

# Специальные методы обработки двумерных данных об электронной плотности плазмы

Желавская Ирина

Московский физико-технический институт  
Факультет управления и прикладной математики  
Кафедра интеллектуальных систем

Научные руководители д.ф.-м.н. Ю. Шприц, д.ф.-м.н. М. Спасоевич

Москва,  
2015 г.

## Задача

Разработать алгоритм автоматического определения электронной плотности плазмы при помощи измерений параметров электромагнитных волн в плазме, сделанных на спутниках-близнецах Ван Аллена миссии NASA (Van Allen Probes).

## В работе предложена

Процедура, использующая нейронные сети прямого распространения для решения задачи нелинейной регрессии для определения электронной плотности плазмы.

## Новизна

Данное исследование – одна из первых попыток применения алгоритмов машинного обучения к данным космической физики.

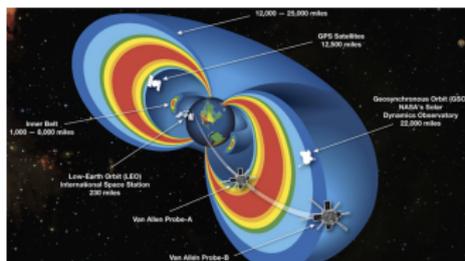
# Предметная область: миссия NASA по исследованию радиационных поясов Земли

## Цель миссии

Изучение поясов ван Аллена (или радиационных поясов) и предоставление наиболее подробных данных об этих поясах, когда-либо нам доступных.

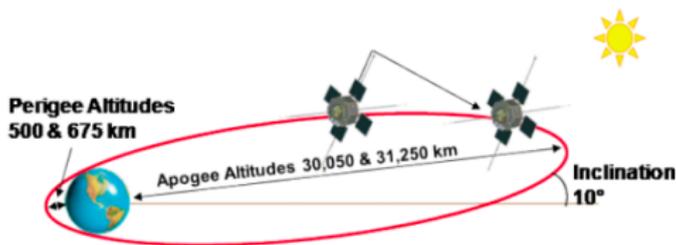
## Плотность плазмы

Плотность плазмы - фундаментальный параметр плазмы, необходимый для моделирования и анализа электромагнитных волн нестабильных явлений в плазмосфере. Во всех предыдущих миссиях определялась вручную. С ростом объемов и разрешения данных, для данной миссии такой способ определения плотности – очень затруднителен.

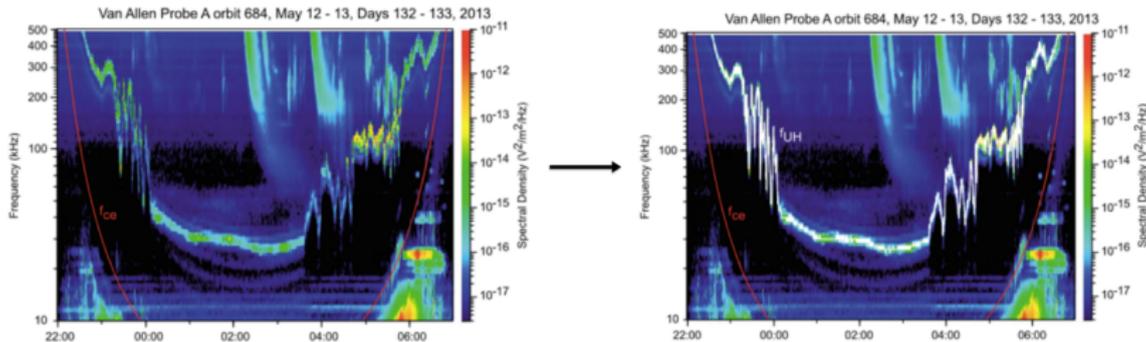


## Факты о миссии

- Два спутника
- Идентичные орбиты, период – 9 часов
- Длительность миссии – более 2х лет (запущены 30 августа 2012 года)
- Набор из 5 инструментов на борту (инструменты позволяют собирать данные об электро-магнитном поле, о составе поясов в высоком разрешении)



