Predicting Ionospheric Dynamics At Low Latitudes Using Neural Networks:
Applications to Ionogram and Spread-F Forecasting

Enrique Rojas
Cornell University

Marco Milla
Pontificia Universidad Catolica del Peru

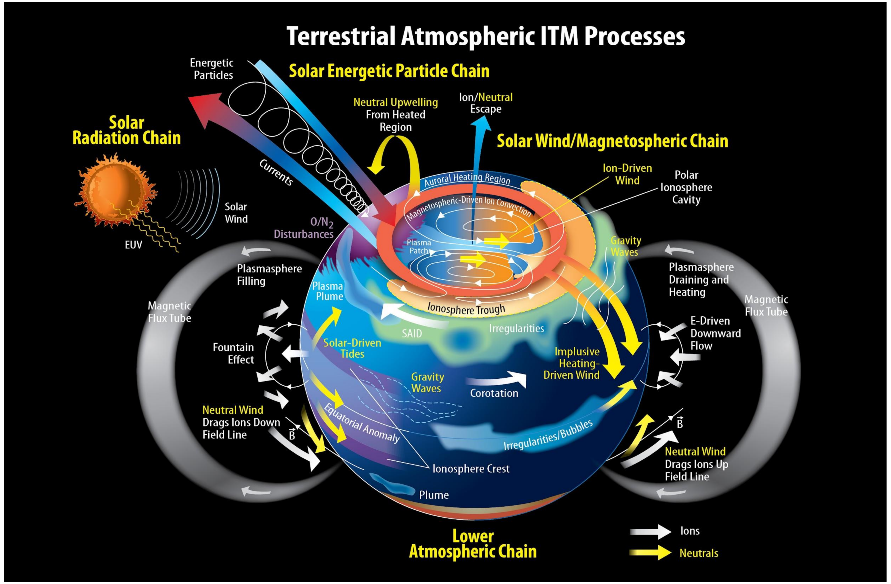
Jhassmin Aricoche
Jicamarca Radio Observatory

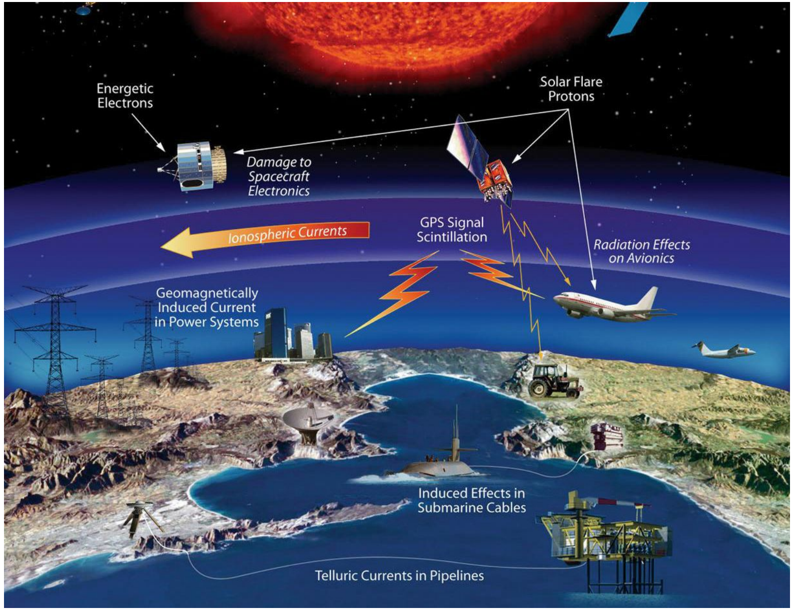
Reynaldo Rojas
Universidad de Ingenieria y Tecnologia

Content:

1. **Introduction**.
2. Projects:
   2.1 Learning from ionosondes: Predicting ionograms.
   2.2 Spread-F forecasting: Predicting occurrence.
3. Summary and conclusions.

1. Introduction





. Developments in technology => Big Data => Train very powerful statistical models.

. Neural Networks try to capture F: X ➔ Y , but we have G(Y).

. We tested regression and classification models.

Content:

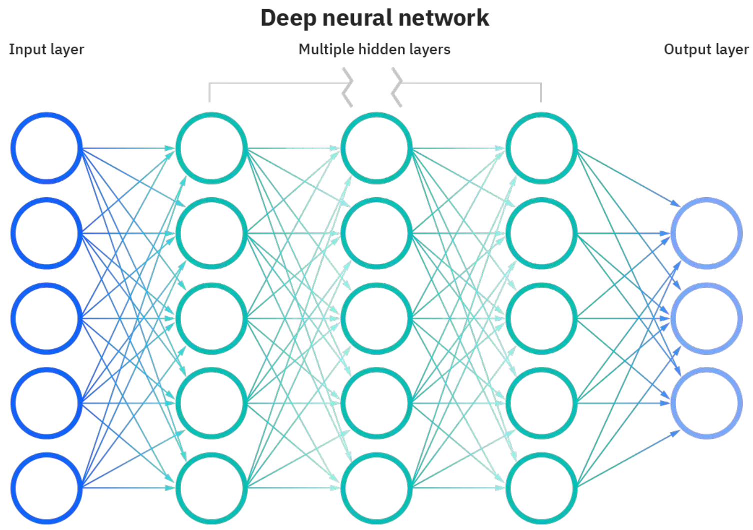
1. Introduction.
2. Projects:

