

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (государственный университет)  
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ  
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМ. А. А. ДОРОДНИЦЫНА РАН  
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»

Желавская Ирина Сергеевна

**Специальные методы обработки двумерных данных  
об электронной плотности плазмы**

010656 — Математические и информационные технологии

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

**Научные руководители:**

д. ф.-м. н. Юрий Шприц

д. ф.-м. н. Мария Спасоевич

Москва

2015

MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION  
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY (state university)  
DEPARTMENT OF APPLIED MATHEMATICS AND CONTROL  
DORODNICYN COMPUTING CENTRE, RAS (CC RAS)  
SPECIALIZED DEPARTMENT «INTELLIGENT SYSTEMS»

Irina Zhelavskaya

**Specialized methods of two-dimensional electron  
plasma density data processing**

010656 — Mathematics and information technology

MASTER'S GRADUATE QUALIFICATION WORK

**Thesis advisors:**

Yuri Shprits, PhD

Maria Spasojevic, PhD

Moscow

2015

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Objectives and Motivation . . . . .	5
1.2	Outline . . . . .	5
<b>2</b>	<b>Relevant Research</b>	<b>7</b>
2.1	Plasma Density . . . . .	7
2.1.1	Basic Definitions . . . . .	7
2.1.2	Application to Wave Particle Simulations . . . . .	8
2.2	The Van Allen Probes Mission . . . . .	8
2.2.1	Mission Brief Overview and Science Objectives . . . . .	8
2.2.2	Spacecraft Brief Overview . . . . .	9
2.2.3	The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) Instrumentation Suite . . . . .	11
2.2.4	The Fluxgate Magnetometer (MAG) . . . . .	13
2.2.5	The Waves Instrument . . . . .	14
2.3	Determining the Electron Density from Upper-Hybrid Band Resonance Fre- quency . . . . .	15
2.4	AURA – An Automated Upper-hybrid Resonance Detection Algorithm . .	16
<b>3</b>	<b>Nonlinear Regression with Neural Networks</b>	<b>19</b>
3.1	Problem Statement . . . . .	19
3.2	Neural Networks for Regression . . . . .	20
3.3	Feedforward Neural Networks . . . . .	20
<b>4</b>	<b>Data</b>	<b>22</b>
4.1	Van Allen Probes Data Processing . . . . .	22
4.2	L2 Data: Van Allen Probes Plasma Wave Observations . . . . .	23

4.3	L4 Data: Electron Densities . . . . .	24
4.4	Data Examples . . . . .	25
<b>5</b>	<b>Methodology</b>	<b>27</b>
5.1	Problem Statement . . . . .	27
5.2	Design Matrix Description . . . . .	28
5.3	Neural Network Design and Implementation . . . . .	29
5.3.1	Neural Network Design Workflow . . . . .	29
5.3.2	Data Cleaning . . . . .	30
5.3.3	Data Split . . . . .	30
5.3.4	Data Preprocessing . . . . .	31
5.3.5	Training Routine . . . . .	32
5.3.6	Validation . . . . .	33
5.3.7	Neural Network Usage . . . . .	34
<b>6</b>	<b>Results</b>	<b>36</b>
6.1	Statistical Results . . . . .	36
6.2	Comparison with AURA . . . . .	37
6.3	Results Evaluation . . . . .	38
6.4	Application . . . . .	42
<b>7</b>	<b>Conclusions and Further Work</b>	<b>43</b>
7.1	Conclusions . . . . .	43
7.2	Further Work . . . . .	43
<b>A</b>	<b>Data Types</b>	<b>44</b>
	<b>Bibliography</b>	<b>50</b>

## Abstract

The electron plasma density,  $n_e$ , is a fundamental parameter of plasma, a critical parameter for wave particle simulations and key to understanding the role of plasma waves in the formation and decay of the Earth's radiation belts. It is a crucial parameter that is required for the prediction of the evolution of the space environment and is also important for the analysis of satellite anomalies that have occurred in the past. In August 2012 NASA (The National Aeronautics and Space Administration) launched two identical spacecraft, namely the Van Allen probes, to provide the most detailed picture of the Earth's radiation belts ever seen. The information collected from these probes allows better protection of satellites and understanding of how space weather affects communications and technology on Earth.

There are numerous existing techniques to measure electron plasma density either directly or indirectly. Determining the electron density from intense upper-hybrid band emissions is considered to be the most reliable passive technique. The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) on board the Van Allen Probes mission makes routine electric field measurements in the frequency range of 10 to 487 kHz in order to identify the frequency of the upper hybrid resonance band, thus providing an accurate estimation of the electron density. In previous missions, the plasma resonance bands were manually identified, and there have been few attempts to do robust, routine automated detection.

This thesis presents an algorithm for automatic inference of the electron number density from plasma wave measurement made onboard the Van Allen Probes mission. It accomplishes this by using feedforward neural networks to infer upper hybrid resonance frequency  $f_{uhr}$ , used then to determine the electron number density. The design and implementation of the algorithm are described, as well as the results of the inference. The comparison with the Automated Upper-hybrid Resonance Detection Algorithm (AURA) [1] is also presented.

# Chapter 1

## Introduction

### 1.1 Objectives and Motivation

The goal of this research is to design an algorithm for automatic determination of the electron number density from plasma wave measurement made onboard the Van Allen Probes mission. This NASA mission is a part of the Living With a Star program and it aims to provide the most detailed picture of the Earth's radiation belts ever seen [8]. The radiation belts are donut-shaped belts around the Earth that consist of very highly energised particles. Amounts of radiation in these belts can be very damaging to anything exposed to it (satellites, etc.). The information collected from the Van Allen probes allows better protection of satellites and understanding of how space weather affects communications and technology on Earth.

Electron density (plasma density) is the fundamental parameter of plasma that is required for the prediction of the evolution of the space environment and is also important for the analysis of satellite anomalies that have occurred in the past. In all previous missions, it was identified manually, and this is a very hard and time-consuming process that soon will be impossible to handle due to the growth of available data. That is why the possibility to obtain plasma densities values automatically is essential for many future simulations of the Earth surroundings.

### 1.2 Outline

The rest of this thesis is organized as follows. In Chapter 2 we present the background material about the Van Allen Probes mission, plasma density, technique employed in this

work to derive electron densities and the Automated Upper-hybrid Resonance Detection Algorithm (AURA). In Chapter 3 we describe neural networks apparatus used to solve the problem at hand. Chapter 4 contains a thorough description of the data obtained from the spacecraft and used in this work. Chapter 5 describes the methodology and the workflow of neural network design and implementation. Finally, we present results of the computational experiments in Chapter 6 and conclude and discuss further work in Chapter 7.

# Chapter 2

## Relevant Research

### 2.1 Plasma Density

This section gives the basic definitions and background required for understanding what plasma density is and why this parameters is so important for the scientific community.

#### 2.1.1 Basic Definitions

**Plasma.** Plasma is one of the four fundamental states of matter, the others being solid, liquid, and gas. A plasma has properties unlike those of the other states. Plasma is a partially ionized gas containing an equal number of positive and negative charges, as well as some other number of none ionized gas particles [4].

**Electrical conductivity of plasma.** The presence of a significant number of charge carriers makes plasma electrically conductive so that it responds strongly to electromagnetic fields. Like gas, plasma does not have a definite shape or a definite volume unless enclosed in a container. Unlike gas, under the influence of a magnetic field, it may form structures such as filaments, beams and double layers [4].

**Plasma density.** The term "plasma density" by itself usually refers to the "electron density", that is, the number of free electrons per unit volume. The electron density is how many free (not bound to an atom) electrons there are in a given volume. Since all plasmas have some degree of ionization, this means that there are electrons that have been stripped from atoms, and are moving around, while the atoms are converted into ions [4].

**Plasma density units.** In plasma physics, choice of units can depend on the size of the plasma, or simply preference. Typically scientists and engineers use electrons per

$cm^3$  or electrons per  $m^3$ . In the literature, the electron density is written as a number followed by  $cm^{-3}$  or  $m^{-3}$ . The choice of  $cm^3$  or  $m^3$  stems from using either CGS units or MKS units. It often depends on the length and volume scales that apply to the situation. So, for laboratory-sized plasmas,  $cm^3$  are often used, while for things like inter-stellar or inter-galactic medium,  $m^3$  makes more sense.

## 2.1.2 Application to Wave Particle Simulations

The electron plasma density,  $n_e$ , is a fundamental parameter of plasma, a critical parameter for wave particle simulations and key to understanding the role of plasma waves in the formation and decay of the Earth's radiation belts. It is a crucial parameter that is required for the prediction of the evolution of the space environment and is also important for the analysis of satellite anomalies that have occurred in the past.

The following list describes several important applications of plasma density:

- Plasma density is particularly important in deposition processing and etching process since rate of etching and deposition rates are function of plasma density [5];
- Other processing parameters like uniformity, number of radicals and processing time depend on plasma density [6];
- Variation in plasma density can provide notification about instabilities in plasma due to pressure disturbance and unstable input power [5];
- Ion density and other plasma parameter like Debye length, plasma frequency, and dielectric property of plasma can be measured from electron density [7].

## 2.2 The Van Allen Probes Mission

### 2.2.1 Mission Brief Overview and Science Objectives

On August 30, 2012 NASA launched the Van Allen Probes (formerly known as the Radiation Belt Storm Probes (RBSP)) spacecraft to provide the most detailed picture of Van Allen radiation belts ever seen [2], [8].

The *Van Allen radiation belts* were discovered in 1958 and are two (and sometimes more) giant swaths of radiation in the near-Earth space environment. They are comprised of high-energy ion and electrons that are trapped on Earth encircling orbits as a result of

the dipolar structure of the Earth's intrinsic magnetic field. The inner belt is fairly stable, but the outer belt swells and shrinks with solar activity. When the outer belt swells, this region of dangerous radiation expands to include the orbits of the International Space Station and many other satellites. Space weather can disrupt satellites, cause power grid failures and disrupt GPS communications.

Scientists know that the belts shrink and swell in response to incoming energy from the sun, but they do not know exactly how. The Van Allen Probes, the second mission in NASA's Living With a Star program, helps scientists to understand that. The mission's general objectives are to [2], [8]:

- Discover which processes (singly or in combination) accelerate and transport the particles in the radiation belt, and under what conditions;
- Understand and quantify the loss of electrons from the radiation belts;
- Determine the balance between the processes that cause electron acceleration and those that cause losses;
- Understand how the radiation belts change in the context of geomagnetic storms.

The Van Allen Probes takes its place as part of a fleet of spacecraft that may someday help predict space weather before it even impacts Earth's surroundings [9].

## 2.2.2 Spacecraft Brief Overview

The Van Allen Probes mission consists of two identical spacecraft designed and built at the Johns Hopkins University Applied Physics Laboratory (JHU/APL) [2].

These two spacecraft fly in nearly identical eccentric orbits that cover almost the entire radiation belt region. The spacecraft orbits have apogee altitudes between 30,050 kilometers and 31,250 km, perigee altitudes between 500 km and 675 km, and inclination of 10 degrees from the equator (see Figure 2-1) [2].

Both spacecraft have nine-hour orbital periods. The orbits are designed to allow one observatory to lap the other approximately every 75 days. Apogees that differ slightly (by approximately 100 km) allow for simultaneous measurements to be taken over the full range of observatory separation distances during the course of the mission. Thus, scientists can, for the first time, measure changes in the radiation belts over both space and time.

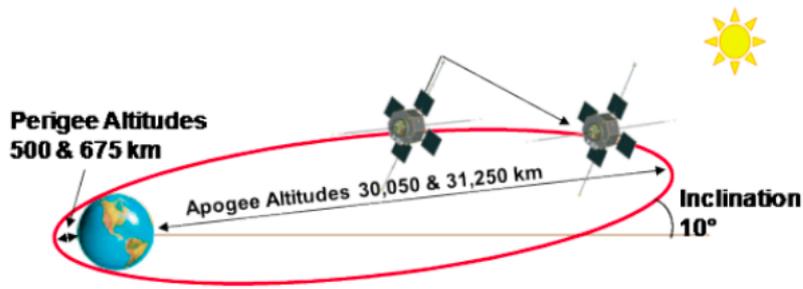


Figure 2-1: Spacecraft orbits [2].

Several facts about spacecraft are given below [2].

**Dimensions.** Each satellite main body is approximately 1.8 m across by 1.3 m tall (the spacecraft span with EFW instrument wire booms fully deployed is 101.7 m, tip to tip, and the spacecraft height with EFW instrument axial booms deployed is 12 m, tip to tip).

**Mass.** The total mass of each spacecraft at launch is 647.6 kilograms for spacecraft A and 666.6 kg for spacecraft B; which includes spacecraft systems with mass of 444 kg on spacecraft A and 463 kg on spacecraft B; instruments with mass of 129.6 kg on each spacecraft and fuel with mass of 56 kg onboard each spacecraft. Spacecraft B has more mass because it carries more mechanical systems and pieces related to attachment to and separation from both the launch vehicle and spacecraft A. Apart from those differences, the design of the two spacecraft is identical.

**Ground Stations.** Twin satellites communicate with Earth using three ground stations. The primary ground station is the 18-meter satellite dish at the Johns Hopkins University Applied Physics Laboratory in Laurel, Md. The two other ground stations are part of the Universal Space Network (USN) and are 13-meter dishes located at different points on the globe (Hawaii and Australia) to maximize coverage.

**Power.** The spacecraft battery provides enough power for full science operations during times when the spacecraft are in Earth's shadow and sunlight cannot reach their solar panels. The daily eclipse times vary throughout the mission, depending on when the spacecraft launch, though the longest eclipse period is 115 minutes.

**Instruments.** Each RBSP spacecraft carries an identical set of five instrument suites (Figure 2-2):

- The Energetic Particle, Composition, and Thermal Plasma Suite (ECT);
- The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS);

- The Electric Field and Waves Suite (EFW);
- The Radiation Belt Storm Probes Ion Composition Experiment (RBSPICE);
- The Relativistic Proton Spectrometer (RPS).

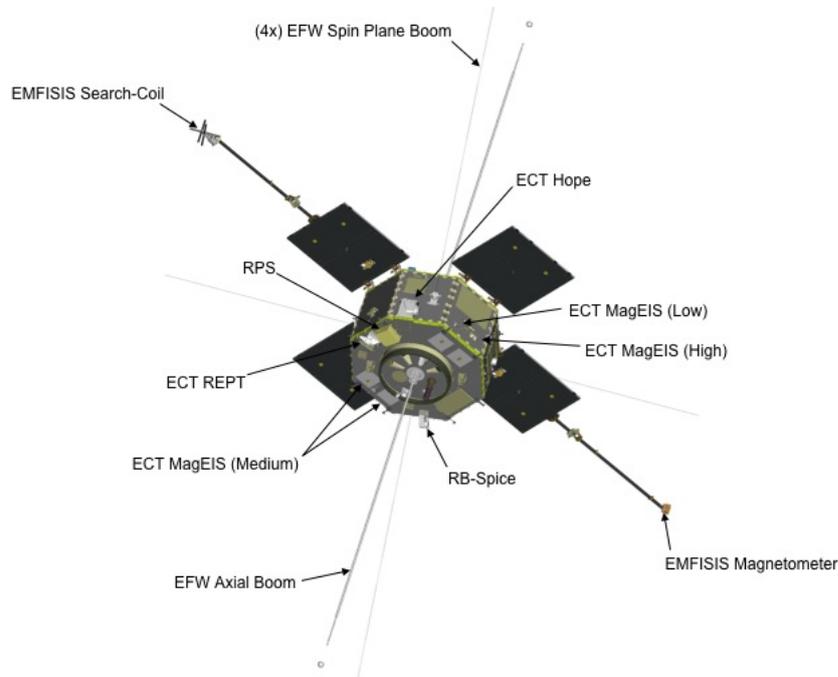


Figure 2-2: The Van Allen Probes spacecraft and instruments; instruments are in identical locations on both spacecraft [3].

