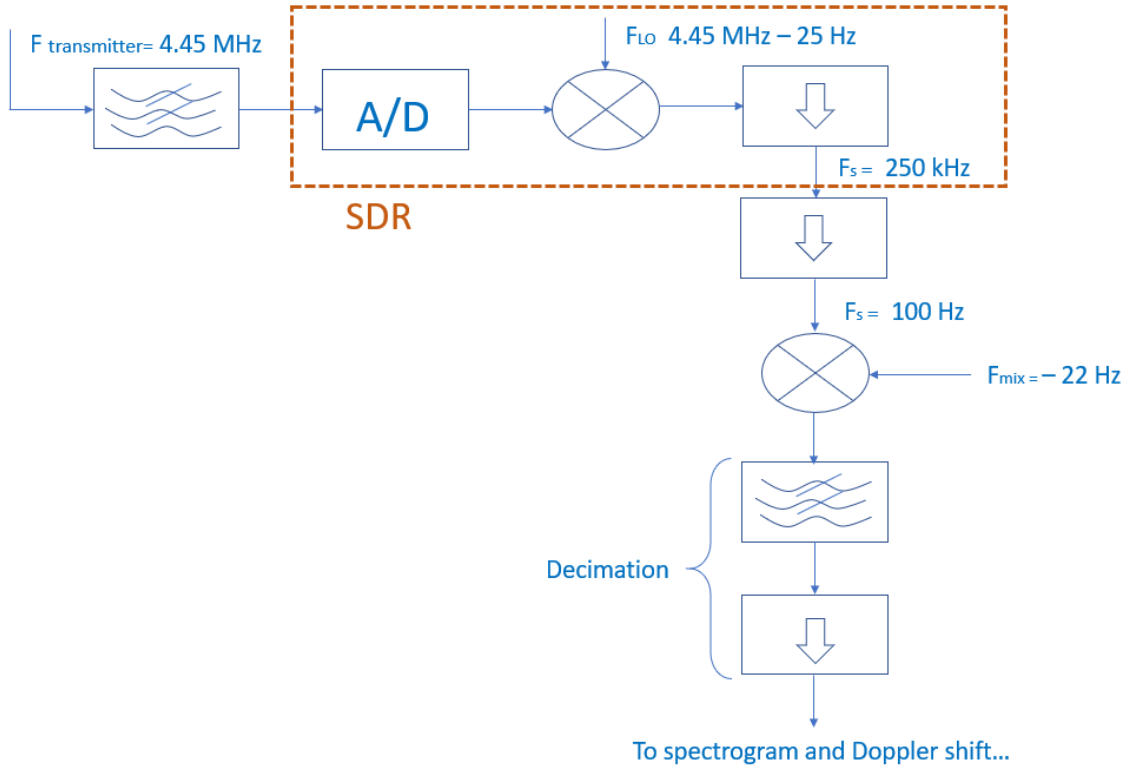

Figure A.8: Block diagram of the PRIDE processing operations

Method	Description	Continuity	Comments
`'linear'`	Linear interpolation. The interpolated value at a query point is based on linear interpolation of the values at neighboring grid points in each respective dimension. This is the default interpolation method.	C^0	• Requires at least 2 points • Requires more memory and computation time than nearest neighbor
`'nearest'`	Nearest neighbor interpolation. The interpolated value at a query point is the value at the nearest sample grid point.	Discontinuous	• Requires at least 2 points • Modest memory requirements • Fastest computation time
`'next'`	Next neighbor interpolation. The interpolated value at a query point is the value at the next sample grid point.	Discontinuous	• Requires at least 2 points • Same memory requirements and computation time as `'nearest'`
`'previous'`	Previous neighbor interpolation. The interpolated value at a query point is the value at the previous sample grid point.	Discontinuous	• Requires at least 2 points • Same memory requirements and computation time as `'nearest'`
`'pchip'`	Shape-preserving piecewise cubic interpolation. The interpolated value at a query point is based on a shape-preserving piecewise cubic interpolation of the values at neighboring grid points.	C^1	• Requires at least 4 points • Requires more memory and computation time than `'linear'`
`'cubic'`	Cubic convolution used in MATLAB® 5.	C^1	• Requires at least 3 points • Points must be uniformly spaced • This method falls back to `'spline'` interpolation for irregularly-spaced data • Similar memory requirements and computation time as `'pchip'`
`'v5cubic'`	Same as `'cubic'`.	C^1	
`'makima'`	Modified Akima cubic Hermite interpolation. The interpolated value at a query point is based on a piecewise function of polynomials with degree at most three. The Akima formula is modified to avoid overshoots.	C^1	• Requires at least 2 points • Produces fewer undulations than `'spline'`, but does not flatten as aggressively as `'pchip'` • Computation is more expensive than `'pchip'`, but typically less than `'spline'` • Memory requirements are similar to those of `'spline'`
`'spline'`	Spline interpolation using not-a-knot end conditions. The interpolated value at a query point is based on a cubic interpolation of the values at neighboring grid points in each respective dimension.	C^2	• Requires at least 4 points • Requires more memory and computation time than `'pchip'`