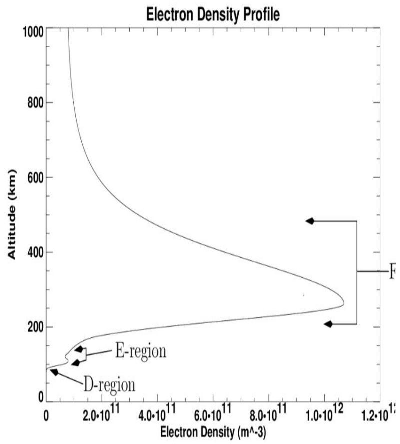
   2.1 Learning from ionosondes: Predicting ionograms.

   2.2 Spread-F forecasting: Predicting occurrence.
3. Summary and conclusions.

2.1 Learning from ionosondes: Predicting ionograms.

Geophysical Parameters





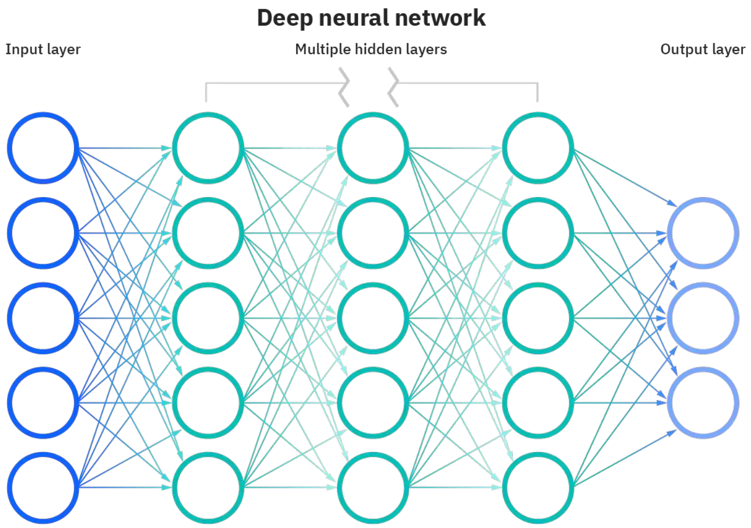
Jhassmin Aricoche, *JRO*

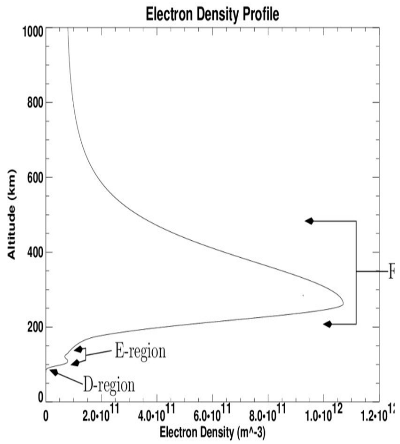
. Long-term goal: teach a NN to estimate electron density profiles

2.1 Learning from ionosondes: Predicting ionograms.

Geophysical Parameters

X

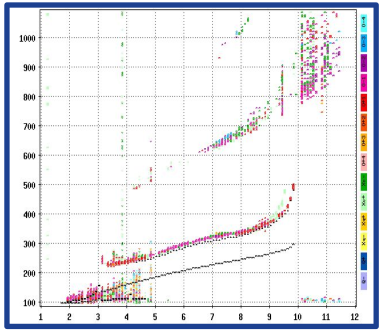




Y

But only have...

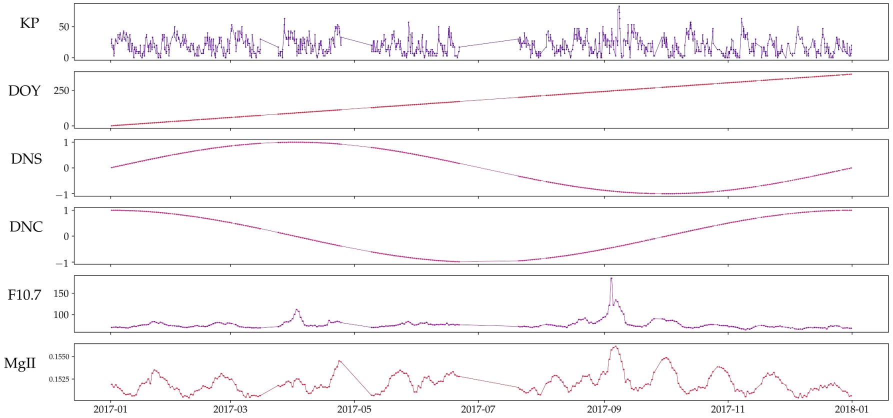
G(Y)



Will start with this!