## Определение электронной плотности с помощью интенсивных верхнегибридных излучений на спектрограммах:



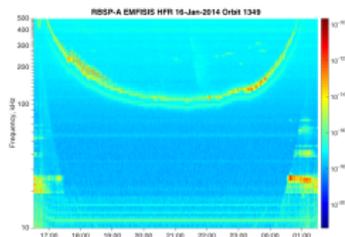
- Верхняя гибридная частота:  $f_{uhr} = \sqrt{(f_{ce}^2 + f_{pe}^2)}$

- Электронная циклотронная и плазменная частоты:

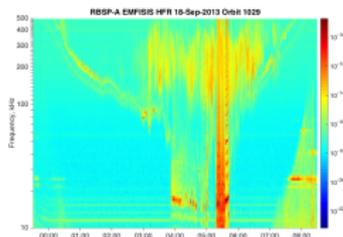
$$f_{ce} = \frac{|q_e|B}{2\pi m_e}, \quad f_{pe} = \frac{1}{2\pi} \sqrt{\frac{q_e^2 n_e}{m_e \epsilon_0}}$$

- Плотность плазмы:  $n_e = \frac{f_{uhr}^2 - f_{ce}^2}{8980^2}$

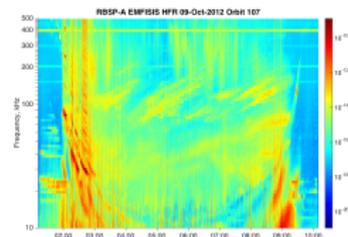
Деление данных на 3 типа было введено в работе Kurth et al., *Electron densities inferred from plasma wave spectra obtained by the Waves instrument on Van Allen Probes*, J. Geophys. Res., 120, 904-919, 2015.



Тип А: 70% орбит  
визуально ясный сигнал



Тип В: 20% орбит  
необходима  
интерпретация сигнала



Тип С: 10% орбит  
скрытый сигнал

## Методы определения верхнегибридной резонансной частоты

- 1 Определение верхнегибридной частоты вручную на спектрограммах;
- 2 AURA – Automated Upper-hybrid Resonance Detection Algorithm.

## AURA

- 1 Публикация: Kurth et al., *Electron densities inferred from plasma wave spectra obtained by the Waves instrument on Van Allen Probes*, J. Geophys. Res., 120, 904-919, 2015;
- 2 Модель основана на правиле гистерезиса: каждый последующий пик спектра, соответствующий  $f_{uhr}$ , находится рядом с пиком, определенным на предыдущем шаге;
- 3 Неудача  $F$  – необходимость исправления неправильно определенных точек вручную;
- 4 Вероятность неудачи для орбиты:
  - Типа А:  $Pr_F < 25\%$
  - Типа В:  $25\% < Pr_F < 50\%$
  - Типа С:  $50\% < Pr_F$

# Постановка задачи предлагаемого метода

## Дано

- Матрица признаков  $\mathbf{X} \in \mathbb{R}^{M \times N}$
- Вектор целевых переменных  $\mathbf{y} \in \mathbb{R}^M$

## Требуется построить

Нелинейную регрессионную модель  $f : (\mathbf{w}, \mathbf{X}) \mapsto \mathbf{y} \in \mathbb{R}^M$ ,  
где  $\mathbf{w}$  – вектор параметров модели.