Figure A.9: Summary and utilisation of the main interpolation functions on Matlab (Matlab documentation)

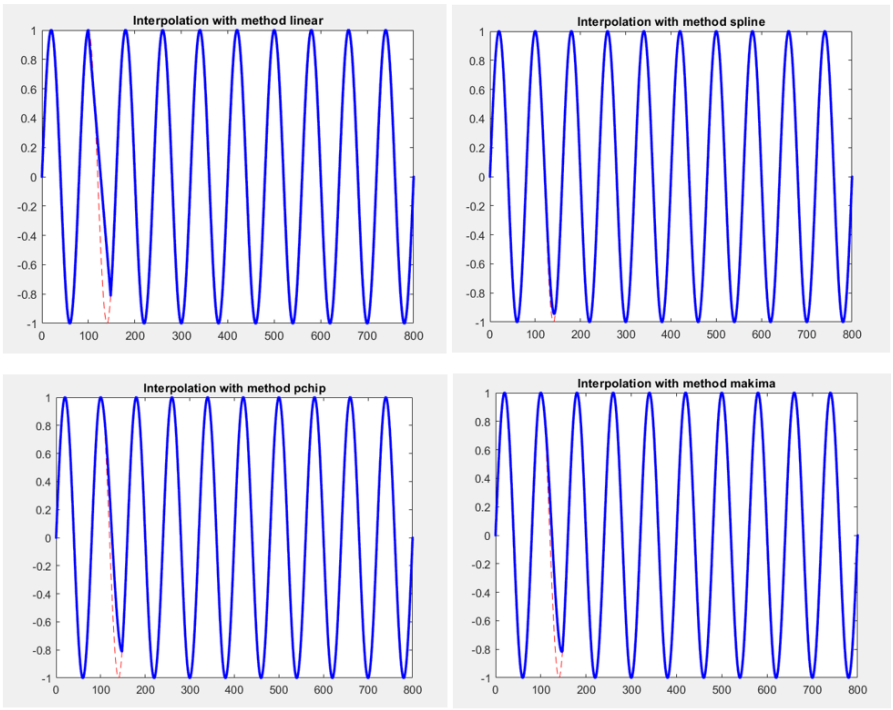

Figure A.10: Different interpolation techniques with a gap of 6%