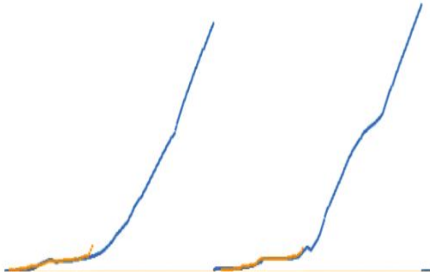
. Looking for F: X → Y , but we have G(Y).

2.1 Learning from ionosondes: Predicting ionograms.



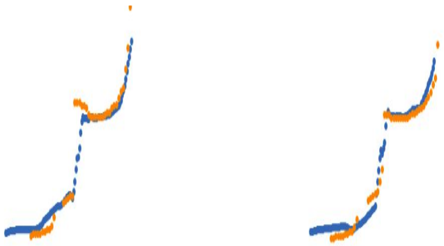
NN that predicts virtual heights.

NN_1



NN that predicts critical frequency.

NN_2

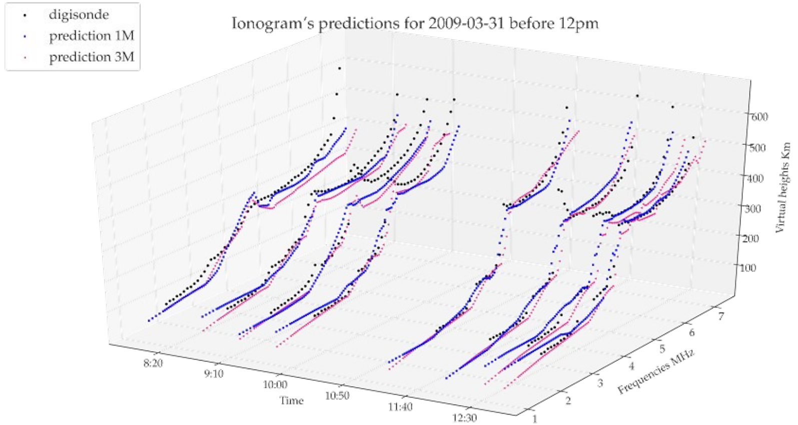


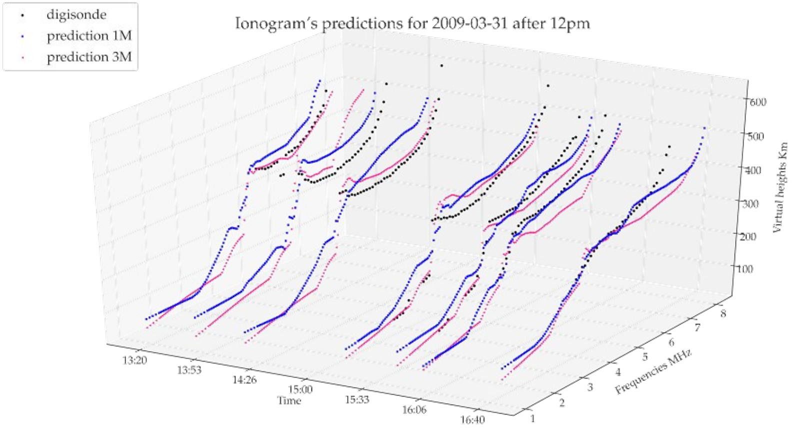
. Geophysical parameters and time.

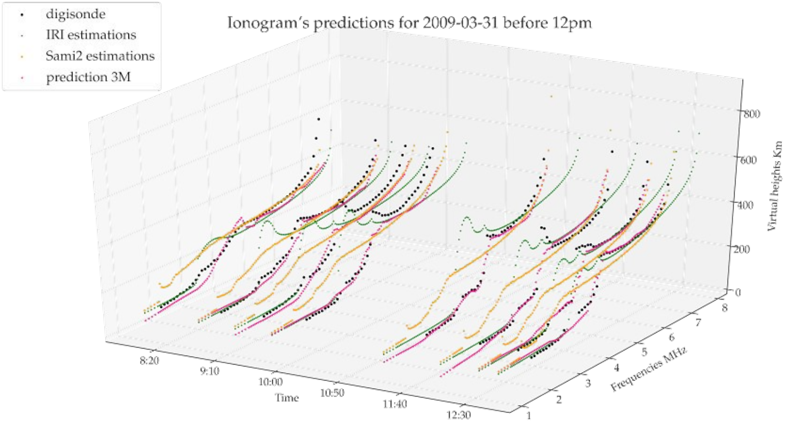
. Training strategy:

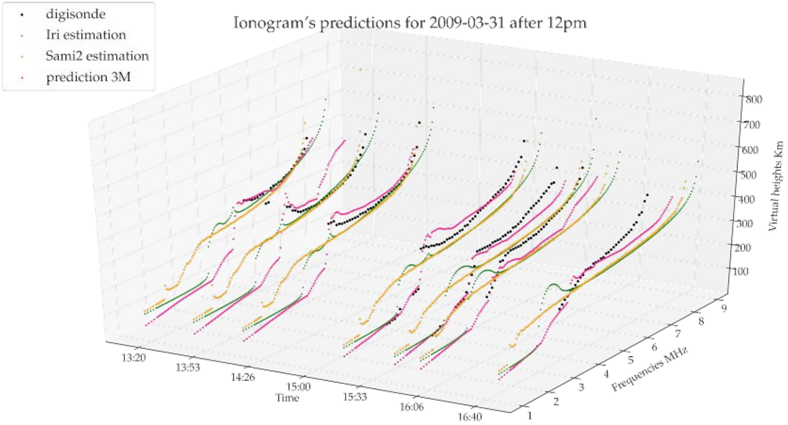
Training (90%)	Validation (10%)	Testing

2.1 Learning from ionosondes: Predicting ionograms.

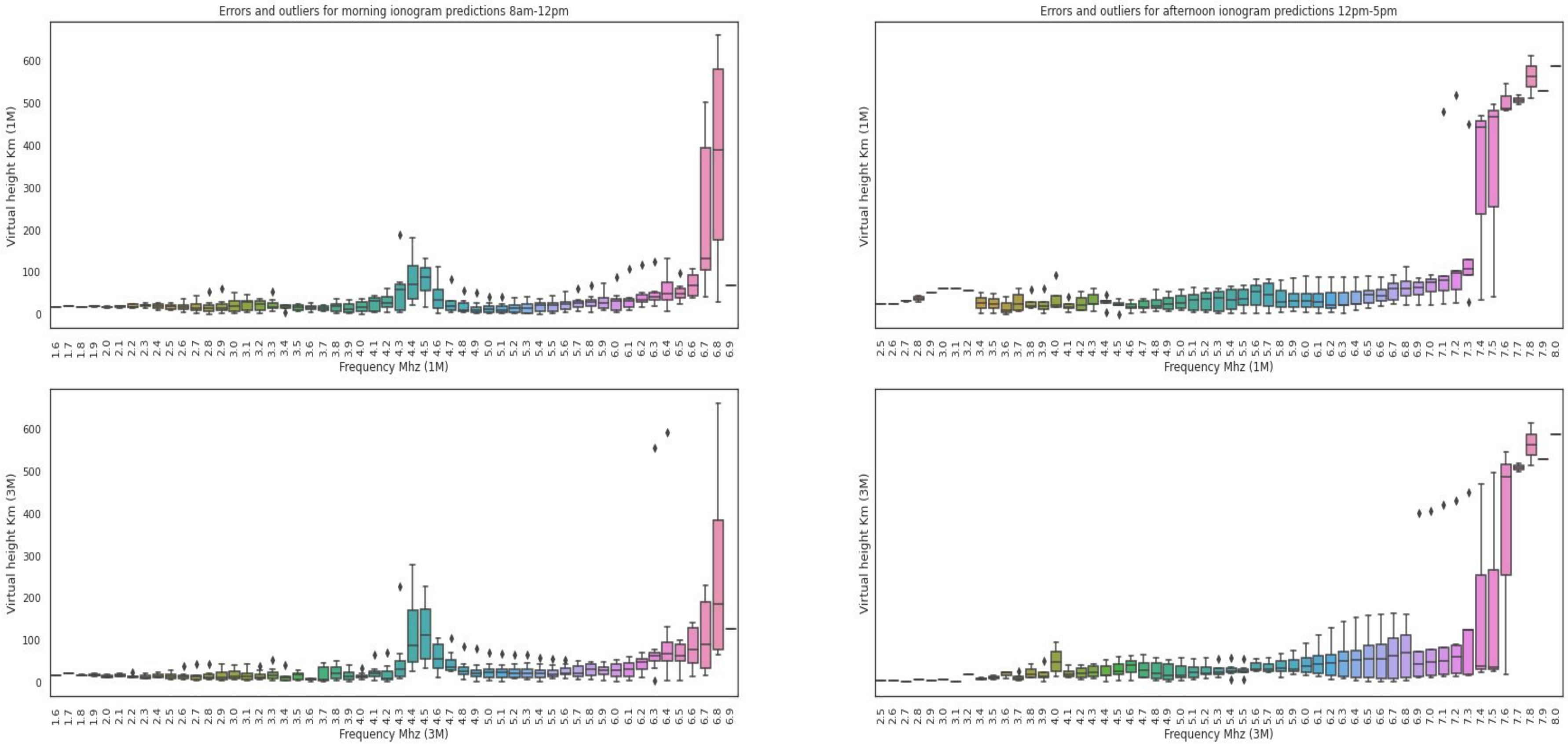








2.1 Learning from ionosondes: Predicting ionograms.



. Not a significant increase in accuracy by adding two months of data but virtual heights around critical frequency seem to improve a little bi.

2.1 Learning from ionosondes: Predicting ionograms.

Evaluation of neural network models to estimate ionograms (RMSE Km)				
Metrics	Model 2 (1 month of data)	Model 2 (3 months of data)	IRI ESTIMATIONS	SAMI2 ESTIMATIONS
Solstice of a Solar Minimum (December 2009)	43.47	51.69	87.23	81.15
Equinox of a solar minimum (March 2009)	25.64	30.37	82.86	70.07
Solstice of a solar maximum (June 2014)	53.04	40.20	54.45	91.68
Equinox of a solar maximum (March 2013)	32.46	31.15	67.0	49.23

Open questions:

. How to implement the loss function to connect ionograms and electron densities?