## Необходимо найти

Оптимальный набор параметров:

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

где функция потерь:

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

## Дано

- Набор измерений (данные со спутников)  $\mathfrak{S} = \{\mathfrak{s}_1, \dots, \mathfrak{s}_M\}$
- Словарь, содержащий экспертные знания о данных  $V = V(\mathfrak{S})$
- Набор экспертно-заданных примитивных функций  $G = \{g_1, \dots, g_N\}$ , таких что:

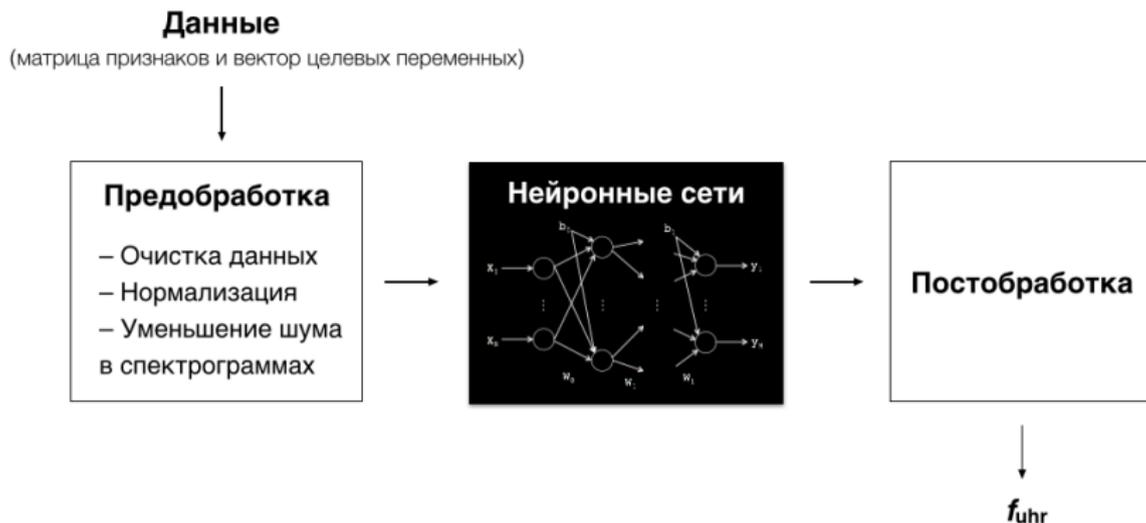
$$g_j : (\mathbf{b}_j, \mathfrak{s}_i, V) \mapsto x_{ij} \in \mathbb{R}^1$$

## Требуется построить

Матрицу признаков:

$$\mathbf{X} = \{x_{ij}\}, \quad i = \{1, \dots, M\}, j = \{1, \dots, N\}.$$

# Предлагаемый метод



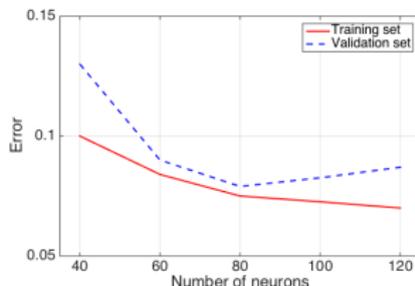
# Постановка эксперимента

## Данные

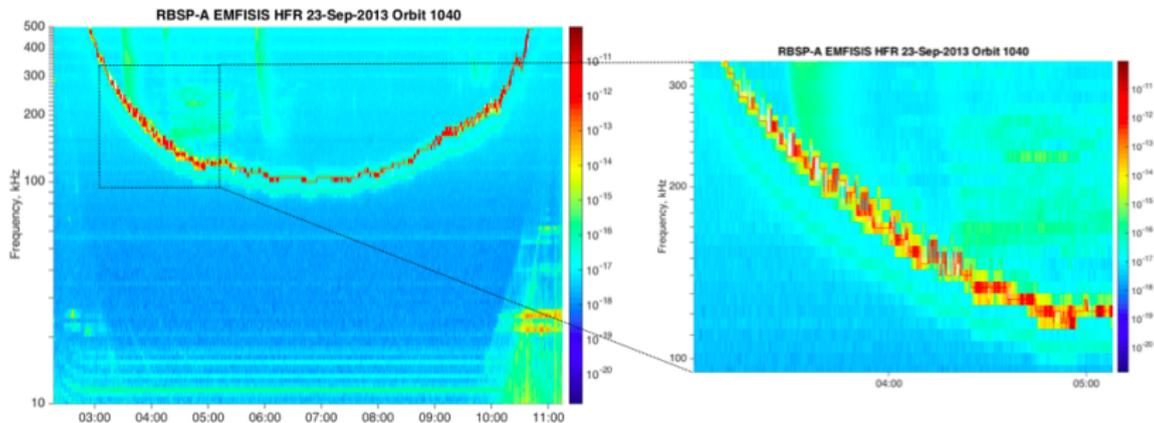
- Размер набора данных: 4,077,933 точки (после предобработки)
- Отношение деления данных на обучающее - валидационное - тестовое подмножества: 0.34 : 0.33 : 0.33
- Архитектура нейронной сети (НС): НС прямого распространения с 1 скрытым слоем, содержащим 80 нейронов (модель выбрана на валидационном множестве)