In this thesis, data measurements from the EMFISIS instrumentation suite will be considered. The specifics of its design are presented next.

### 2.2.3 The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) Instrumentation Suite

**Science Objectives.** The Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) focuses on the important role played by magnetic fields and plasma waves in the processes of radiation belt particle acceleration and loss [10]. It provides the essential plasma wave measurements of magnetic and electric fields to understand the physics of the interactions between the waves and charged particles that are responsible for acceleration, transport and loss of radiation belt particles. Additionally, the EMFISIS measurements of steady or slowly varying magnetic fields will provide information on much lower frequency phenomena such as the variations in the ring current that encircles Earth, and ultra-low frequency waves which transport particles. These field

measurements, combined with excellent coverage of particle measurements from other instruments and investigations, are revealing the clearest picture ever obtained of radiation belt physics [2].

**Design Details.** The EMFISIS instrumentation suite measurements of DC magnetic fields and a set of wave electric and magnetic field 3D measurements covering the frequency range from 10 Hz up to 12 kHz (and up to 500 kHz).

The need to measure wave properties across the relevant frequency ranges and to measure the DC magnetic field both for background field determination as well as to measure very low frequency waves drives the design to measure 3D vector quantities. For waves between 10 Hz and 12 kHz, EMFISIS makes 3D measurements of both the electric and the magnetic field. Below 10 Hz, EMFISIS measures only the magnetic field. However, this measurement can be combined with the 3D EFW electric field data to have a full set of electromagnetic vector quantities. Above 12 kHz, only a single electric component is measured from 10 to 500 kHz with decreasing response above 400 kHz due to frequency roll-off in the EFW signals.

The highest frequency for the EMFISIS 3D wave measurements is set by the desire to fully measure both lower and upper band whistler-mode chorus<sup>1</sup>. This sets the upper frequency response at 12 kHz. The desire to measure the *upper hybrid line*<sup>2</sup> and the intensity of electron cyclotron harmonics drives the requirement to measure the electric field up to at least 400 kHz. A single electric field component is sufficient to satisfy this. Any of the three EFW dipole pairs can be used for the high frequencies, but it is expected that one of the spin-plane dipoles will provide the best sensitivity and lowest noise. The high frequency receiver is designed to measure up to 500 kHz, but the signal provided by EFW rolls off, significantly, above 400 kHz [13].

**Suite Overview.** EMFISIS comprises two sensors: a tri-axial fluxgate magnetometer (MAG) and a tri-axial AC magnetic search coil magnetometer (MSC). MAG consists of three sensors to measure the background steady or slowly varying magnetic fields and ultra-low frequency waves, and MSC senses quickly varying wave magnetic fields. Additionally, to measure wave electric fields, the Waves instrument uses signals from the

---

<sup>1</sup>Whistler mode waves are right-hand circularly polarized electromagnetic waves propagating at frequencies less than the local gyrofrequency (the frequency at which non-relativistic charged particles circle the magnetic field) [11].

<sup>2</sup>In plasma physics, an upper hybrid oscillation is a mode of oscillation of a magnetized plasma. It consists of a longitudinal motion of the electrons perpendicular to the magnetic field.

The frequency of long wavelength oscillations is a "hybrid", or mix, of the electron plasma and electron cyclotron frequencies,  $\omega^2 = \omega_{pe}^2 + \omega_{ce}^2$  [12]. It is a higher frequency than either because the additional restoring force leads to a higher oscillation frequency.

EFW (AC electric fields) experiment and a tri-axial search coil magnetometer. Signals from these sensors are detected with receivers in a Main Electronics Box (MEB), which collects and processes all of the measurements. More specifically, the Main Electronics Box consists of [2], [13]:

- The Low Voltage Power Supply – converts primary spacecraft power to voltages used by the rest of the suite;
- The Central Data Processing Unit – controls the suite and handles data transfer to and from the central spacecraft systems;
- The Waves component (four boards):
  - two FFT engine boards;
  - the AC electric field receiver;
  - the AC magnetic field receiver;
- The MAG drive, sampling, and heater control board.

The specification of the MAG and the Waves instruments will be reviewed more thoroughly next, since the data from these instruments are exploited in this work.

## 2.2.4 The Fluxgate Magnetometer (MAG)

The EMFISIS Magnetometer is a wide-range, high performance triaxial fluxgate magnetometer system. Its signal processing, analog to digital converter (A/D) and interface electronics are implemented in a single electronics card that is protected by the Main Electronics Box of the Van Allen Probes Spacecraft. The Magnetometer's Flux Gate and Search Coil Booms are attached to two opposing solar arrays of the spacecraft and extend 3 meter from the edge of the Spacecraft Bus. MAG measures 3D vector magnetic fields essential for understanding of particle distributions that are critically dependent on the local and global magnetic fields and their time variation providing a full vector magnetic field vector at a rate of 64 vectors/s, corresponding to a frequency range of 0 – 30 Hz. Magnetometer specifications are given in Table 2.1 [14].

Table 2.1: Magnetometer Specifications.

Features	Specifications		
Data Rate	$\sim 3$ kbs, depending on compression		
Sampling Cadence	64 vectors/s		
Ranges	$0 \pm 256$ nT	$0 \pm 4096$ nT	$0 \pm 65536$ nT
Resolution	$0 \pm 0.001$ nT	$0 \pm 0.16$ nT	$0 \pm 2$ nT
Accuracy	0.1 nT (sensor alone, with spacecraft 5 nT)		
Frequency Range	0 – 30 Hz		

### 2.2.5 The Waves Instrument

The Waves magnetic sensor is comprised of three search coil antennas mounted in a tri-axial configuration installed on the boom opposite from the magnetometer boom to reduce any interference. The axial booms of Van Allen Probes have a length of 12 meters from tip to tip while the spin plane booms are 40 and 50 meters in length [3].

The goal of the EMFISIS Waves Instrument is quantitative determination of the effect of plasma waves in the radiation belts on particles present in the Belts. The Waves instrument provides two types of wave measurements [13]:

1) A three axis electric and magnetic field measurement covering the frequency range of 10Hz to 12kHz. The electronics and the electric field signals for the Waves Instrument are provided by the EFW experiment and consists of differential voltages from opposing EFW spherical sensors, and by the Waveform Receiver (WFR) of the Waves instrument. By combining both electric and magnetic measurements, digitized simultaneously by common electronics, Waves provides a full vector of electric and magnetic field measurement that enables calculation of key quantities such as Poynting flux, polarization, planarity, and ellipticity for key wave modes such as VLF hiss, magnetosonic equatorial noise, and chorus [3].

2) A single axis electric field measurement over the frequency range 10–400 kHz provided by the High Frequency Receiver (HFR) of the Waves instrument in order to measure the spectrum of electron cyclotron harmonic emissions and the frequency of the upper hybrid resonance band to determine background plasma electron density.

There are also two modes of operation: survey and burst modes. Burst modes provide

continuous waveforms or very high temporal resolution spectral matrices (30 ms resolution), while in a survey mode spectral densities are collected every 6 s. In survey mode, all digital waveforms are Fourier transformed using a floating point fast Fourier transform (FFT) on board [13].

## 2.3 Determining the Electron Density from Upper-Hybrid Band Resonance Frequency

There are numerous existing techniques to conduct these measurements either directly, such as direct particle counting techniques with various types of plasma instruments, or indirectly (quasi-thermal noise spectroscopy [15]; using the spacecraft potential as a proxy for the electron density [16]; determining the electron density from intense upper-hybrid band emissions [17] etc.). Determining the electron density from intense upper-hybrid band emissions is still considered to be the most reliable passive technique [18]. In this thesis, this method is employed.

We use the technique described by Mosier et al. [19] to derive electron number density,  $n_e$ , profiles from the intense plasma noise bands typically observed near the upper hybrid frequency. The upper hybrid frequency,  $f_{uhr}$  (where *uhr* in the subscript stands for Upper-Hybrid Resonance) is a combination of the electron cyclotron frequency,  $f_{ce}$ , and the electron plasma frequency,  $f_{pe}$ :

$$f_{uhr} = \sqrt{(f_{ce}^2 + f_{pe}^2)}. \quad (2.1)$$

The electron cyclotron frequency is a function of the magnetic field strength,  $B$ , and the electron plasma frequency is a function of the electron number density,  $n_e$ . In SI system these frequencies are given as

$$f_{ce} = \frac{|q_e| B}{2\pi m_e}, \quad f_{pe} = \frac{1}{2\pi} \sqrt{\frac{q_e^2 n_e}{m_e \epsilon_0}}, \quad (2.2)$$

where  $q_e$  is the charge of electron,  $m_e$  is the mass of an electron, and  $\epsilon_0$  is the permittivity of free space.

In the EMFISIS data, the intense UHR noise band typically extends from  $f_{uhr}$  down to  $f_{pe}$  with a generally less intense emission extending down to the z-mode cutoff frequency given by:

$$f_z = \frac{1}{2} \left[ (f_{ce}^2 + 4f_{pe}^2)^{\frac{1}{2}} - f_{ce} \right]. \quad (2.3)$$

The upper limit of the noise band at  $f_{uhr}$  is usually the most pronounced cutoff in the data.

An example measurement is shown below in Figure 2-3. The top panel is the survey mode HFR data. The upper hybrid resonance band is indicated. The magnetic field strength is directly measured by the magnetometer onboard the spacecraft and thus,  $f_{ce}$ , can be directly determined and is indicated on the spectrogram. The bottom two panels show slices of the electric field power spectral density as a function of frequency at two specific times. The middle panel is from 7.9 UT when the spacecraft was in the high-density region. The estimated values  $f_{uhr}$ ,  $f_{pe}$ , and  $f_z$  are indicated. The lower panel shows a slice from 9.9 UT when the spacecraft was in the lower density region. In the lower density region, strong half harmonics of the electron cyclotron,  $(n + \frac{1}{2})f_{ce}$ , can be observed, and the  $f_z$  cutoff is not typically observed.

From 8.75 to 9.25 UT, there are fluctuations in the UHR band. These indicate the presence of density irregularities, and it is important to capture these effects.

In the example above, the UHR band was fairly clear. At other times, the spectrum is contaminated with a variety of plasma emissions. The main challenge of developing a robust automated algorithm is to understand and properly deal with the different types of contaminating emissions. Often the UHR band in the high-density region is fairly clear, but there is more noise in the lower density region. You can see more examples of orbits in Appendix A.

## 2.4 AURA – An Automated Upper-hybrid Resonance Detection Algorithm

The Van Allen Probes EMFISIS team at the University of Iowa devised the Automated Upper-hybrid Resonance Detection Algorithm (AURA) to extract electron plasma densities from wave observations of the upper-hybrid resonance band,  $f_{uhr}$ , in dynamic spectra [1].

The authors use a technique similar to that described in 2.3 to derive electron densities from the upper-hybrid resonance band,  $f_{uhr}$ . To detect  $f_{uhr}$  AURA employs a restricted searching approach to detect a relative peak frequency signal embedded in an individual

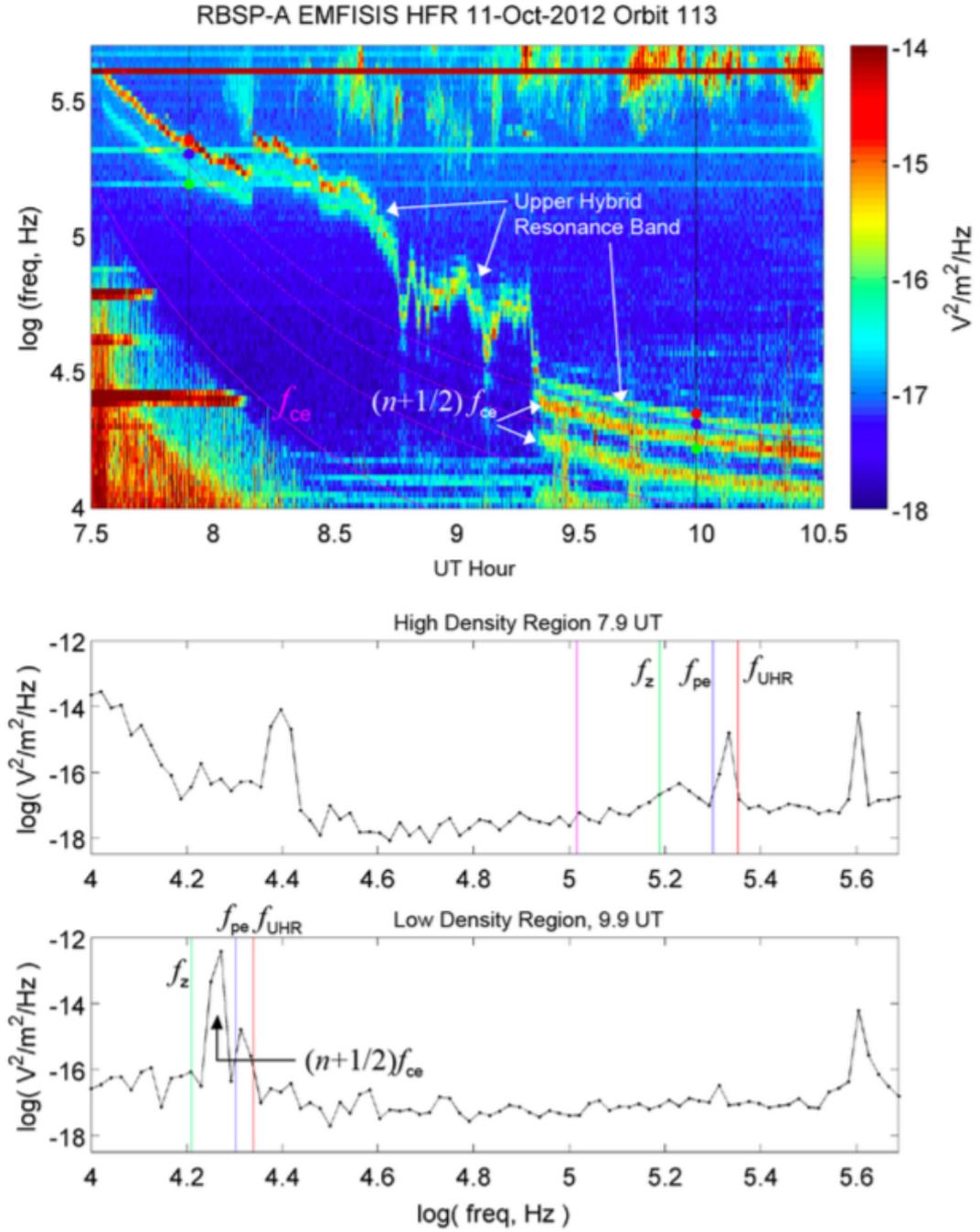


Figure 2-3: An example measurement. The top panel is the survey mode HFR data. The bottom two panels show slices the electric field power spectral density as a function of frequency at two specific times.

spectrum. This approach is based on a rule of hysteresis that assumes each successive spectrum contains a peak frequency associated with  $f_{uhr}$  near the previously identified spectral record (i.e., there is not a large change in frequency of the binned signal within a 6 s sampling time step).

AURA is used for a semi-automated  $f_{uhr}$  extraction process. It does an excellent job of automatically finding  $f_{uhr}$  with little or no manual intervention when the spectrum

is simple. For example, these are such orbits for which the spacecraft never leaves the plasmasphere or when geomagnetic conditions are very quiet.

To assess algorithm's performance, the authors split orbits data into 3 types: A, B, and C:

- Type A refers to 70% of the HFR orbit spectrograms from mission start (September 2012) to full MLT coverage (July 2014) successfully digitized into density records with less than 1/4 visually determined as needing correction (a single orbit at 6 sec cadence may contain as many as 4000 points).
- Type B constitutes 20% of the data set that is contaminated by 1/4 to 1/2 misidentified records, but nevertheless is quickly remedied by manual inspection.
- The final category, Type C, coincides with high geomagnetic activity where severe convection erodes the plasmasphere to low L shells and is observed as strong emission after a plasmopause crossing.