. What is the optimal data-driven architecture?

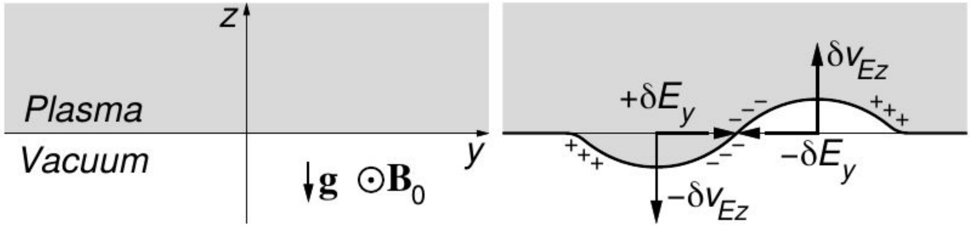
Content:

1. Introduction.
2. Projects:

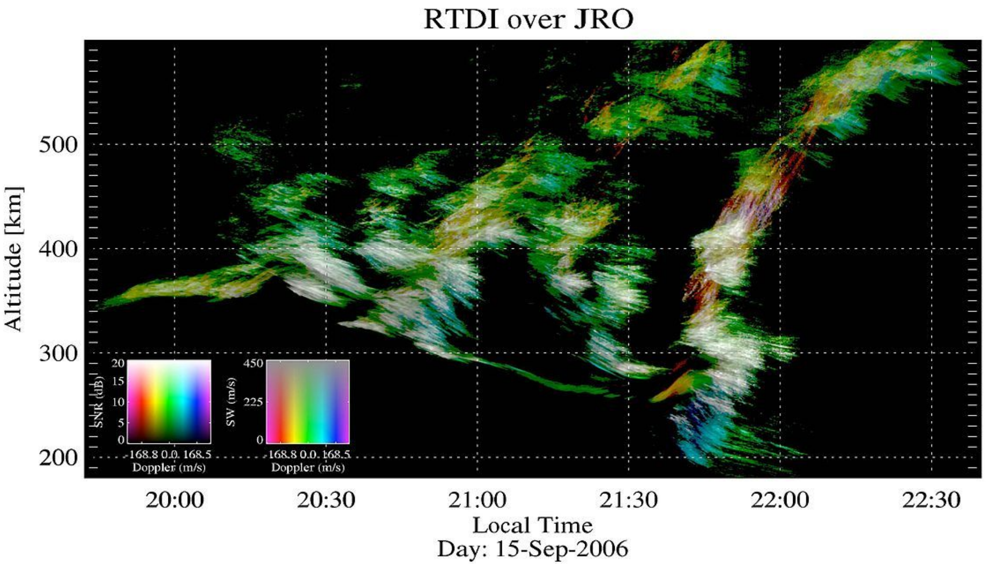
   2.1 Learning from ionosondes: Predicting ionograms.

   2.2 Spread-F forecasting: Predicting occurrence.
3. Summary and conclusions.

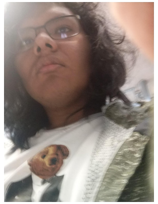
2.2 Spread-F forecasting: Predicting occurrence.



. Main mechanism for the Generalized Rayleigh Taylor (GRT) instability.