## Валидация

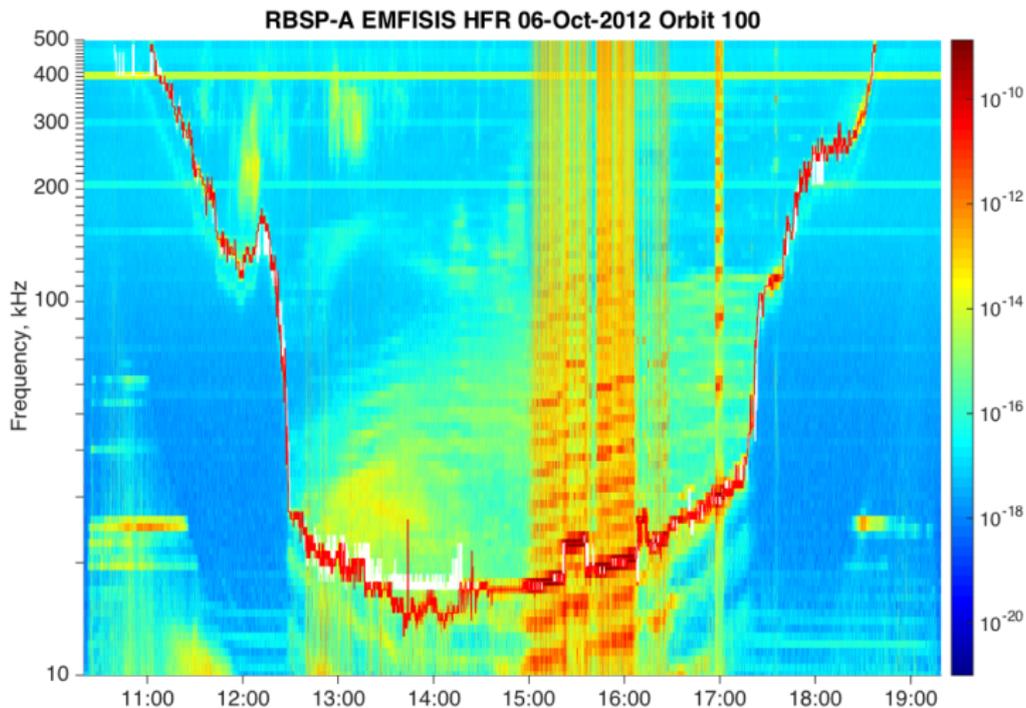
Зависимость ошибки моделей от количества нейронов в скрытом слое для обучающего и валидационного множеств