According to the results presented in the paper, the highest level of success, class A, requires less than 25% of the data points to be corrected for a given orbit. Class B requires 25 to 50% of the data points to be corrected. The remaining class C represent challenges, typically occurring during high geomagnetic activity in which a large portion of the outer part of the orbit presents interpretational difficulties in finding the upper hybrid band for a number of reasons. Typically, these orbits exhibit very low densities in the plasma trough.

# Chapter 3

## Nonlinear Regression with Neural Networks

Regression analysis is used when one wants to predict a continuous dependent variable from a number of independent variables. Nonlinear regression is a form of regression analysis, in which observational data are modelled by a function, which is a nonlinear combination of the model parameters and depends on one or more independent variables. In this thesis, we solve exactly such problem, i.e. we predict plasma density from a set of independent variables (features), using feedforward neural networks, one of the models in machine learning that are widely-used and quite effective for many problems. In this chapter, the theory behind them is described.

### 3.1 Problem Statement

Let there be given a sample set of  $M$  pairs  $(\mathbf{x}_i, y_i)$ . Also, the regression model  $f(\mathbf{w}, \mathbf{X})$  is given, which depends on a set of parameters  $\mathbf{w} = (w_1, \dots, w_N)$  and independent variables  $\mathbf{X}$ .

It is necessary to find a set of parameters that would minimise the sum of residuals:

$$S = \sum_{i=1}^M r_i^2, \quad (3.1)$$

where a residual is  $r_i = y_i - f(\mathbf{w}, \mathbf{x}_i)$ , for  $i = 1, \dots, M$ .

## 3.2 Neural Networks for Regression

The problem stated in the previous section consists essentially in finding a mapping  $f$  from  $\mathbf{X}$  to  $\mathbf{y}$ :

$$\mathbf{y} = f(\mathbf{w}, \mathbf{X}), \quad (3.2)$$

where *transfer function* (TF)  $f$  can be any function, and it depends on  $\mathbf{X}$  and on a set of parameters  $\mathbf{w}$ . This empirical mapping can be performed using conventional tools (linear and nonlinear regression).

In many real-world problems, linear models are not sufficient to capture the real-world phenomena, and thus nonlinear models are necessary. In this case, TF  $f$  is a nonlinear function, and we need to implement a particular type of nonlinearity a priori. This may not always be possible, because we may not know in advance what kind of nonlinear behavior a particular TF demonstrates, or this nonlinear behavior may be different in different regions of the TF's domain. If an inappropriate nonlinear regression function is chosen, it may represent a nonlinear TF with less accuracy than with its linear counterpart [20].

In this situation (TF is nonlinear and the form of nonlinearity is not known), we need a more flexible, self-adjusting approach that can accommodate various types of nonlinear behavior representing a broad class of nonlinear mappings. Neural networks (NNs) are well-suited for a very broad class of nonlinear approximations and mappings.

## 3.3 Feedforward Neural Networks

The Feedforward Neural Network is one of the most common and the first type of artificial neural networks devised. Artificial neural network is basically a complex mathematical function (model), which mimics a biological neural networks in the living brain in the simplest fashion [21]. Similarly to linear classification and regression methods, neural networks yield outputs in the following form [22]:

$$y(\mathbf{x}, \mathbf{w}) = f\left(\sum_{j=1}^N w_j \phi_j(\mathbf{x})\right), \quad (3.3)$$

where  $f$  is a nonlinear activation function,  $\mathbf{w}$  is a vector of weights, and  $\phi$  are nonlinear basis functions.

The Feedforward Neural Network (FNN) is the most basic and widely used artificial neural network. FNNs have achieved success in a number of domains (e.g., [23], [24], [25]),

most notably in large vocabulary continuous speech recognition [26], where they were directly responsible for considerable improvements over previous highly-tuned, state-of-the-art systems.

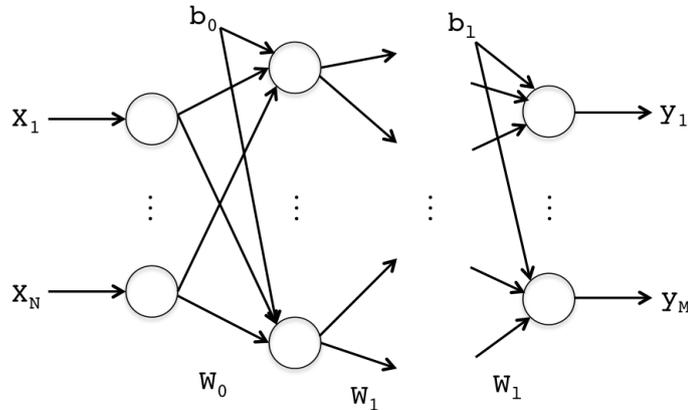


Figure 3-1: General architecture of the Feedforward Neural Network.

General architecture of the FNN is presented in Fig. 3-1. It consists of a number of layers of artificial neurons that are arranged into a layered configuration. The input layer consists of the inputs to the network. Then follows a hidden layer, which consists of any number of neurons placed in parallel. Each neuron performs a weighted summation of the inputs, which then passes a nonlinear activation function. There might be several hidden layers in the network. The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer. This summation on the output is called the output layer [21].

Formally, a feedforward neural network with  $l$  hidden layers is parameterized by  $l + 1$  weight matrices ( $W_0, \dots, W_l$ ) and  $l + 1$  vectors of biases ( $b_0, \dots, b_l$ ). The ultimate goal then is to find the optimal set of parameters (weights and biases) of the network. It is done by training, during which parameters of the network are adjusted incrementally until the training data satisfy the desired mapping as well as possible.

# Chapter 4

## Data

In this thesis, data from the Electric and Magnetic Field Instrument Suite and Integrated Science (EMFISIS) instrumentation suite was used. Description of the instrument can be found in Section 2.2.3, and description of the data produced by the instrument and used in this thesis is presented in this chapter.

### 4.1 Van Allen Probes Data Processing

Data acquired by the individual spacecraft is stored aboard the vehicle before being downlinked to the Mission Operations Center which then relays the data to the Science Operations Center for each of the instruments. It is then processed and categorised by levels. There are 5 Data Levels [14], [13]:

- *L0 Data (Level 0)* is the raw data downlinked from the spacecraft. This data is in binary format and data volume varies between the different instruments: ECT: 450MB/day; EMFISIS: 500MB/day; EFW: 130MB/day; 414MB/day; RPS: 50MB/-day.
- *L1 Data* is achieved at the Science Operations Center by initial data processing. It is larger in volume and includes CDF or CSV formatting. For some instruments, L1 data is available within 8 hours, but for others L1 data will be available 14 days after receipt.
- *L2 Data* include calibrated data and initial analysis and is available 1 month after acquisition.

- *L3 Data* includes in depth data analysis with additional values gained from raw data. Events are time-stamped, calibrated and all background noise is removed.
- *L4 Data products* are high level data sets that include global maps, phase space density data as well as in depth analysis. L4 data is available after one year of data processing for all instruments (except RPS, for which L4 Data in the form of Global Maps is available within 9 days).

In this thesis, L2 and L4 Data products were used for building a model to infer plasma densities. The detailed description of these products is given in the next sections.

## 4.2 L2 Data: Van Allen Probes Plasma Wave Observations

L2 Data used here are Plasma Wave Observations obtained by the EMFISIS instrument suite. We used the data measurement from the magnetometer (MAG) and from the HFR component of the Waves instrument. The data were retrieved from the EMFISIS official website [14].

The data is presented as a set of files in Common Data Format (CDF) format (this is a self-describing data format for the storage of scalar and multidimensional data in a platform- and discipline-independent way [27]). The data is time-ordered and non-duplicating. All the data were parsed into MATLAB files.

Tables 4.1 and 4.2 shows the data products used after parsing and their brief description. These data were used to create features to fit the regression model to infer plasma densities. This process is described in Chapter 5. It is also worth mentioning that all the data at the web server is organised in term of days (for each day there are files with corresponding information). In this work, we arranged the data in terms of orbits, i.e. for each particular orbit there is corresponding data. In the tables below and throughout all the work this arrangement is used.

Table 4.1: HFR data used.

Feature name	Dimensions	Units	Description
Epoch	$[1 \times n]$	$ns$	Default time
Spectra	$[82 \times n]$	$V^2/m^2/Hz$	HFR Spectra Data
Frequencies	$[82 \times 1]$	$Hz$	HFR center band frequencies

Table 4.2: MAG data used.

Feature name	Dimensions	Units	Description
Epoch	$[1 \times n']$	$ns$	Default time
$f_{ce}$	$[1 \times n']$	$Hz$	Electron cyclotron frequency
$L$	$[1 \times n']$	Earth-radii	Magnetic field lines
MLT	$[1 \times n']$	Hours	Magnetic local time

### 4.3 L4 Data: Electron Densities

L4 Data products used here are upper hybrid frequencies and electron densities. L4 products were generously provided through the courtesy of William Kurth and the EMFISIS team.

The densities were derived using AURA, an Automated Upper-hybrid Resonance detection Algorithm, described earlier in Section 2.4. This data is in CDF format as well and is also time-ordered and non-duplicating.

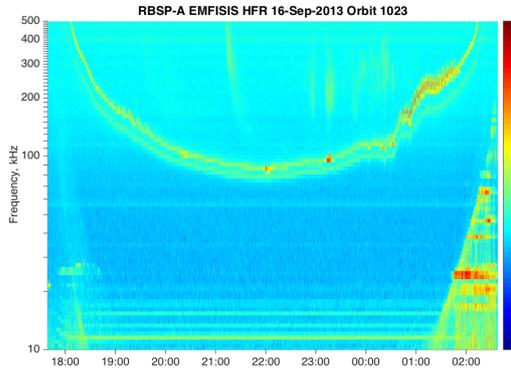
Table 4.3 shows the data products used. These data were used as target variables in the training and for comparison analysis.

Table 4.3: L4 data used.

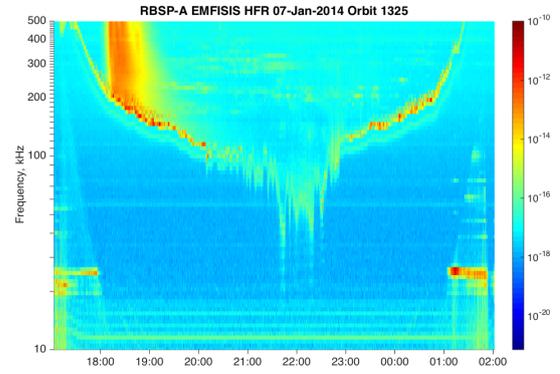
Feature name	Dimensions	Units	Description
Epoch	$[1 \times n'']$	$ns$	Default time
$f_{uhr}$	$[1 \times n'']$	$Hz$	Upper hybrid frequency
$n_e$	$[1 \times n'']$	$m^{-3}$	Plasma density

## 4.4 Data Examples

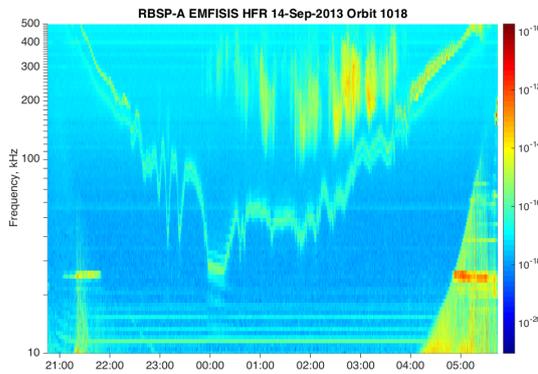
Example spectrograms of Waves HFR data are presented in Figure 4-1. More examples of spectrograms can be found in Appendix A.



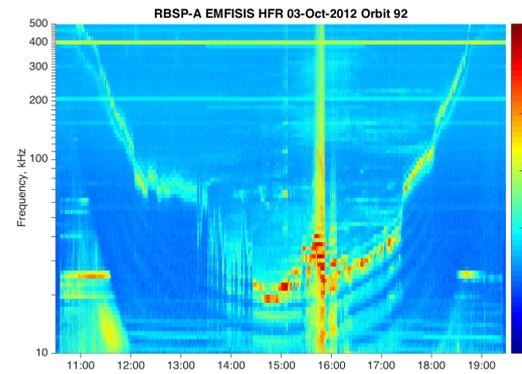
(a)



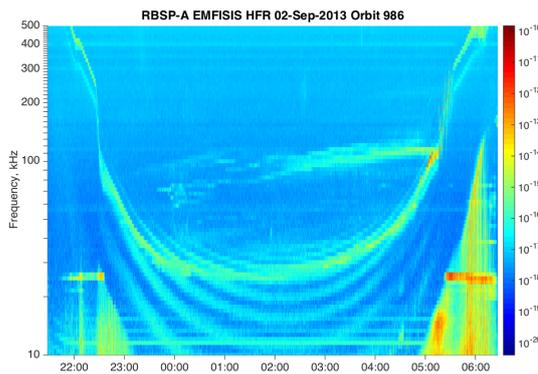
(b)



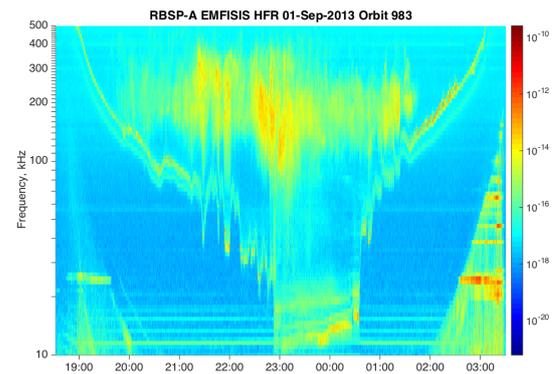
(c)



(d)



(e)



(f)

Figure 4-1: Example spectrograms of Waves HFR data.

# Chapter 5

## Methodology

This chapter contains a thorough description of the experimental set-up used in this thesis project, as well as the software written to process the collected data. For a more general description of the theory behind neural networks, please refer to Chapter 3.

### 5.1 Problem Statement

A set of measurements  $\mathfrak{S}$  is given such that

$$\mathfrak{S} = \{\mathfrak{s}_1, \dots, \mathfrak{s}_M\}. \quad (5.1)$$

The element  $\mathfrak{s}_i$  of the set  $\mathfrak{S}$  is a measurement from EMFISIS (includes spectra,  $f_{ce}$ ,  $L$ , MLT) at a particular moment of time. Let  $\mathbf{y} = \{y_1, \dots, y_M\}$  be a set of target variables, which are upper hybrid frequencies  $f_{uhr}$ .

Together with the set  $\mathfrak{S}$  a set  $V = V(\mathfrak{S})$  is given. The set  $V = V(\mathfrak{S})$  is called a vocabulary and contains knowledge about the set of measurements. The vocabulary is obtained as the result of measurement structure analysis and used for model generation.

By  $G = \{g_1, \dots, g_N\}$  we denote an expert-given set of primitive functions such that each function  $g_j$  maps an object  $\mathfrak{s}_i$  to an element  $(i, j)$  of the design matrix  $\mathbf{X}$ :

$$g_j : (\mathbf{b}_j, \mathfrak{s}_i, V) \mapsto x_{ij} \in \mathbb{R}^1, \quad (5.2)$$

where  $\mathbf{b}_j$  is the set of parameters of the primitive function  $g_j$ .

Let  $f$  be a nonlinear mapping from  $\mathbf{X}$  to  $\mathbf{y}$  such that

$$f : (\mathbf{w}, \mathbf{X}) \mapsto \mathbf{y} \in \mathbb{R}^M, \quad (5.3)$$

where  $\mathbf{w}$  is a set of parameters of the model.