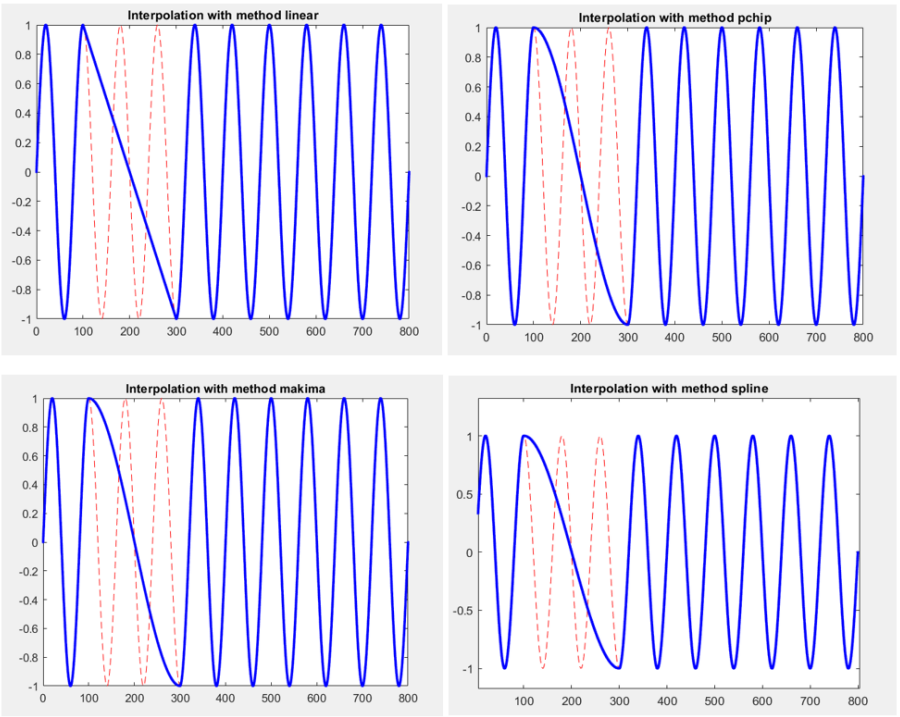

Figure A.11: Different interpolation techniques with a gap of 25%

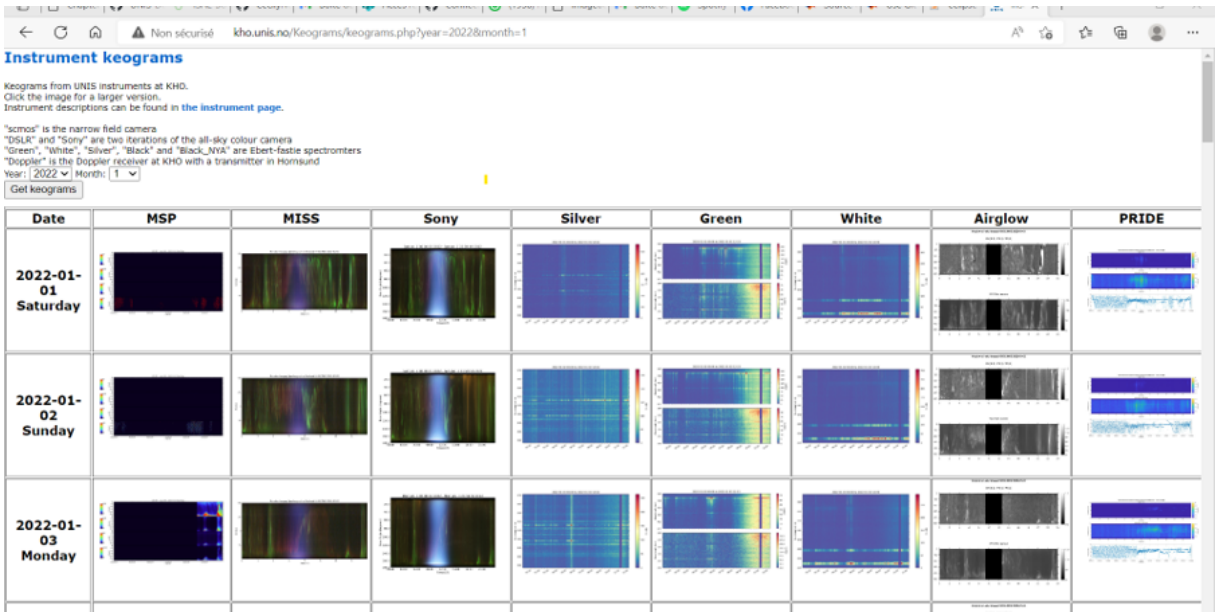

Figure A.12: Keograms from UNIS instruments at KHO. PRIDE plots are on the right side