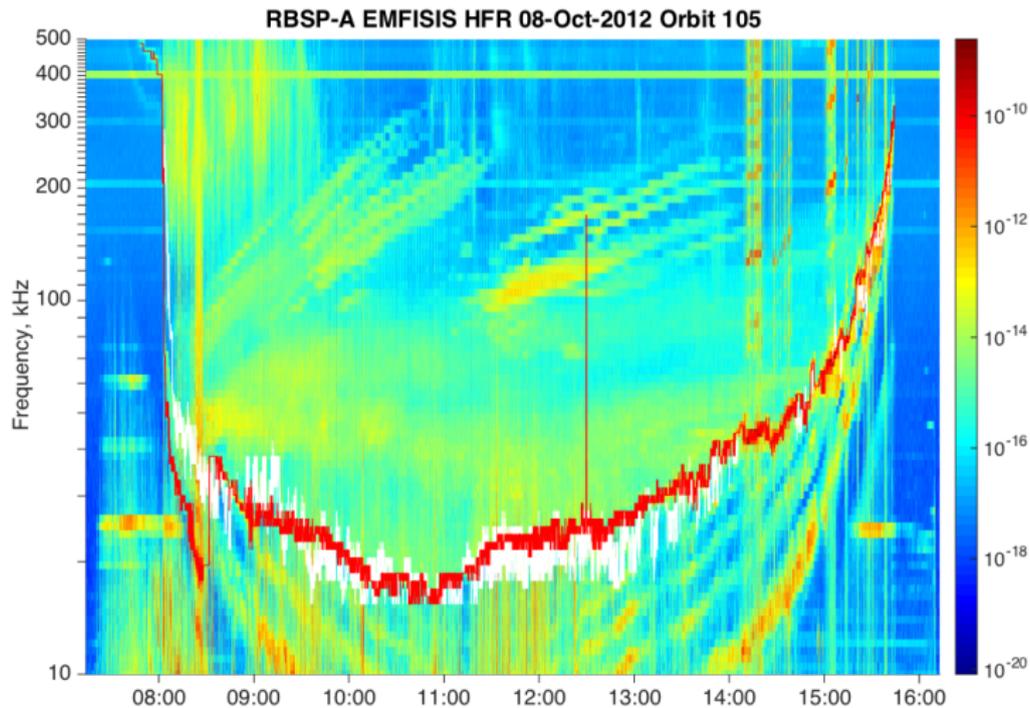
# Примеры определения $f_{uhr}$ : Тип А



# Примеры определения $f_{uhr}$ : Тип В



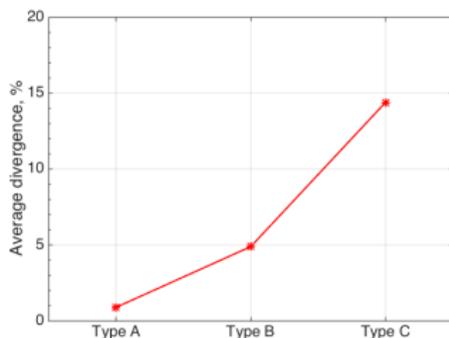
# Примеры определения $f_{uhr}$ : Тип С



Вероятность неудачи (необходимости ручного исправления) для орбиты:

| Метод | Тип А         | Тип В                | Тип С         |
|-------|---------------|----------------------|---------------|
| AURA  | $Pr_F < 25\%$ | $25\% < Pr_F < 50\%$ | $Pr_F > 50\%$ |

Среднее расхождение значений электронной плотности  $n_e$ , определенным с помощью предложенного метод и алгоритма AURA (на тестовом подмножестве)



- 1 Предложена процедура автоматического определения электронной плотности плазмы по измерениям параметров электромагнитных волн в плазме, сделанных на спутниках-близнецах Ван Аллена миссии NASA (Van Allen Probes).
- 2 Проведены вычислительные эксперименты.
- 3 Произведено сравнение предложенной процедуры с алгоритмом AURA. Показано, что предложенный алгоритм – надежная автоматическая процедура, которая может быть использована в качестве альтернативы для определения плотностей плазмы по данным миссии NASA Van Allen Probes и будущих миссий.
- 4 Разработанный метод позволит разработать эмпирические модели плотности плазмы в условиях, когда данные не доступны.