To find the optimal parameters  $\hat{\mathbf{w}}$  we need to minimize a loss function  $S(\mathbf{w}|f, \mathbf{X}, \mathbf{y})$  such that

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^N} S(\mathbf{w}|f, \mathbf{X}, \mathbf{y}), \quad (5.4)$$

where the loss function is given as

$$S(\mathbf{w}|f, \mathbf{X}, \mathbf{y}) = \frac{1}{M} \sum_{i=1}^M (f(\mathbf{w}, \mathbf{x}_i) - y_i)^2. \quad (5.5)$$

## 5.2 Design Matrix Description

The input data is a pattern in per se. If the key attributes or features characterizing the data can be extracted, the problem encountered can be easily solved. However, feature extractions are usually dependent upon the domain-specific knowledge.

As was mentioned in the problem statement that an expert-given set of primitive functions  $G = \{g_1, \dots, g_N\}$  produces design (or feature) matrix  $\mathbf{X} \in \mathbb{R}^{M \times N}$  from the given set of measurements  $\mathfrak{S}$ . The element  $\mathfrak{s}_i$  of the set  $\mathfrak{S}$  is a measurement from EMFISIS at a particular moment of time that includes spectra,  $f_{ce}$ ,  $L$ , MLT (these parameters are described in detail in Chapter 4) with a corresponding  $f_{uhr}$ . Here, we will describe features produced by these functions.

In the feature extraction process, we relied on assumptions what characteristics and measurement are related and influence the upper hybrid frequencies values. Table 5.1 shows the 88 features produced.

All features are time dependent. As described in Chapter 4,  $f_{ce}$ ,  $L$  and MLT measurement were provided by the magnetometer, and the spectra – by the HFR component of the Waves instrument. These instruments have different cadence, i.e. the frequency of taking routine measurement, which means that the number of measurement for the duration of an orbit is not the same for the two instruments. In order to provide  $f_{ce}$ ,  $L$  and MLT measurement for each measurement in the spectra, linear interpolation was performed.

As target variables  $\log_{10} f_{uhr}$  were used.

Table 5.1: Features used in the design matrix.

Feature	Description	Possible values
$\log_{10} spectrum$	Decimal logarithm of the spectrum	$\mathbb{R}$
$\log_{10} f_{ce}$	Decimal logarithm of electron cyclotron frequency	$\mathbb{R}$
$L$	Magnetic field lines	$\mathbb{R}$
$K_p$ index	Global geomagnetic storm index	Ranges from 0 to 9
MLT	Magnetic local time	Ranges from 0:00 to 24:00
$fbin_{max}$	Frequency bin of the largest element of the Waves HFR spectrum	$\{1, 2, \dots, 82\}$
$n$	Order of the Waves HFR spectrum in the spectrogram	$\mathbb{R} \in [0; 1]$

At the moment, there is data available for 2,425 orbits. Spectral data for each orbit comprises approximately 5,000 spectra (measurements are conducted with a 6 seconds cadence). It is total of  $\sim 12,000,000$  data points. But the dataset of available upper hybrid frequencies is labeled for 1091 orbits. Thus, the total number of samples in the constructed design matrix turned out to be  $\sim 5,100,000$  points.

### 5.3 Neural Network Design and Implementation

Feedforward Neural Networks were employed to solve the problem stated above. For more detailed discussion and rationale behind this choice please refer to Chapter 3. This section describes the experimental setup and the workflow for neural network design.

#### 5.3.1 Neural Network Design Workflow

The general workflow for the neural network design process has the following major steps:

1. Collect data;
2. Clean data;

3. Split data;
4. Preprocess data;
5. Create the network;
6. Train the network;
7. Validate the network;
8. Use the network.

Data collection was discussed in Chapter 4. Details of the other steps are discussed in the following sections.

### 5.3.2 Data Cleaning

Sometimes, data may contain missing or corrupted records. Data cleaning is the process of detecting and correcting or removing corrupt or inaccurate records from a data set.

Here, in the design matrix obtained by features extraction, some data points contained NaNs <sup>1</sup>. NaNs in data are caused by linear interpolation that as was already mentioned was applied to some measurement for the reason of different cadence of instruments. All data points containing NaNs were removed from the dataset.

Dataset after cleaning comprised 4,077,993 time points.

### 5.3.3 Data Split

Data was randomly split into three parts: training, validation, and test sets. Training set is used to train neural networks, validation set is used for model selection. Test set was left aside for the whole process of training, feature and model selection. It was used in the very end to evaluate the performance of the model.

The ratio of the split is 0.34 : 0.33 : 0.33. Generally, the split is something like 0.7 : 0.15 : 0.15, or even 0.8 : 0 : 0.2 when the dataset is small (in this case cross-validation is usually conducted for model and feature selection). But in our case, the large size of the dataset allows such almost even split [28].

---

<sup>1</sup>NaN stands for not-a-number – numeric data type value representing an undefined or unrepresentable value.

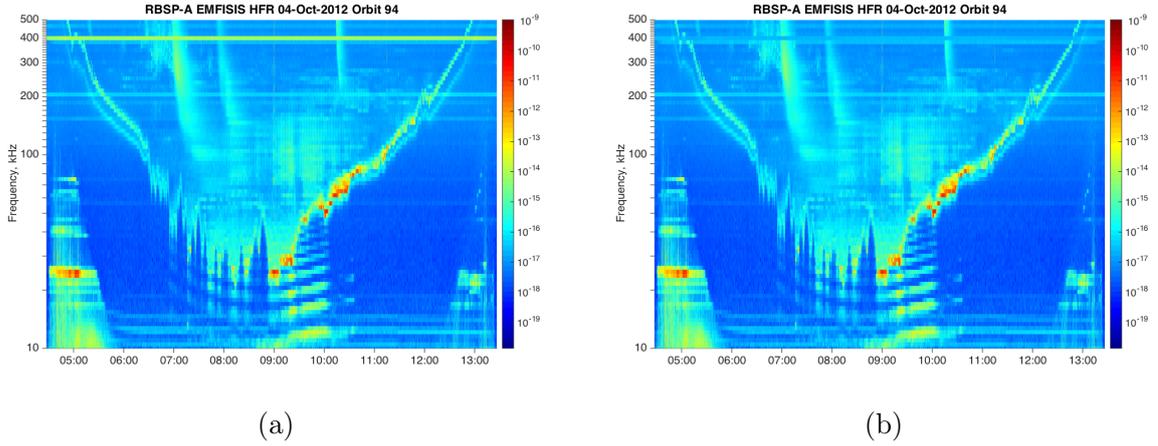


Figure 5-1: An example spectrogram of Waves HFR data (a) before and (b) after noise line removal.

### 5.3.4 Data Preprocessing

Data preprocessing converts raw data and signals into data representation suitable for application through a sequence of operations. The objectives of data preprocessing include size reduction of the input space, data normalization, noise reduction, and feature extraction. Feature extraction process was described in the previous section. Here, noise reduction and data normalization will be discussed.

#### Noise Reduction

A sequence of data may involve useful data, noisy data, and inconsistent data. Preprocessing is applied to reduce the noisy and inconsistent data [28].

After careful data analysis and manual inspection, it was identified that some data contains noise lines that are caused by interference from other instruments on board (an example is shown in 5-1a). Some of these noise lines can significantly contaminate the calculations of  $f_{uhr}$ . Such noise lines were removed from the spectra. An example of the spectra before and after the removal is presented in Figure 5-1.

#### Normalization

For many practical problems, the units used to measure each of the input variables can skew the data and make the range of values along some axes much larger than others. This results in unnecessarily complex relationships by making the nature of the mapping along some dimensions much different from others. This difficulty can be circumvented by normalizing (or scaling) each of the input variables so that the variance of each variable

is equal. Scaling the data equalises the importance of variables. A large value input can dominate the input effect and influence the model accuracy of the neural network system [29].

Several algorithms can be used to normalize the data. Here, *min-max normalization* was used.

Min-max normalization is a linear scaling algorithm. It transforms the original input range into a new data range (typically 0-1). It is given as

$$y_{new} = \left( \frac{y_{old} - min_1}{max_1 - min_1} \right) (max_2 - min_2) + min_2, \quad (5.6)$$

where  $y_{old}$  is the old value,  $y_{new}$  is the new value,  $min_1$  and  $max_1$  are the minimum and maximum of the original data range, and  $min_2$  and  $max_2$  are the minimum and maximum of the new data range (here 0 and 1 accordingly). Since the min-max normalization is a linear transformation, it can preserve all relationships of the data values exactly.

In terms of features, "row-wise" normalization was applied to the spectra only. Other features ( $f_{ce}$ ,  $L$ , etc.) were normalized in a "non-row-wise" way. For such features, it is worth mentioning that normalization is applied to the training set before being used to train the network, and exactly the same pre-processing should be done to the test set, if we are to avoid peculiar answers from the network.

### 5.3.5 Training Routine

After the data has been collected, the next steps are to create the network and to train it. This section will describe the experimental setup and the specific parameters used for training will be described in the next two sections.

Matlab Neural Network Toolbox was used to create and train the model. This toolbox provides functions and apps for modeling complex nonlinear systems that are not easily modeled with a closed-form equation. Neural Network Toolbox supports supervised learning with feedforward networks [30].

Using Matlab's function `feedforwardnet(hiddenSizes)` a feedforward neural network was created. This function creates a multilayer feedforward network consisting of a series of layers specified by `hiddenSizes` that is a row vector of one or more hidden layer sizes.

To train the network `train(net, X, y)` was used, where `net` is the network created, `X` is the training sample, and `y` are the target variables. It trains a network according

to `net.trainFcn` and `net.trainParam` specified by the user (or using default values). Parameter `trainFcn` defines algorithm used for training the network. Parameter `net.trainParam` indicates the training parameter values such as maximum number of epochs to train, performance goal, learning rate, minimum performance gradient, momentum constant, etc. Parameters used in this setup are shown in Table 5.2.

Table 5.2: Training parameter values used in the experimental setup.

Parameter	Value	Description
<code>net.trainFcn</code>	Scaled conjugate gradient backpropagation algorithm	Training algorithm
<code>net.trainParam.goal</code>	1e-5	Performance goal
<code>net.trainParam.min_grad</code>	1e-5	Minimum performance gradient
<code>net.trainParam.epochs</code>	1000	Maximum number of epochs to train

Scaled conjugate gradient backpropagation algorithm was chosen for training here, since it performs well over a wide variety of problems, particularly for networks with a large number of weights. The SCG algorithm is faster for large networks than any other algorithms presented in the toolbox, and it has relatively modest memory requirements [31].

### 5.3.6 Validation

Validation of networks with different sets of parameters was conducted after the training procedure to determine optimal parameters of the model. During the series of experiments with different neural networks architectures, the best suitable model was chosen. Figure 5-2 shows the plot of the error of different models vs. the number of neurons in their hidden layers. Red solid curve shows the results for the training set, blue dashed curve shows results for the validation set. The neural network with 80 neurons in the hidden layer was selected as the best model since it shows the best performance on the validation

set, and as the number of neurons in the hidden layer increases the performance decreases.

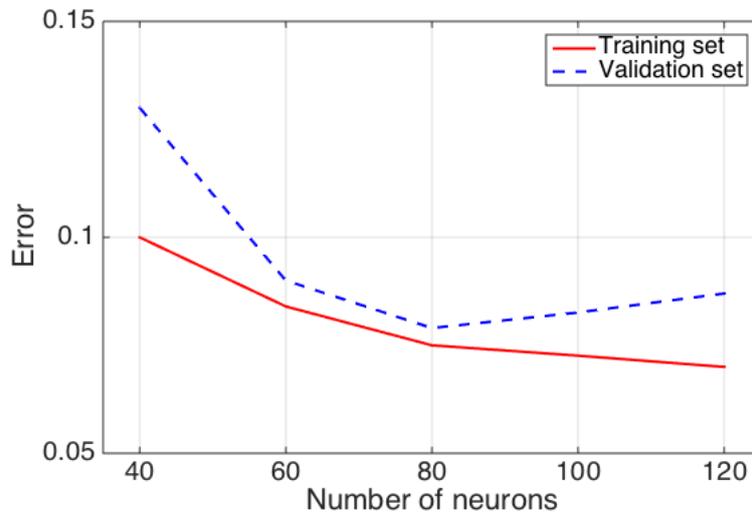
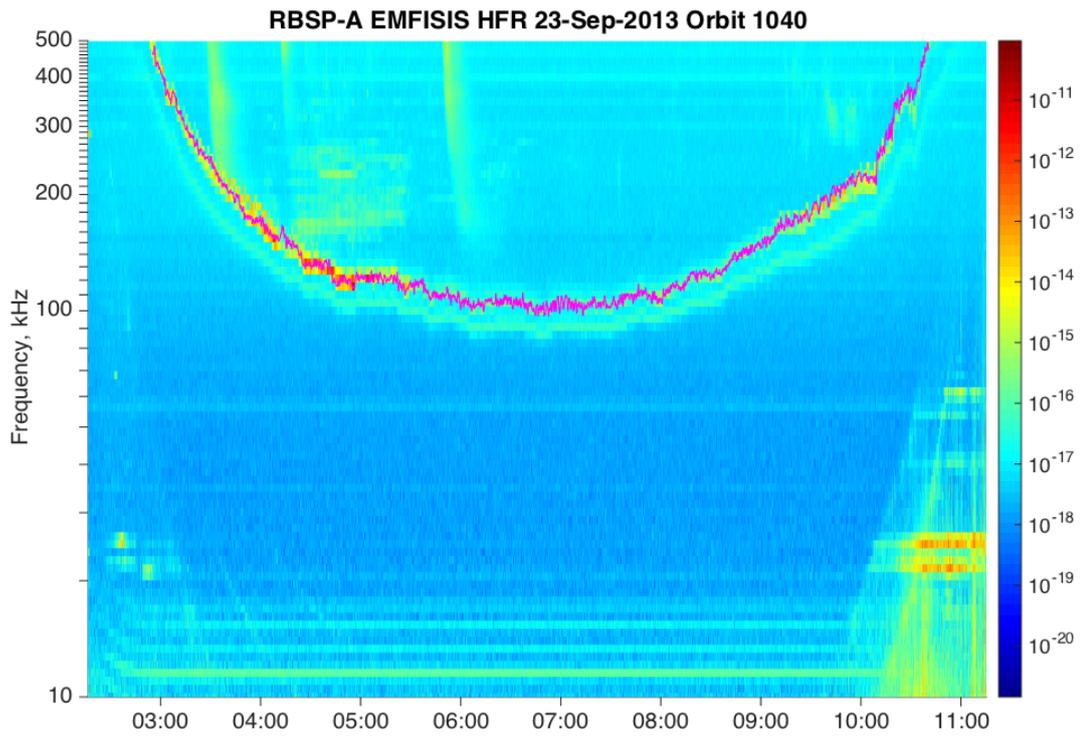


Figure 5-2: Error of different models vs. the number of neurons in their hidden layers. Red solid curve shows the results for the training set, blue dashed curve shows results for the validation set.