A.1 Observations of Waves in the Ionosphere

```
theorical_lenght=size(time_frac,2);
lenght_of_signal=size(time_frac_test,2);

missing_points=[];
amount_of_missing_points=[];
if lenght_of_signal~=theorical_lenght
    for i= 1:(lenght_of_signal-1)
        if (time_frac_test(i+1)-time_frac_test(i))>0.0000040001
            m=floor(round((time_frac_test(i+1)-
            time_frac_test(i))*fs-1));
            missing_points=[missing_points i];
            amount_of_missing_points=[amount_of_missing_points m];
        end
    end
      for i= size(missing_points,2):-1:1
          i
        for k=1:amount_of_missing_points(i)
            time_frac_test=[time_frac_test(1:missing_points(i)+k-1)
   time_frac_test(missing_points(i)-1)+(k+1)*(1/fs) time_frac_test(
   missing_points(i)+k:end)];
            sig_1_frac_test=[sig_1_frac_test(1:missing_points(i)+k-1) 0
   sig_1_frac_test(missing_points(i)+k:end)];
            lenght_of_signal=size(time_frac_test,2)
         end
      end
end

size_final=size(time_frac_test,2);
if size_final~=theorical_lenght
    eps=theorical_lenght-size_final;
    while eps>0
        time_frac_test=[time_frac_test time_frac_test(end)+1/fs];
        sig_1_frac_test=[sig_1_frac_test 0];
        eps=eps-1;
    end
end
```

```
year=2021
month=2
day=3

datadir=str(year)+"/0"+str(month)+"/0"+str(day)+"/"
for i in range (0,24):
    if i<10:
        data = load('C:\\Users\\SFF\\Documents\\PFE\\PRIDE_DATA\\PRIDE\\'
     +datadir+'doppler_lyr_20210203_0'+str(i)+'UT.npz')
        lst = data.files
```

```
        print(np.size(data[lst[0]]))
        print(data[lst[0]])
        print(data[lst[1]])
        timestamps= data[lst[0]]
        iq=data[lst[1]]
        savemat('C:\\Users\\SFF\\Documents\\PFE\\PRIDE_DATA\\PRIDE\\'+
        datadir+'doppler_lyr_20210203_0'+str(i)+'UT.mat',
        mdict={'timestamps':data['timestamps'],'iq':data['iq']})
    else:
        data = load('C:\\Users\\SFF\\Documents\\PFE\\PRIDE_DATA\\PRIDE\\'
     +datadir+'doppler_lyr_20210203_'+str(i)+'UT.npz')
        lst = data.files
        print(np.size(data[lst[0]]))
        print(data[lst[0]])
        print(data[lst[1]])
        timestamps= data[lst[0]]
        iq=data[lst[1]]
        savemat('C:\\Users\\SFF\\Documents\\PFE\\PRIDE_DATA\\PRIDE\\'+
        datadir+'doppler_lyr_20210203_'+str(i)+'UT.mat',
        mdict={'timestamps':data['timestamps'],'iq':data['iq']})
```

```
for month=1:12
    if month==1 | month==3 | month==5 | month==7 | month==8 | month==10 |
    month==12
        days_month=31;
    ...
    for day=1:days_month
        if day<10
            day_l='0'+string(day);
        else
         ...
        for i=0:23
        ...
            f_LO=-22; % Hz (from the plot)
            y_LO=exp(1j*2*pi*f_LO*timestamps);
            mix=iq.*y_LO;
            bb=decimate(mix,5);
     %Collecting the PSD and max

            sample=800;
            for i=1:356
                temp=bb(i*200+1:i*200+sample);
                [p,f] = pspectrum(temp,20);
                [M,I]=max(p);
                LEMAXDUMAX=10*log10(p(I));
                FREQUENCEMAXDUMAX=f(I);
                FreqMax=[FreqMax, FREQUENCEMAXDUMAX];
                dBMax=[dBMax, LEMAXDUMAX];
            end
```

```
28              FreqMax=[0, 0, FreqMax, 0, 0];
29              dBMax=[0, 0, dBMax, 0, 0];
30              % Plotting the spectrogram and the Doppler shift
31          fig=figure('Visible',"off");
32
33
34          % Spectrogram
35          ax1 = subplot(3,1,1);
36          pspectrum(iq_decimate,fs/5,'spectrogram','FrequencyResolution'
    ,0.1,'OverlapPercent',75)
37          ylim([-10 10])
38          title('Spectrogram of the signal that has been Decimated (with
    caxis [-140, -90], fr=0.1)')
39          %colorbar(ax1,'off')
40          caxis([-140 -90])
41
42
43          % Doppler shift
44          ax3 = subplot(3,1,2);
45          ax3.Position(3)=0.7368;
46          FinalTime=[1:1:360*24]/6;
47          FinalTimeHours=duration(0,FinalTime,0);
48          plot(FinalTimeHours,FreqMax,'.')
49          title('Doppler shift');
50          xlabel('Time [UT]')
51          ylabel('Frequency [Hz]')
52          xlim([0 FinalTimeHours(end)])
53          try
54              saveas(gcf,'C:\Users\SFF\Documents\PFE\PRIDE_DATA\PRIDE\'+
    year_l+'\'+month_l+'\'+day_l+'\doppler_lyr_'+year_l+month_l+day_l+'.
    jpg');
55              saveas(gcf,'C:\Users\SFF\Documents\PFE\PRIDE_DATA\PRIDE\'+
    year_l'+'\'+month_l+'\'+day_l+'\doppler_lyr_'+year_l+month_l+day_l+'.
    fig');
56              saveas(gcf,'C:\Users\SFF\Documents\PFE\PRIDE_DATA\PRIDE\'+
    year_l+'\'+month_l+'\Whole_month'+'\doppler_lyr_'+year_l+month_l+
    day_l+'.png');
57          catch
58              disp('Error: The file has not been saved')
59          end
```

Résumé —

Mots clés :

ISAE
10, avenue Édouard Belin
BP 54032
31055 Toulouse CEDEX 4