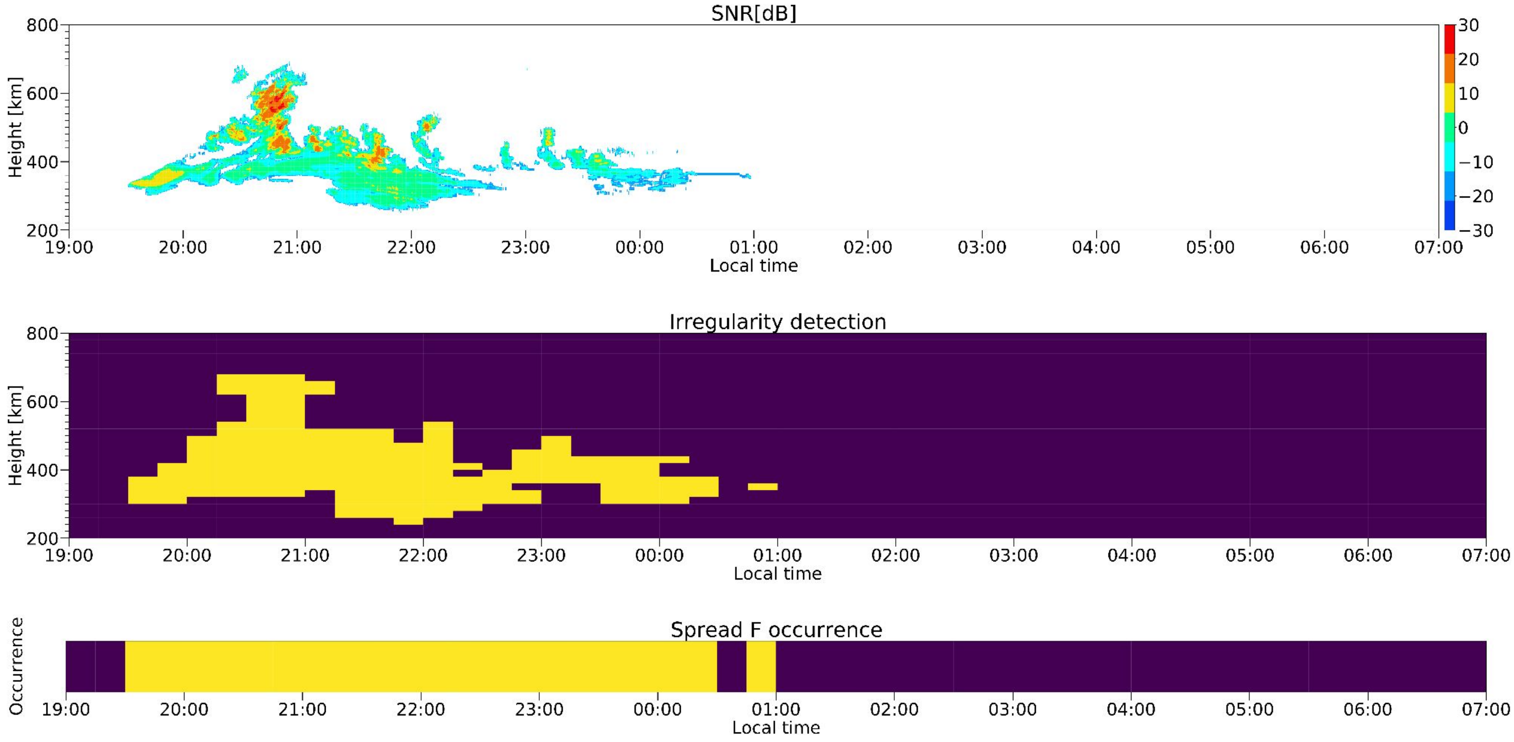
. Neural Networks try to capture F: X → Y



Reynaldo Rojas, *UTEC*

. This is how GRT manifests in Jicamarca's JULIA radar.

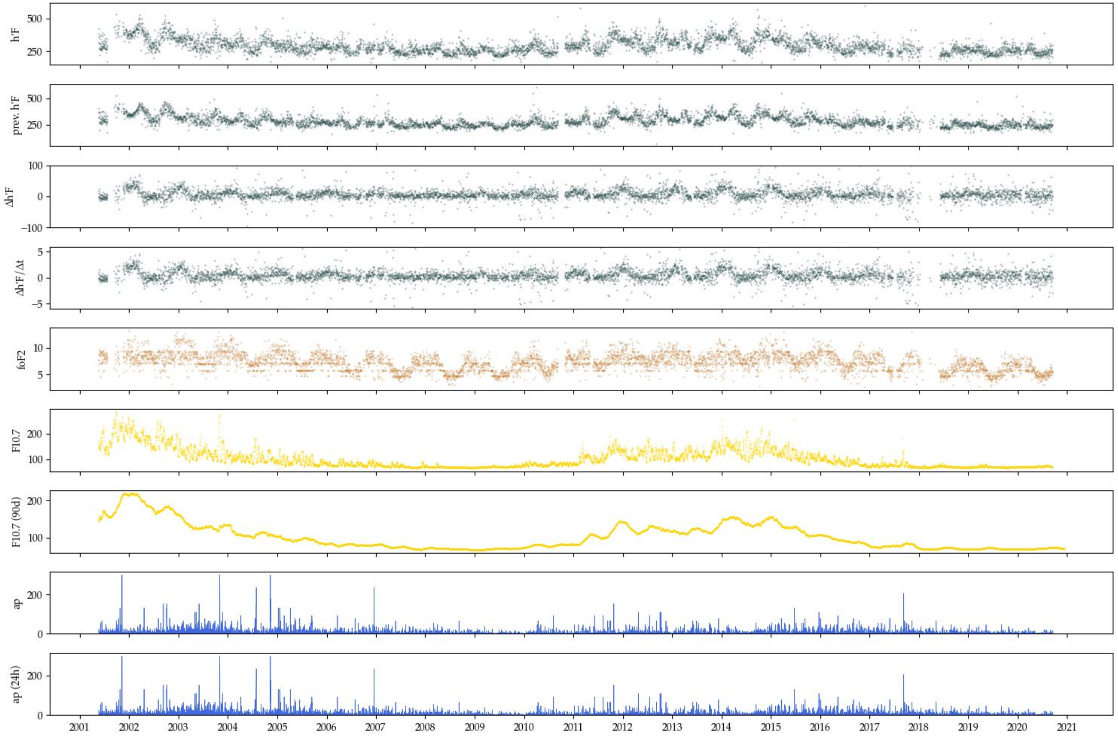
2.2 Spread-F forecasting: Predicting occurrence.