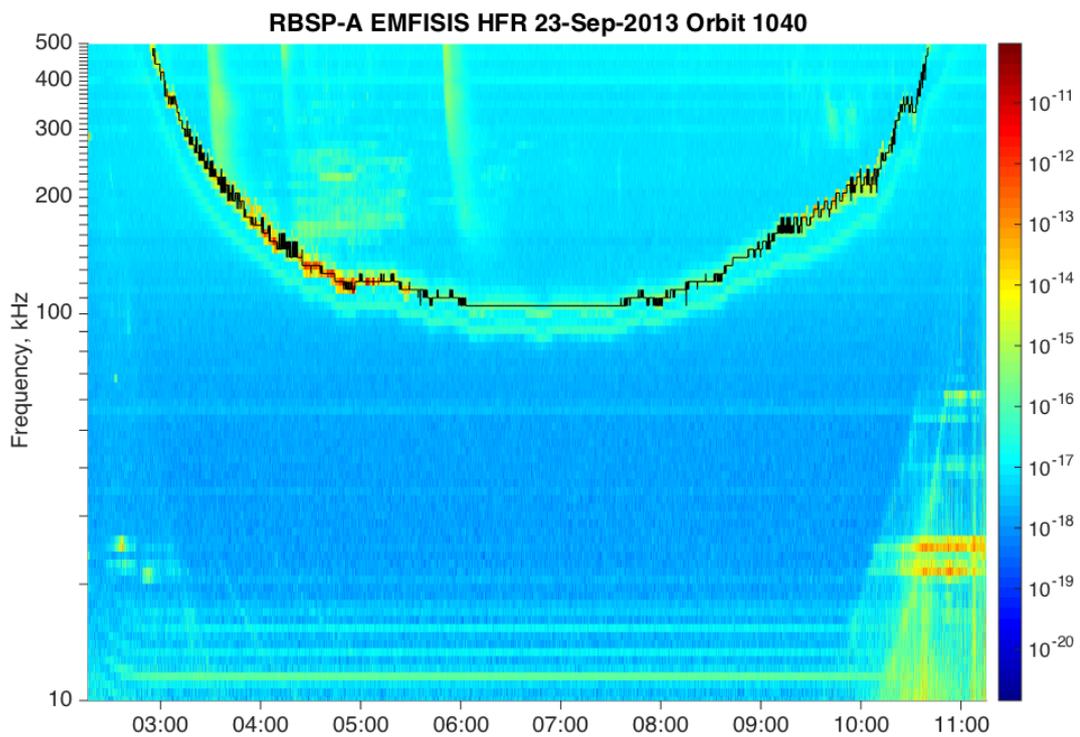
### 5.3.7 Neural Network Usage

The final step of the pipeline under consideration is usage of the neural network. The details of it are described below.

After applying final neural network model to the data, the post-processing procedure of the output was done. The main reasoning behind this step consists in the following. The output of the network represents signal plus noise. We can try to reduce the noise and increase accuracy since we have the knowledge of how the upper hybrid frequency is usually identified manually. Basically, this frequency corresponds to one of the peaks of the spectrum. In many cases, it is the largest peak in the spectrum, but in some cases (when other harmonics are presented in the spectra, and also during geomagnetically active times) it may not be the peak that is maximum in value, as peaks corresponding to other harmonics are larger in value [1]. Thus, we use the neural network output as an approximation of the upper hybrid frequency, and to infer the exact values we search for the maximum value in the corridor around this approximate value of upper hybrid frequency. Example of the output of neural network and post-processed output is presented in Figure 5-3. Computational experiment showed that the accuracy increases by 4.8% (for the selected model on validation set) with this preprocessing.



(a)



(b)

Figure 5-3: An example of the output of the model on the spectrogram of Waves HFR data (a) before and (b) after post-processing step.

# Chapter 6

## Results

This chapter describes the main results of this project. In the first section, statistical results and characteristics of the electron densities determined by the proposed algorithm are given together with AURA's statistics. In the second section, these results are compared with the ones obtained by AURA. Next, expert evaluation of both results are given and the dataset constructed with the aid of the algorithm proposed is described.

### 6.1 Statistical Results

Figure 6-1 shows the average electron density  $n_e$  as determined from the Van Allen Probes mission data by the algorithm proposed in the project and using AURA. The density  $n_e$  is plotted versus  $L$ . As it can be seen from this plot, these two distributions obtained by different algorithms are identical and almost overlap.

Figure 6-2 is like Figure 6-1 except that the data is separated into four groups according to MLT and shown just for the algorithm proposed in the work. The thin solid curve corresponds to the midnight sector  $MLT = 21 - 03$ , the thin dashed curve corresponds to dawn  $MLT = 03 - 09$ , the thick solid curve corresponds to noon  $MLT = 09 - 15$ , and the thick dashed curve corresponds to dusk  $MLT = 15 - 21$ . It is clear that the density is depleted in the dawn sector ( $MLT = 03 - 09$ ). Alternately, one might say that there is a greater probability of plasmatrough density in the dawn local time sector (the plasmopause is typically at its lowest L shell near dawn).

Similarly, in Figure 6-3 the data was separated into two groups, geomagnetically quiet times with  $K_p < 1.2$  (solid blue curve) and active times with  $K_p > 2$  (dashed red curve). Figure 6-3 shows that during more active times ( $K_p > 2$ ), the density is depleted.

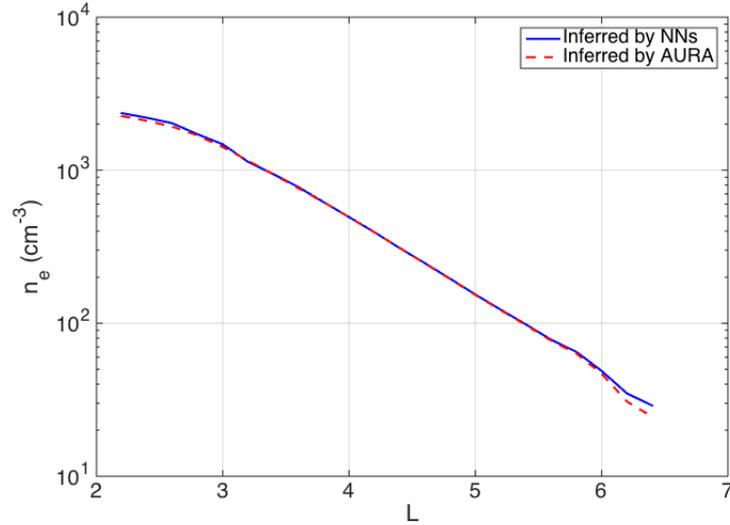


Figure 6-1: Average electron density  $n_e$  as determined from the Van Allen Probes mission data versus  $L$  using the proposed algorithm (blue solid line) and obtained by AURA (red dashed line).

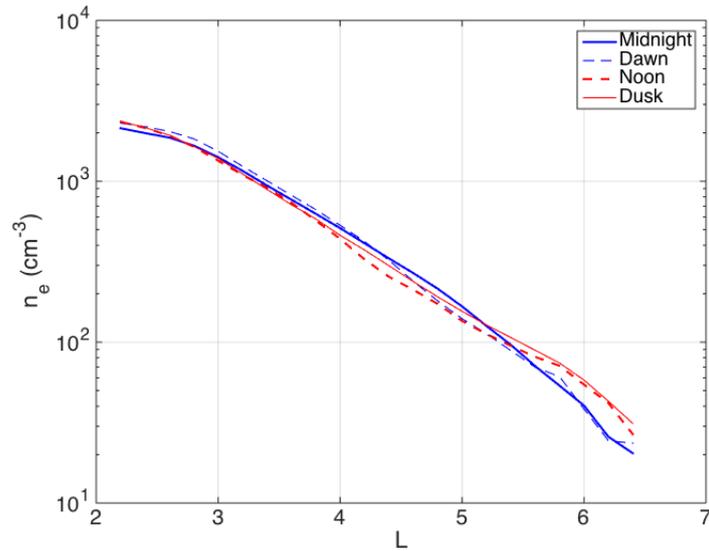


Figure 6-2: Average electron density  $n_e$  as determined from the Van Allen Probes mission data versus  $L$  using the proposed algorithm. The data is plotted for different MLT regions, 21 – 03 (thick solid blue curve), 03 – 09 (thin dashed blue curve), 09 – 15 (thick dashed red curve), and 15 – 21 (thin solid red curve).

## 6.2 Comparison with AURA

For the legitimate comparison with AURA, all available data was categorized into three types (A, B, C). This partition was described in full detail in Section 2.4, but just to remind the reader here is the brief summary of this division:

- Type A: 70% of the HFR orbit spectrograms successfully digitized into density records with less than 1/4 failure to determine the correct densities (visually deter-

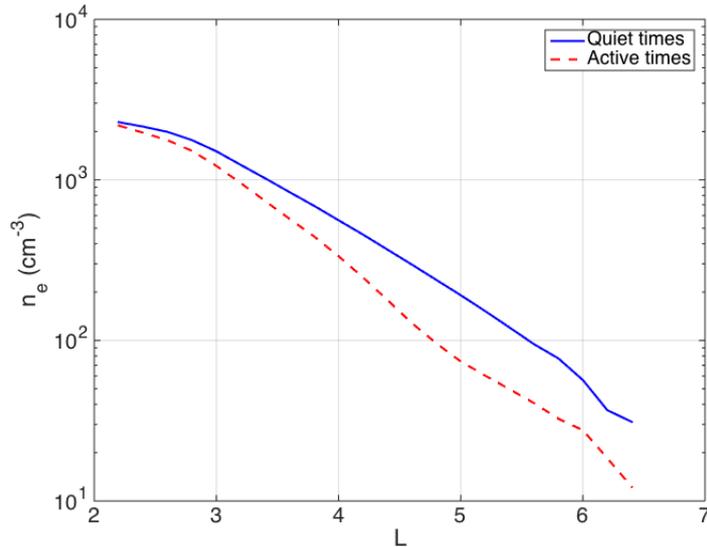


Figure 6-3: Average electron density  $n_e$  as determined from the Van Allen Probes mission data versus  $L$  using the proposed algorithm. Data is plotted for  $K_p < 1.2$  (solid blue curve) and  $K_p > 2$  (dashed red curve).

mined as needing correction).

- Type B: 20% of the data set that is contaminated by 1/4 to 1/2 misidentified records, but is remedied by manual inspection.
- Type C: 10% of data (which coincides with high geomagnetic activity) with failure  $> 1/2$  and concealed signal.

The distribution of the types obtained here is the same as above, i.e. Type A: 70%, Type B: 20%, Type C: 10%. Figure 6-4 shows the plot of the average divergence of the proposed algorithm from the AURA in % vs. data types. From this plot, one can see that the two algorithms yield identical results for orbits of type A since this type corresponds to orbits with frequency profiles that are the easiest to infer. Difference for type B is also not very significant, but difference for type C comprises 14 %. This is due the uncertainty of the upper hybrid frequency determination that appears during the geomagnetically active times when the electron densities are low. Discussion on this divergence in the results continues in the next section.

### 6.3 Results Evaluation

Examples of the work of the algorithm are presented in Figures 6-5 – 6-9. White curves are the  $f_{uhr}$  inferred by the algorithm described in this work. Red curves represents the

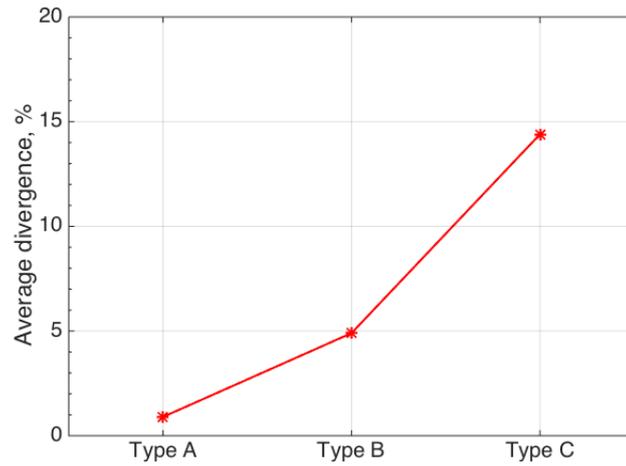


Figure 6-4: Average divergence of the electron density  $n_e$  prediction of the proposed algorithm and AURA as determined from the Van Allen Probes mission data versus types.

results obtained by AURA. Figures 6-5, 6-6 shows the sample results yielded for orbits of type A. As we can see from this plot red curve almost completely overlap with the white, which means that the results of the two algorithms are almost identical (the slight difference can be seen in the zoomed part of Orbit 1040 in Figure 6-7), and there are no difficulties in determining the resonance line by the proposed algorithm.

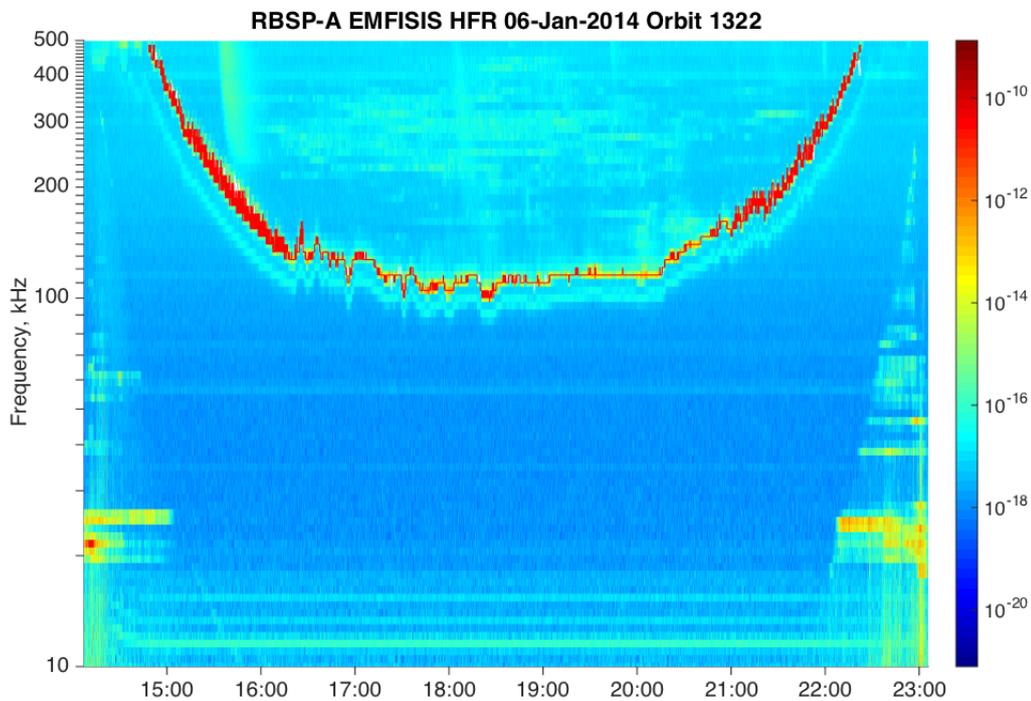


Figure 6-5: Here and thereon red curves represents the results obtained by AURA, white curves are  $f_{uhr}$  inferred by the algorithm described in the thesis. In this case, the results of both algorithms basically overlap (white line cannot be seen).

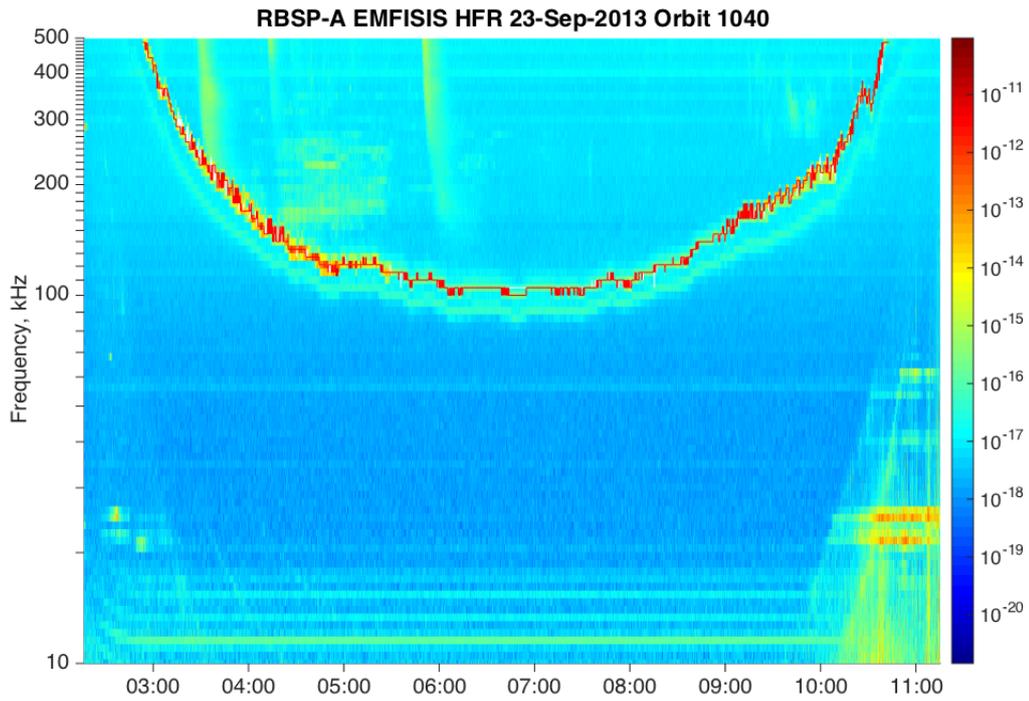


Figure 6-6: For this orbit, the results are almost the same. We can see a slight difference in derived upper hybrid frequencies in Figure 6-7 that is zoomed part of this orbit.