# Features description

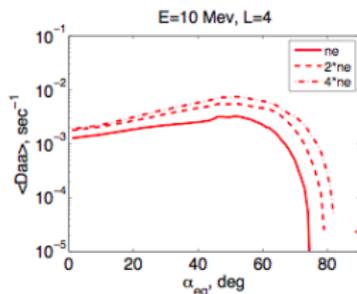
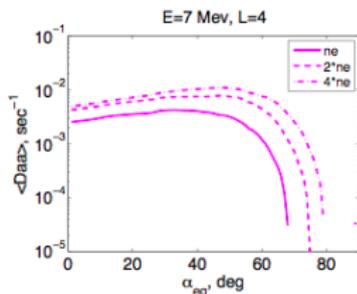
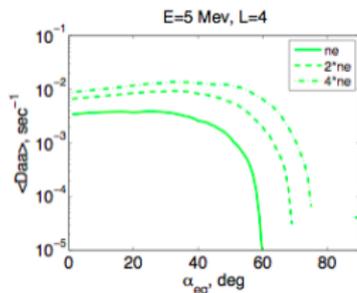
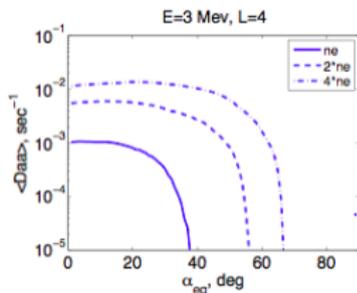
$M \sim 5,100,000$ :

Each spectrogram comprises  $\sim 5,000$  spectra (6 sec cadence of the EMFISIS instrument). Labeled data are available for 1,091 out of 2,425 orbits.

| 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 | $\{1, 2, \dots, 82\}$     |

# Sensitivity to plasma density\*

Bounce-averaged pitch angle diffusion coefficients computed by the oblique EMIC code for a set of ambient plasma densities



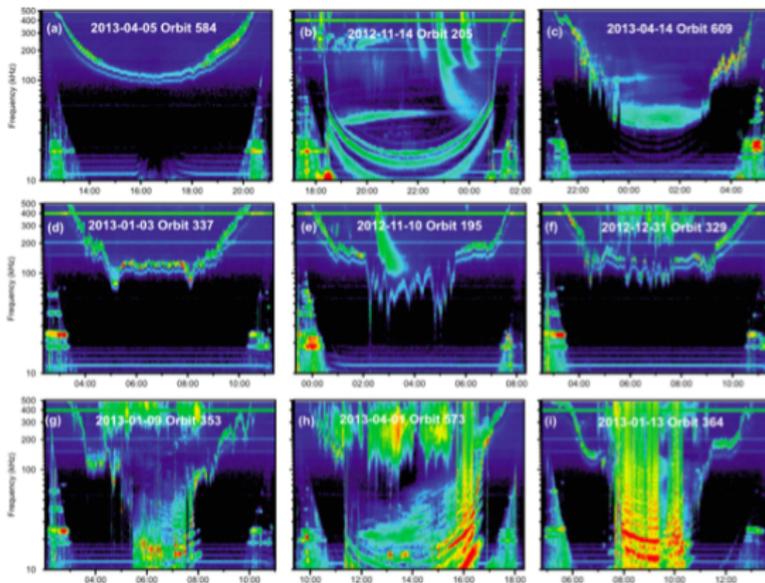
\* Courtesy of Bogdan Uzbekov

## Motivation

- Radiation belts are hazardous for spacecraft and its crew;
- High geomagnetic activity may be hazardous for people and radio electronic devices on Earth;
- The electron plasma density is required for analysis of wave propagation in plasma and instability growth rates.

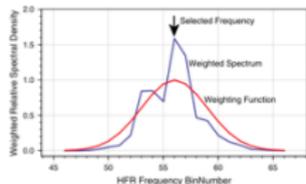
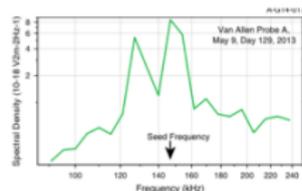
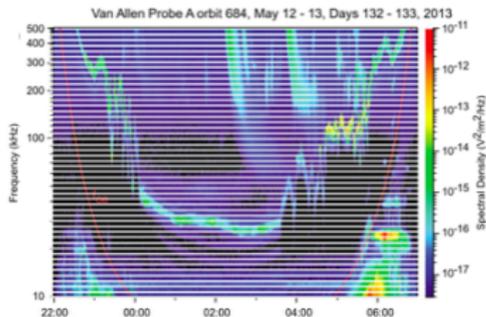
# Data examples