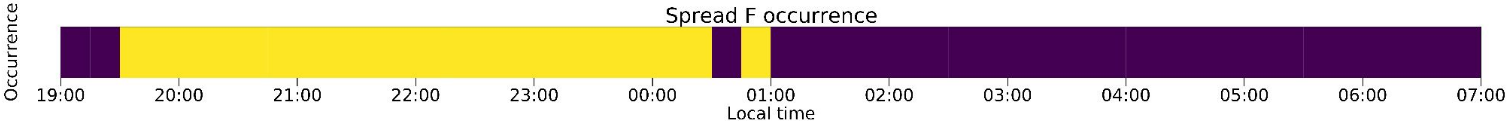
. Followed the approach by Zhan(2018) to bin JULIA RTIs and define Spread-F occurrence.

2.2 Spread-F forecasting: Predicting occurrence.

. h’F was included to try to capture the upwelling effect. It was obtained from the local Digisonde.



X



Y

2.2 Spread-F forecasting: Predicting occurrence.

. Used the Forecasting Ionospheric Real-Time Scintillation Tool (FIRST) model as **reference**.

Forecasting scintillation activity and equatorial spread *F*

David N. Anderson[1] and Robert J. Redmon[2]

[1]CIRES and NOAA/SWPC, University of Colorado Boulder, Boulder, Colorado, USA, [2]NOAA/NCEI, Boulder, Colorado, USA

. Assumption 1: ExB threshold for high scintillation.

. Assumption 2: ExB well correlated to h’F

=> **Threshold in h’F (h’F_{th}) can be used to predict scintillation.**

. How does FIRST work?

1. Measure local h’F at **7:30 pm**.
2. Estimate h’F_{th} using f10.7.
3. If h’F>h’F_{th} then Spread F will occur at **8:30 pm**.

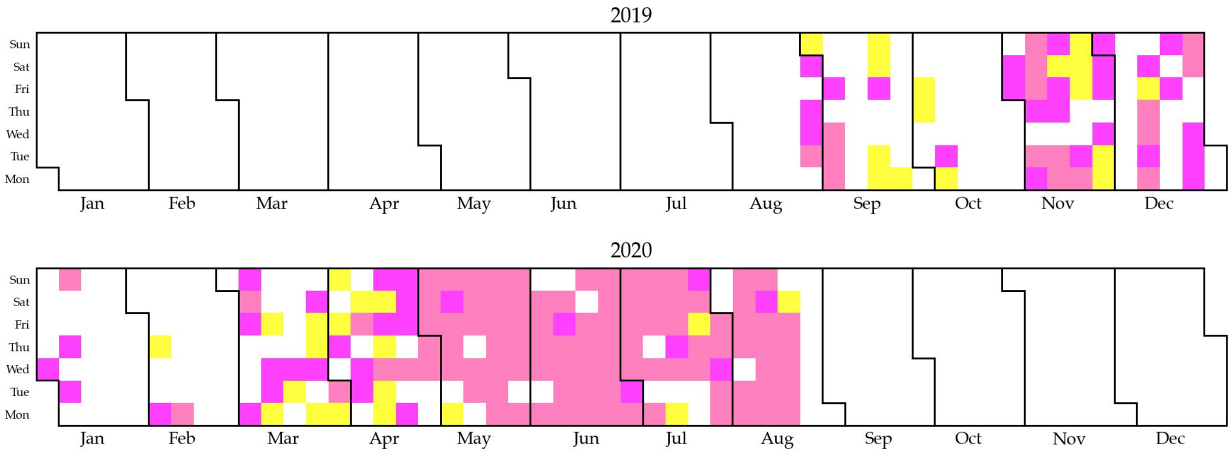
. We will follow the **same procedure** for comparison.