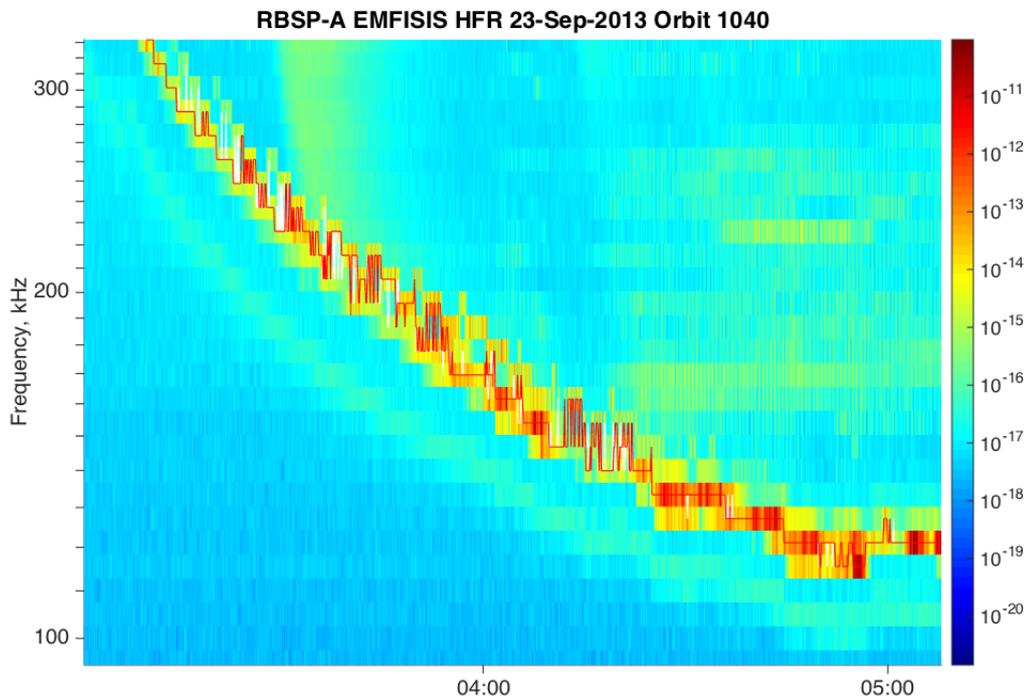


Figure 6-7: Zoomed part of orbit 1040 showing the difference in derived upper hybrid frequencies.

Figures 6-8, 6-9 show more complicated case of types B and C on the examples of orbits 100 and 105. Each case presented in the figure contains explanation in the caption.

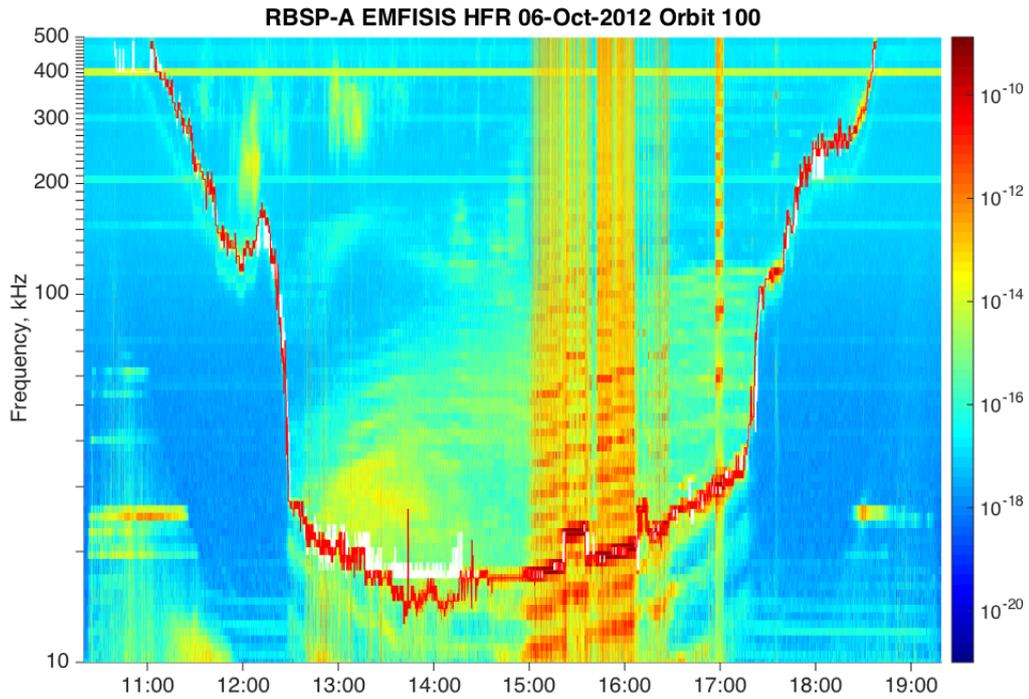


Figure 6-8: This is orbit of type B, which is of the moderate complexity for determining  $f_{uhr}$ . Here, we can see that the results almost identical except for the part from 12:45 until 14:20. Both algorithms shows instability in this region, and it is hard to judge visually which algorithm yields more precise results.

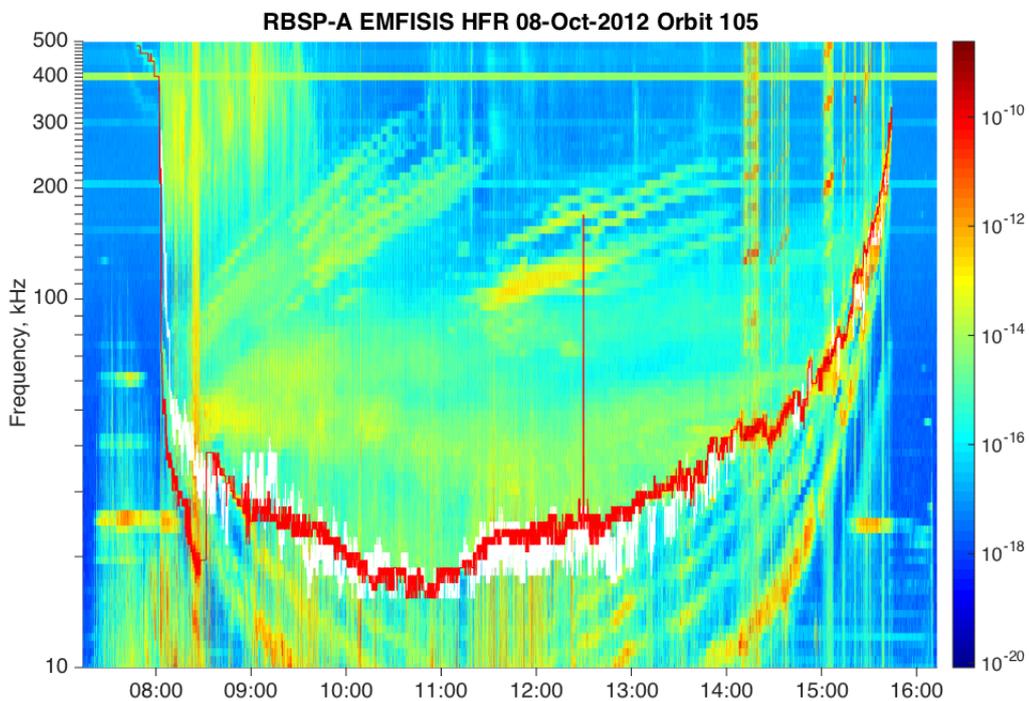


Figure 6-9: This example shows the orbit of type C, which is the most difficult type for determining  $f_{uhr}$ . We can see the difference of the results in the low density region, which represents the most difficult part of the orbit.

## 6.4 Application

Datasets of the electron densities inferred using the algorithm proposed in this work for both RBSP-a and RBSP-b spacecraft for all available orbits are available at

[http://bitly.com/RBSP\\_electron\\_densities](http://bitly.com/RBSP_electron_densities).

Data is presented as .mat files (MATLAB) with a separate file for each orbit. Each file contains the fields presented in Table 6.1. Data is time ordered.

Table 6.1: Data fields.

Field name	Dimensions	Units	Description
Epoch	$[1 \times n]$	<i>ns</i>	Default time
f_uhr	$[1 \times n]$	<i>Hz</i>	Upper hybrid frequency
n_e	$[1 \times n]$	$m^{-3}$	Plasma density
f_ce	$[1 \times n]$	<i>Hz</i>	Electron cyclotron frequency
L	$[1 \times n]$	Earth-radii	Magnetic field lines
Kp	$[1 \times n]$	Ranges from 0 to 9	Global geomagnetic storm index
MLT	$[1 \times n]$	Hours	Magnetic local time

# Chapter 7

## Conclusions and Further Work

### 7.1 Conclusions

In this thesis, the procedure of automated electron densities extraction from spectrograms of Waves HFR measurement was proposed and described. Also, the algorithm was compared with AURA - algorithm for semiautomated densities detection. Computational experiment on the Waves HFR data was conducted and it shows that the algorithms yield identical results on the orbits easy for upper hybrid resonance line detection ( $\approx 1\%$  divergence), and comparable results on the difficult orbits with concealed signal ( $\approx 14\%$  divergence). This means that the algorithm proposed in the thesis is a reliable automatic procedure that can be used as another alternative to derive electron densities from the data obtained from the Van Allen mission and possibly future missions. The dataset of electron densities obtained by this algorithm for all orbits available at the moment was constructed and put online.