Spectrograms obtained from the EMFISIS instruments suite (Electric and Magnetic Field Instrument Suite and Integrated Science)

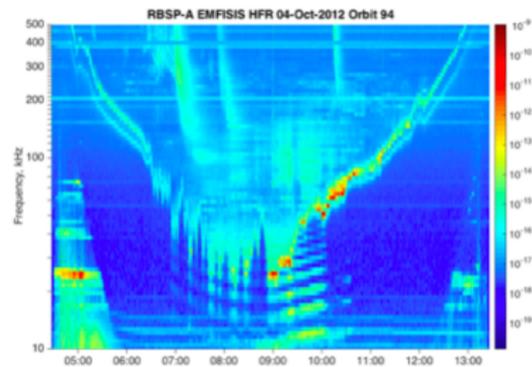
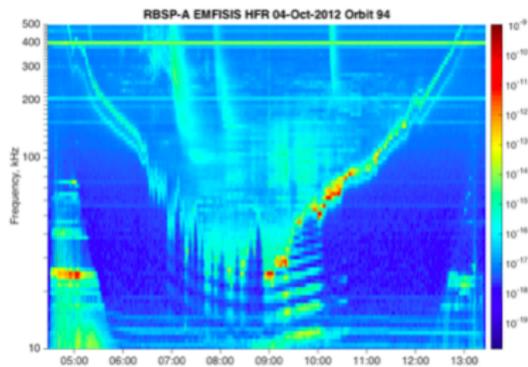


# AURA – Automated Upper-hybrid Resonance detection Algorithm

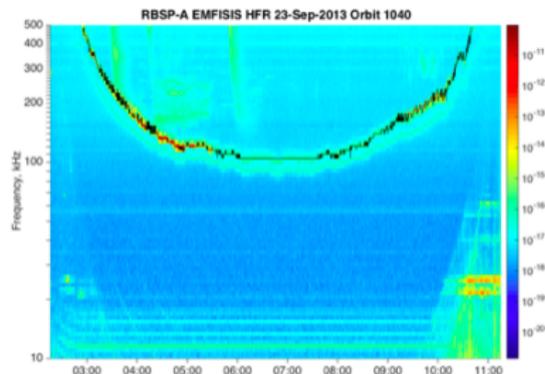
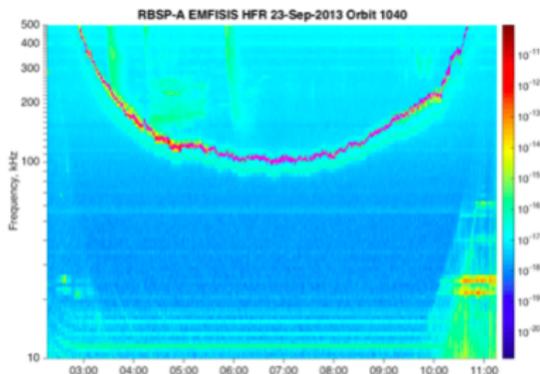
- Model is based on the *rule of hysteresis*: each successive spectrum contains a peak frequency associated with  $f_{uh}$  near the previously identified peak (i.e., there is not a large change in frequency of the binned signal within a 6 s sampling time step);
- 82 log-spaced frequency bins, 5% spectral resolution ( $\Delta f/f$ ), 10% density resolution ( $\Delta n/n$ );
- Data points to be manually corrected for a given orbit:
  - Type A (70% of orbits): failure < 25%
  - Type B (20% of orbits): 25% < failure < 50%
  - Type C (10% of orbits): failure > 50%



## Noise removal



## Postprocessing procedure



# Neural network design workflow

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.

<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.