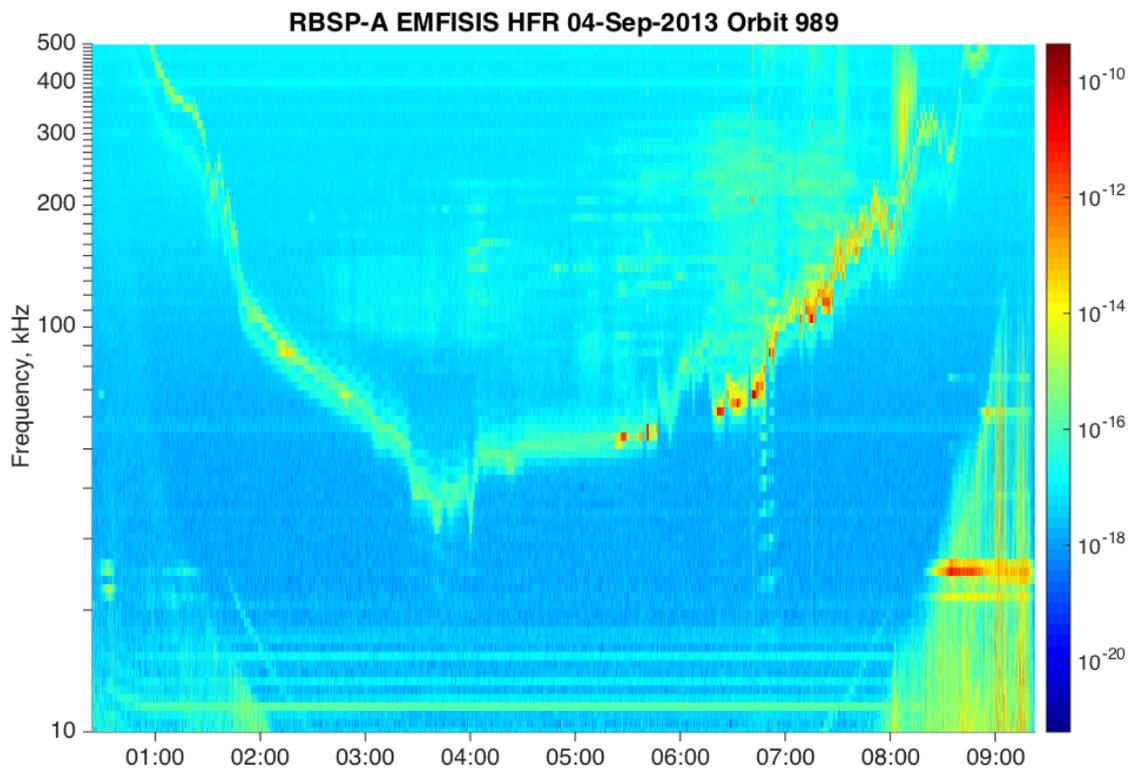
### 7.2 Further Work

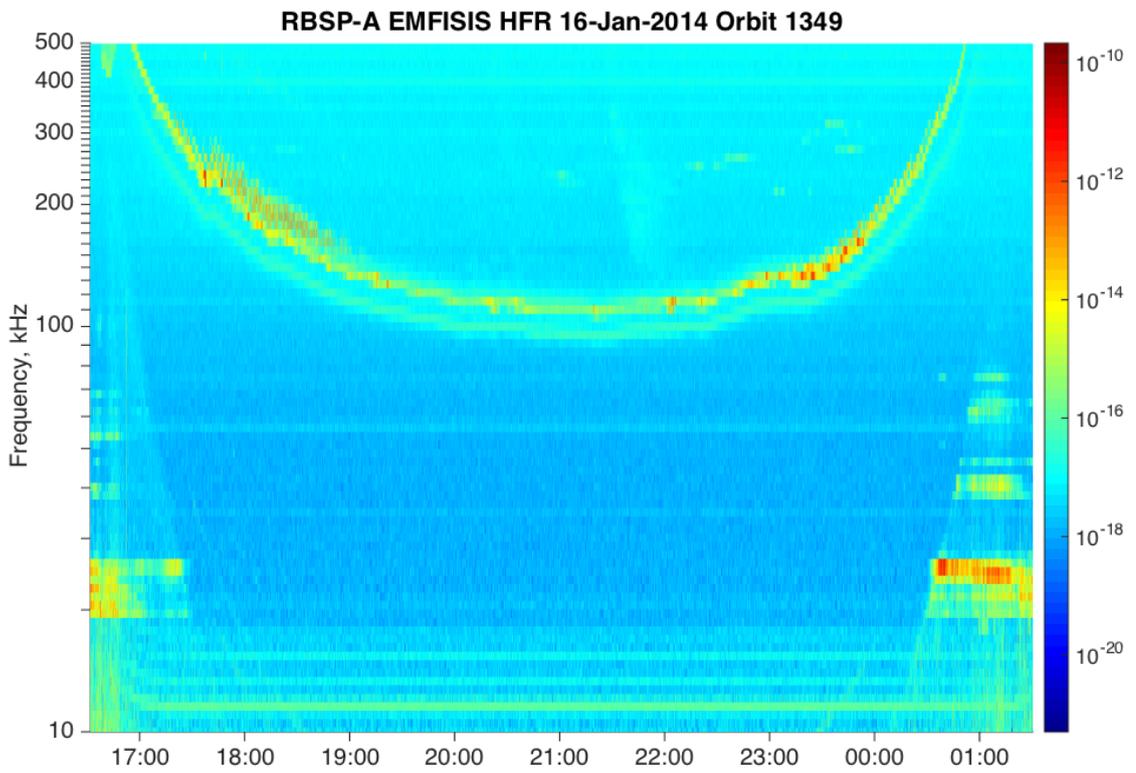
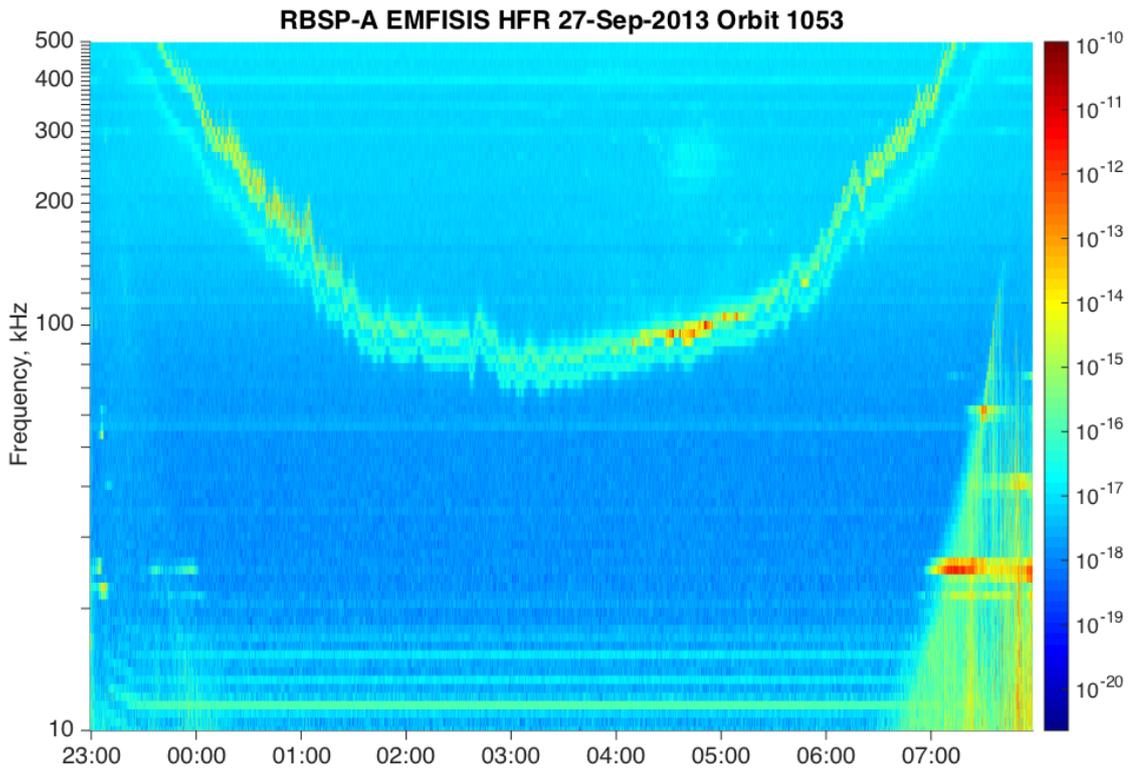
The procedure proposed in the thesis bases the determination of upper hybrid resonance line on neural networks. As future steps, we plan to try other machine learning algorithms for the upper hybrid resonance line detection. Also, we plan to apply the algorithm proposed on data from other missions, making necessary adjustments in the algorithm design, and based on that implement a system for electron densities derivation that can be used over different datasets automatically.

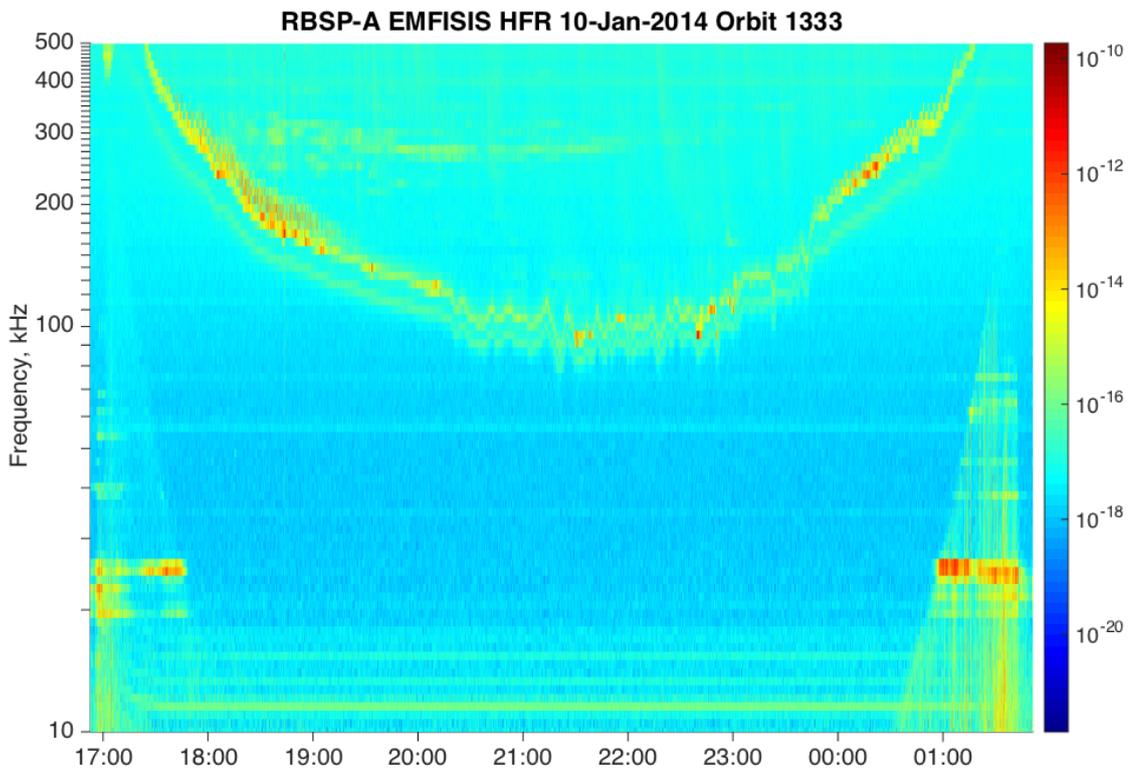
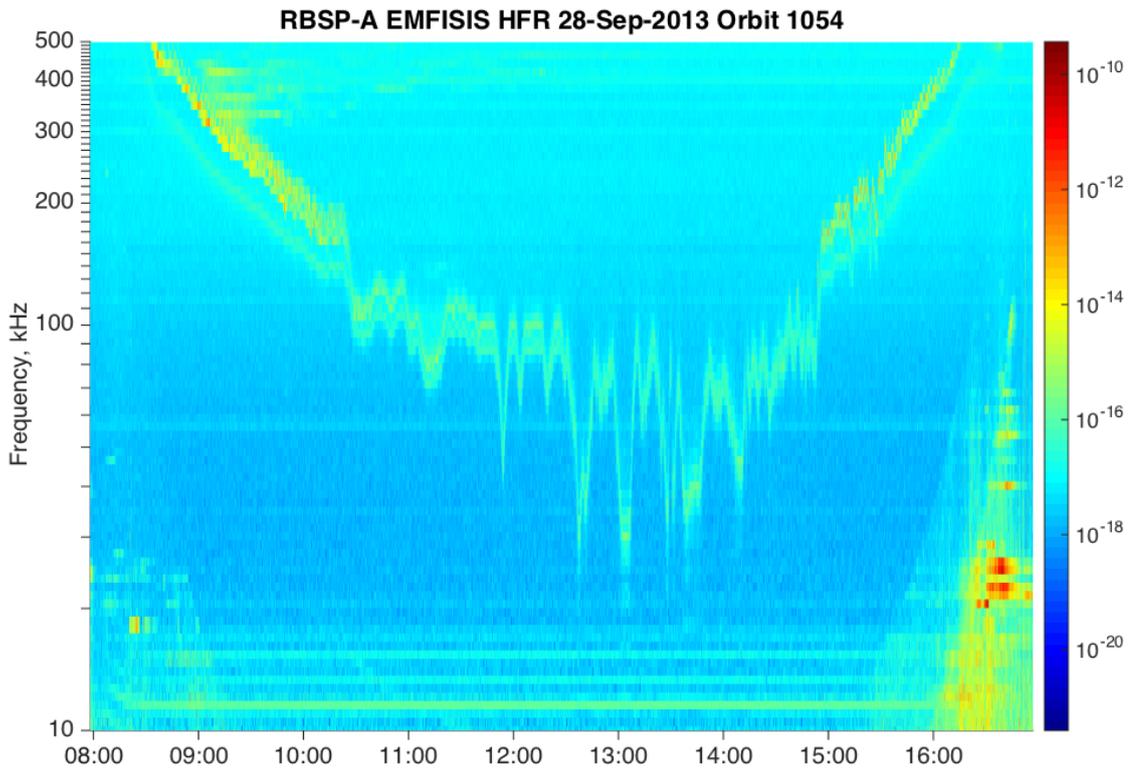
# Appendix A

## Data Types

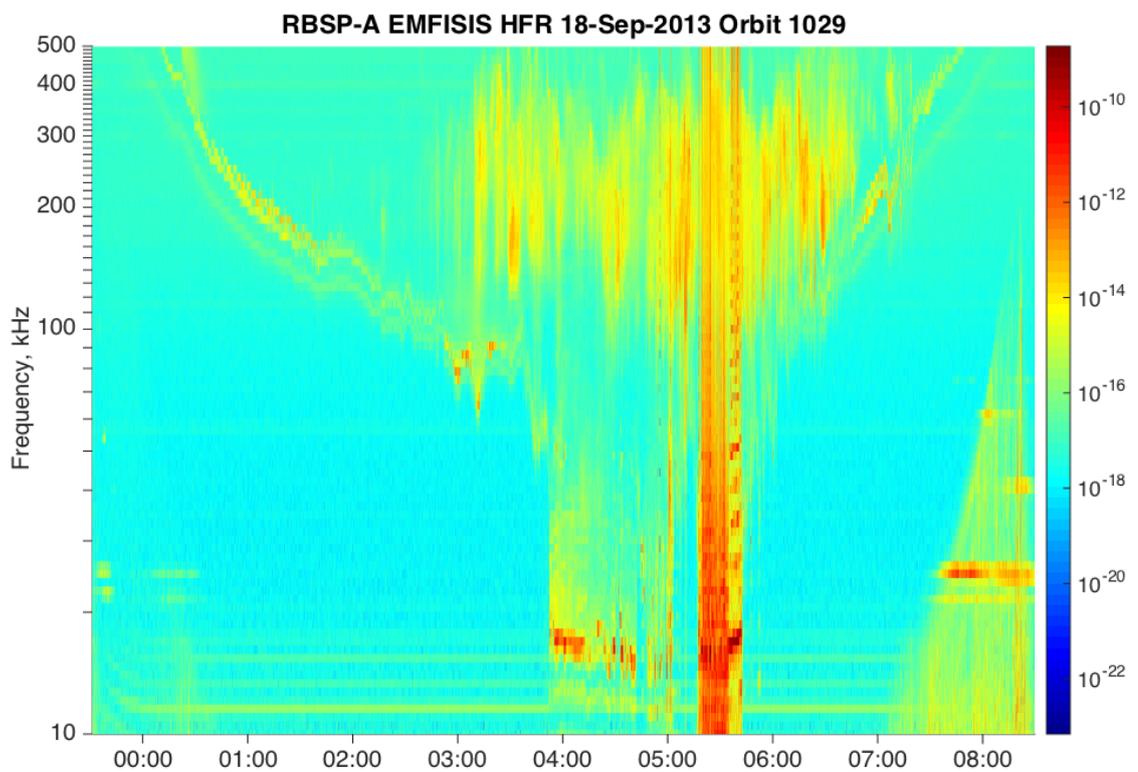
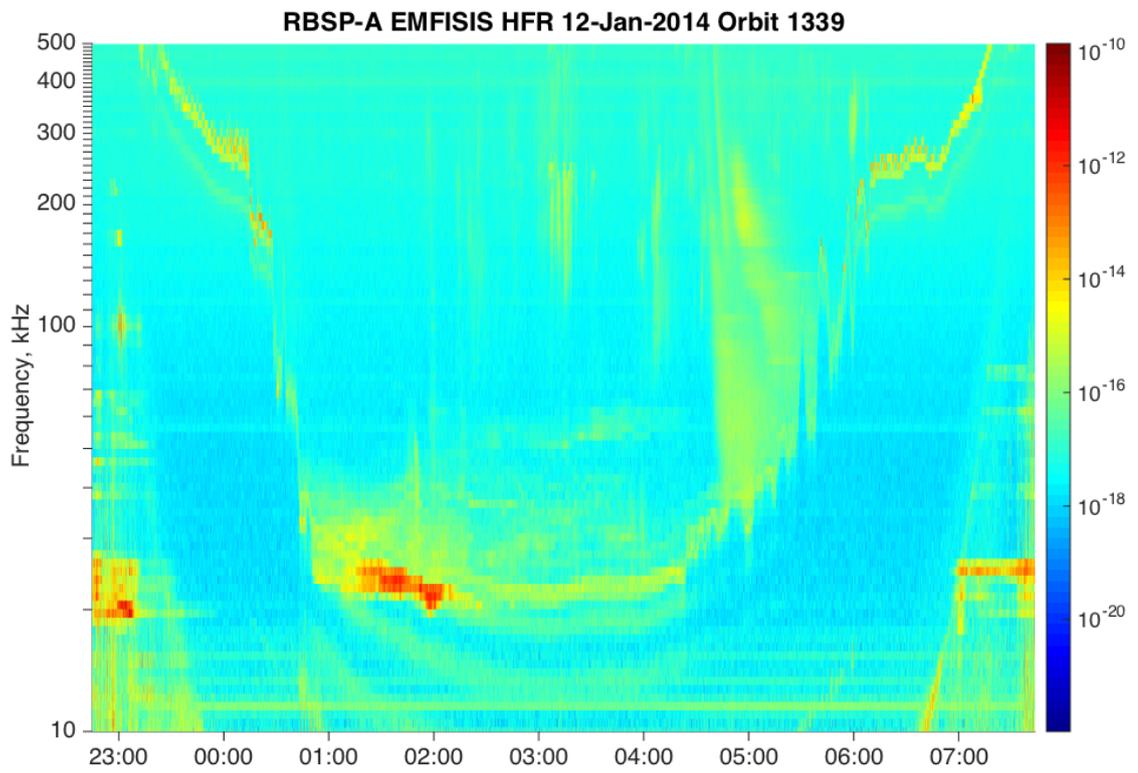
### Examples of Type A



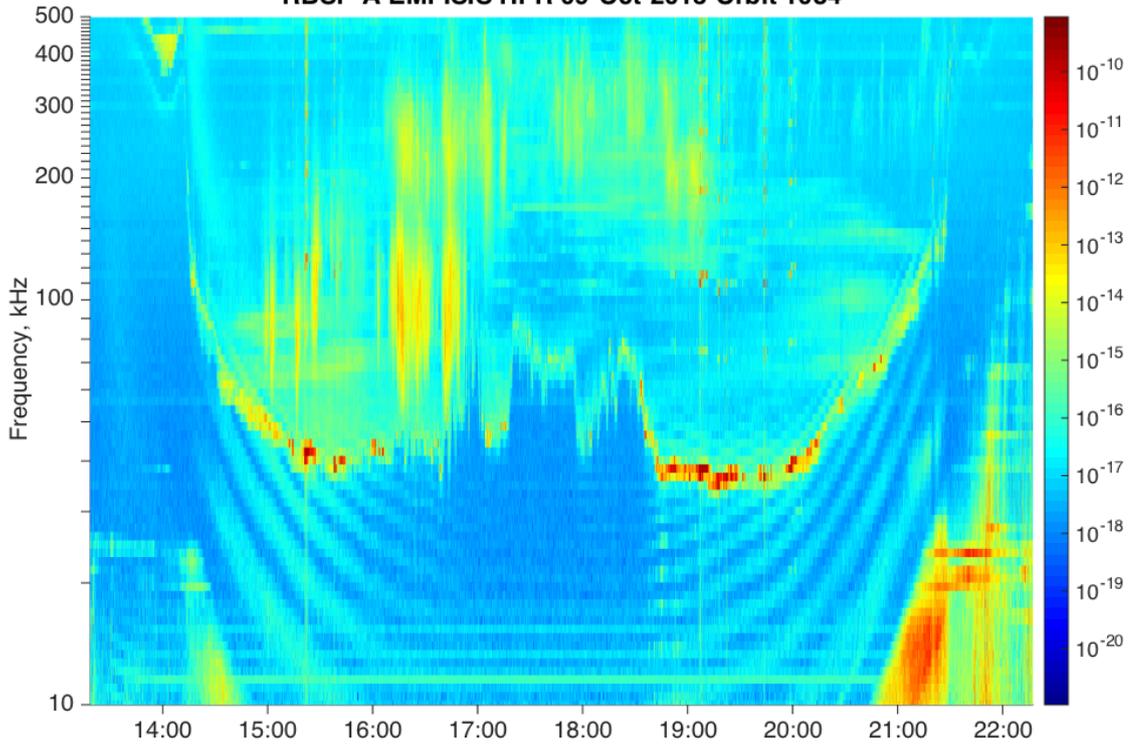




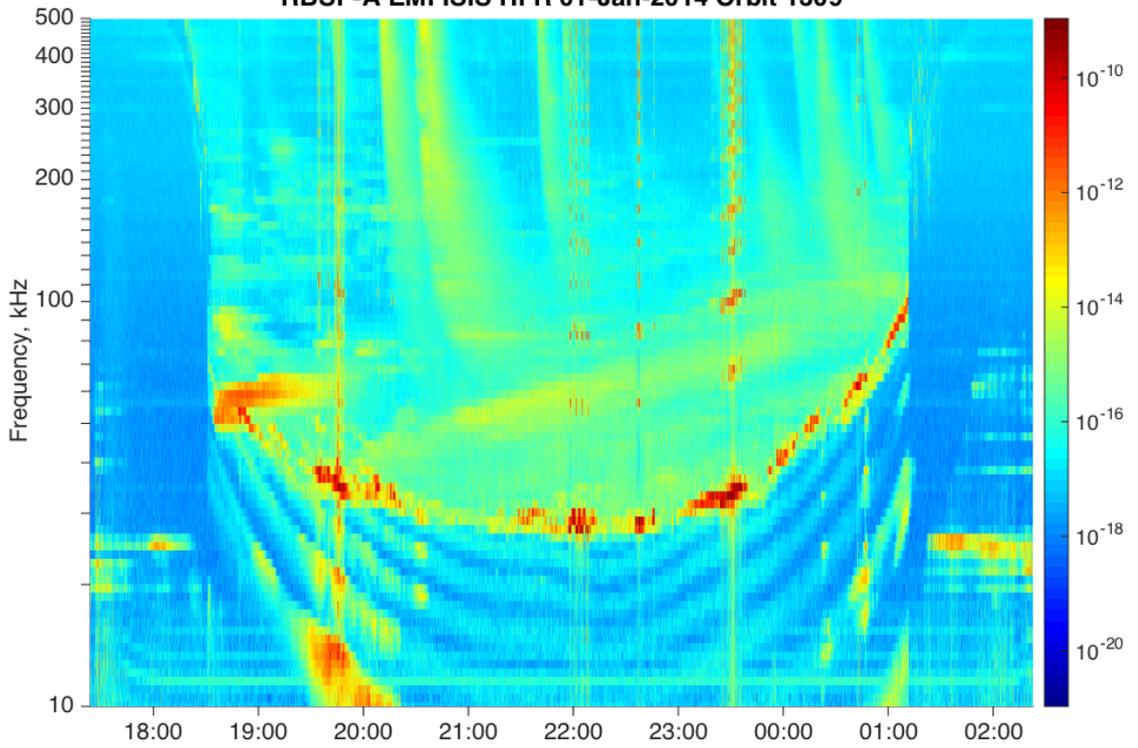
## Examples of Type B



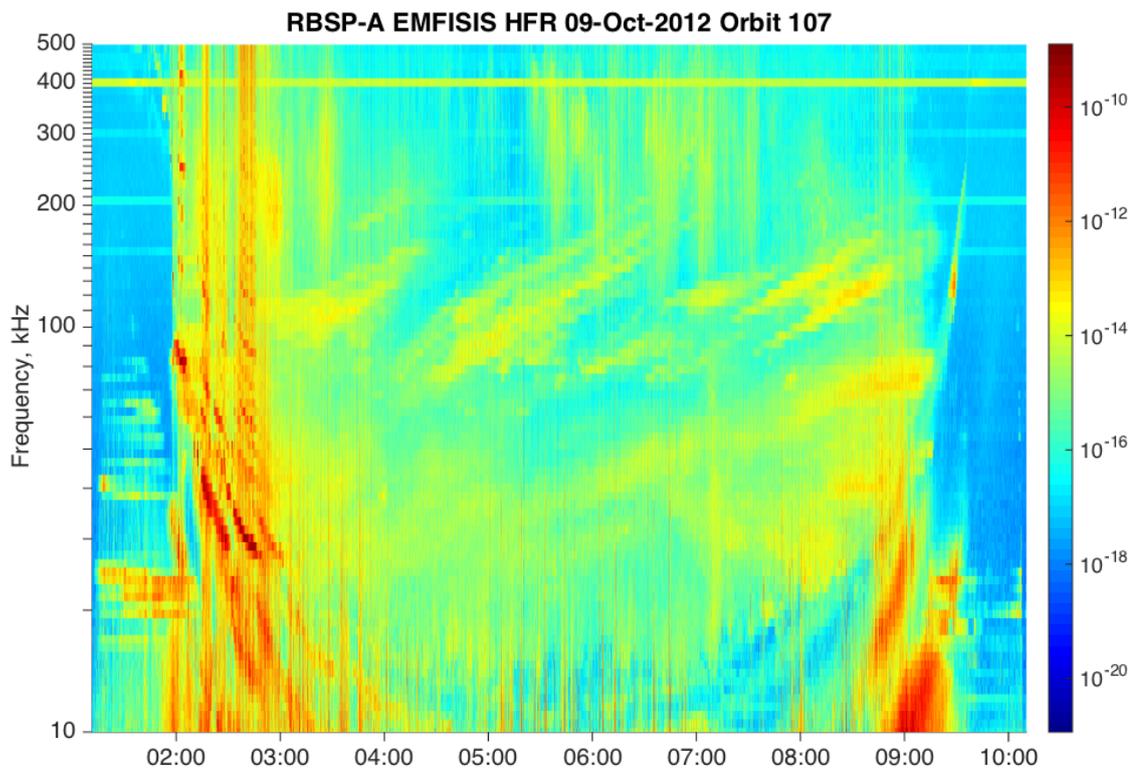
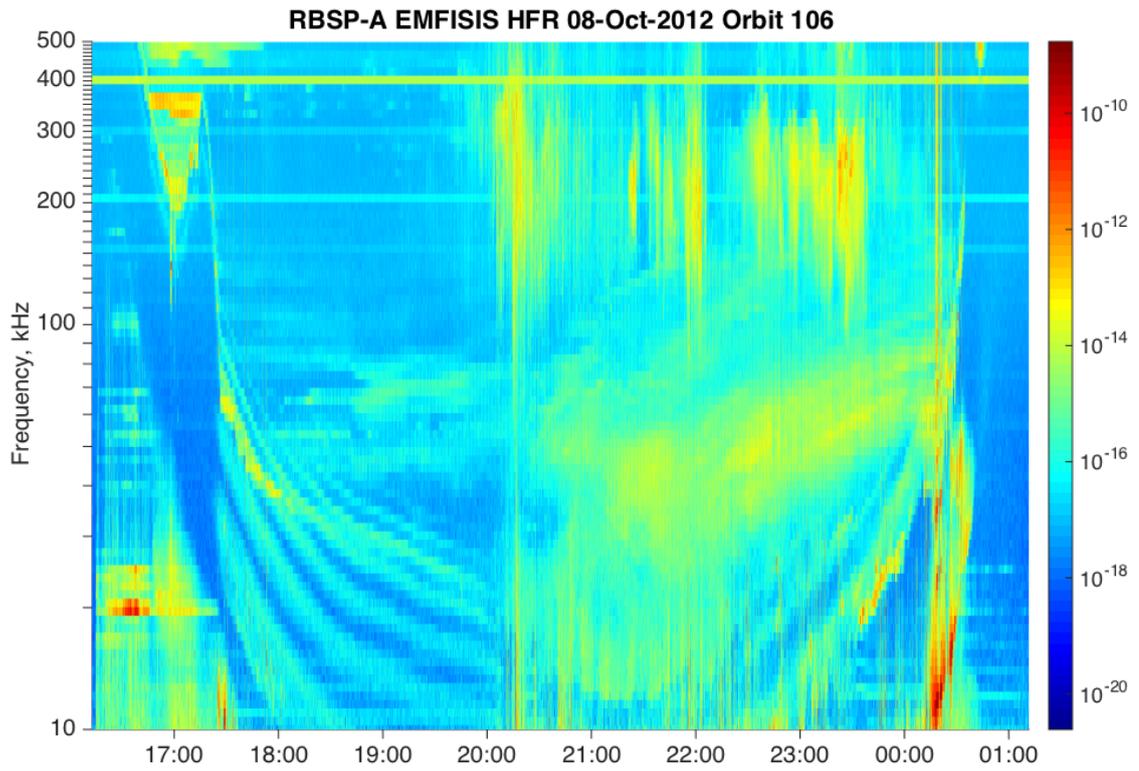
**RBSP-A EMFISIS HFR 09-Oct-2013 Orbit 1084**



**RBSP-A EMFISIS HFR 01-Jan-2014 Orbit 1309**



## Examples of Type C



# Bibliography

- [1] Kurth, W. S., S. De Pascuale, J. B. Faden, C. A. Kletzing, G. B. Hospodarsky, S. Thaller, and J. R. Wygant. Electron densities inferred from plasma wave spectra obtained by the Waves instrument on Van Allen Probes. *J. Geophys. Res. Space Physics*, 120(doi:10.1002/2014JA020857), 2015.
- [2] NASA Press Kit. <http://www.nasa.gov/>.
- [3] RBSP Instrument Information. <http://www.spaceflight101.com/rbsp-instrument-information.html>.
- [4] Wikipedia. [http://en.wikipedia.org/wiki/plasma\\_\(physics\)](http://en.wikipedia.org/wiki/plasma_(physics)).
- [5] M. P. Pandya, H. D. Kabariya, S. K. Karkari, H.K Patel. Survey and comparative analysis of different techniques for electron density measurements in low temperature plasma. *2013 Nirma University International Conference on Engineering (NUiCONE)*, (10.1109/NUiCONE.2013.6780191):1 – 6, Nov. 2013.
- [6] Sung-Ho Jang, Se-Jin Oh, Young-Kwang Lee, Chin-Wook Chung. Real time feedback control of plasma density using a floating probe in semiconductor processing. *Current Applied Physics* 13, (76e79), 2013.
- [7] H. D. Kabariya. Automation of microwave source for diagnosis of electron density in plasma using labview. Master’s thesis, Institute for Plasma Research, Bhat, Gandhinagar, May 2013.
- [8] Wikipedia. [http://en.wikipedia.org/wiki/van\\_allen\\_probes](http://en.wikipedia.org/wiki/van_allen_probes).
- [9] Van Allen Probes Nasa. <http://iphone22.arc.nasa.gov/public/iexplore/missions/pages/96.htm>.
- [10] R. L. Kessel. NASA’s Radiation Belt Storm Probes Mission: From Concept to Reality. *American Geophysical Union*, (10.1029/2012GM001312), 2012.
- [11] University of Calgary. <http://www.ucalgary.ca/above/science/chorus>.
- [12] Wikipedia. [http://en.wikipedia.org/wiki/upper\\_hybrid\\_oscillation](http://en.wikipedia.org/wiki/upper_hybrid_oscillation).
- [13] C.A. Kletzing, W.S. Kurth, M. Acuna, R.J. MacDowall, R.B. Torbert, T. Averkamp, D. Bodet, S.R. Bounds, M. Chutter, J. Connerney, D. Crawford, J.S. Dolan, R. Dvorsky, G.B. Hospodarsky, J. Howard, V. Jordanova, R.A. Johnson, D.L. Kirchner, B. Mokrzycki, G. Needell, J. Odom, D. Mark, R. Pfaff Jr., J.R. Phillips, C.W. Piker, S.L. Remington, D. Rowland, O. Santolik, R. Schnurr, D. Sheppard, C.W. Smith, R.M. Thorne, J. Tyler. The electric and magnetic field instrument suite and integrated science (emfisis) on rbsp. *Space Sci Rev*, 179(10.1007/s11214-013-9993-6):127–181, June 2013.

- [14] The EMFISIS instrument suite description. <http://emfisis.physics.uiowa.edu/>.
- [15] Meyer-Vernet, N. and Perche, C. Tool kit for antennae and thermal noise near the plasma frequency. *J. Geophys. Res.*, 94:2405–2415., 1989.
- [16] C. P. Escoubet, A. Pedersen, R. Schmidt, P. A. Lindqvist. Density in the magnetosphere inferred from ISEE 1 spacecraft potential. *J. Geophys. Res.*, 102:17595–17609, 1997.
- [17] R. E. Denton, J. Goldstein, and J. D. Menietti. Field line dependence of magnetospheric electron density. *Geophys. Res. Lett.*, 29(24)(2205, doi:10.1029/2002GL015963), 2002.
- [18] R. F. Benson, P. A. Webb, J. L. Green, L. Garcia, B. W. Reinisch. Magnetospheric electron densities inferred from upper hybrid band emissions. *Geophys. Res. Lett.*, 31(10.1029/2004GL020847), 2004.
- [19] S. R. Mosier, M. L. Kaiser, and L. W. Brown. Observations of noise bands associated with the upper hybrid resonance by the Imp 6 radio astronomy experiment. *J. Geophys. Res.*, 78:1673–1679, 1973.
- [20] Dr. Youmin Tang. <http://web.unbc.ca/~ytang/chapter6.pdf>.
- [21] Wikipedia. [http://en.wikipedia.org/wiki/feedforward\\_neural\\_network](http://en.wikipedia.org/wiki/feedforward_neural_network).
- [22] MachineLearning.ru. [http://www.machinelearning.ru/wiki/index.php?title=Feedforward\\_neural\\_network](http://www.machinelearning.ru/wiki/index.php?title=Feedforward_neural_network)
- [23] Salakhutdinov, R. and Hinton, G. Semantic hashing. *International Journal of Approximate Reasoning*, 50(7):969–978, 2009.
- [24] Glorot, X., Bordes, A., and Bengio, Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. *ICML-2011*, 2011.
- [25] Krizhevsky, A. and Hinton, G. Using very deep autoencoders for content-based image retrieval. *European Symposium on Artificial Neural Networks ESANN-2011*, 2011.
- [26] Mohamed, A., Dahl, G., and Hinton, G. Acoustic modeling using deep belief networks. *Audio, Speech, and Language Processing, IEEE*, 20(1):14–22, 2012.
- [27] CDF Home Page. <http://cdf.gsfc.nasa.gov/>.
- [28] Neural Network FAQ. <ftp://ftp.sas.com/pub/neural/faq.html>.
- [29] Hongxing Li, C.L. Philip Chen, Han-Pang Huang. *Fuzzy Neural Intelligent Systems: Mathematical Foundation and the Applications in Engineering*. CRC Press, 2000.
- [30] Matlab Neural Networks Toolbox. <http://www.mathworks.com/products/neural-network/>.
- [31] Matlab Neural Networks Toolbox. Choose a Multilayer Neural Network Training Function. <http://www.mathworks.com/help/nnet/ug/choose-a-multilayer-neural-network-training-function.html>.