2.2 Spread-F forecasting: Predicting occurrence.

FIRST

Accuracy: 73%

43	0
58	120



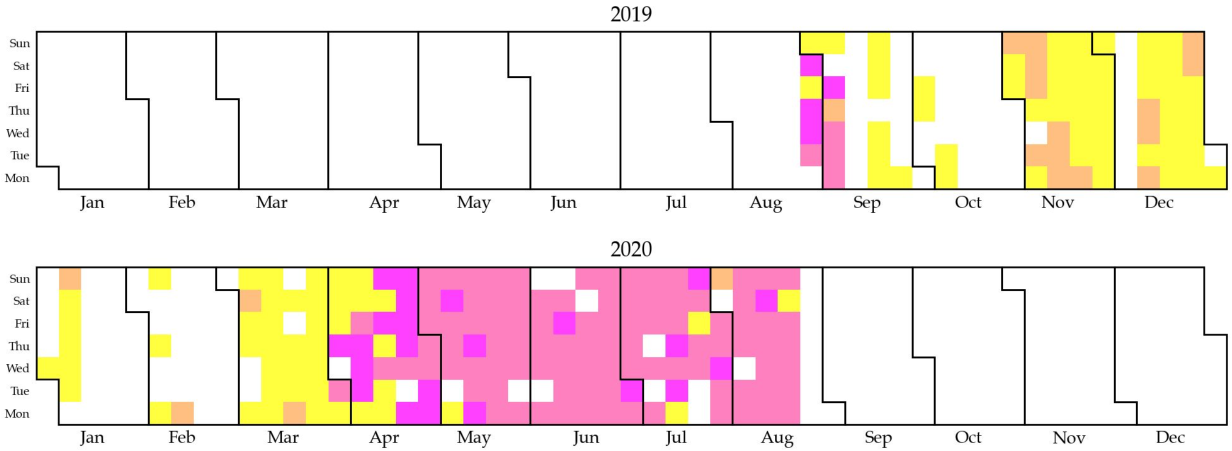
True Positive

False Positive

True Negative

False Negative

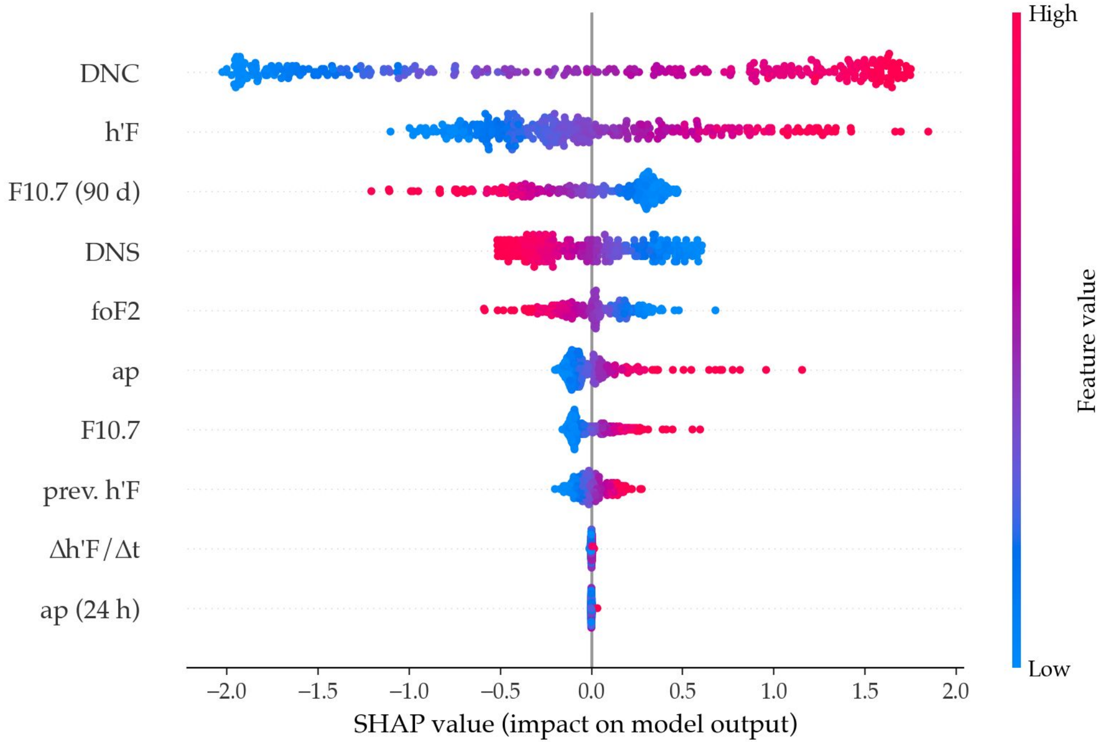
113	23
27	103



Our model

Accuracy: **81%**

2.2 Spread-F forecasting: Predicting occurrence.



. Analyzing NN's sensitivity to inputs using SHAP values.

. Seems consistent with climatology.

Open questions:

. How to understand the discrepancies between scintillation-based models and JULIA data?

. Can inputs be improved to be better predictors of occurrence?

Content:

1. Introduction.
2. Projects:
    2.1 Learning from ionosondes: Predicting ionograms.
    2.2 Spread-F forecasting: Predicting occurrence.
3. **Summary and conclusions.**

3. Summary and conclusions.

. We now have the hardware and the mathematics to revisit well known ionospheric estimation problems.

. Regression NN seem to capture most of the variability of ionograms more accurately than empirical and numerical models.

. Classification NN show promising results in predicting Spread-F occurrence, so far outperforming other linear regression based methods.

. **Future work**:

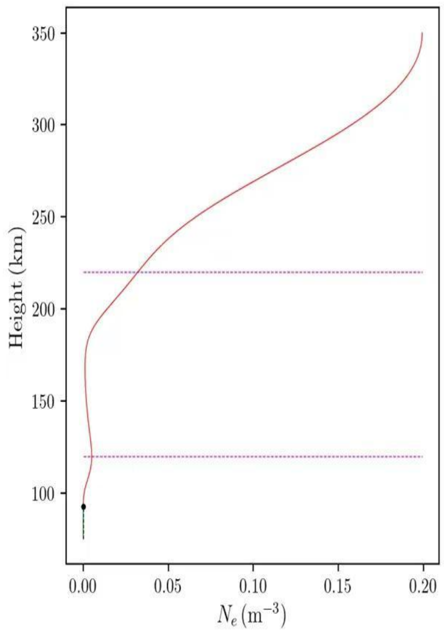
- Learn electron density profiles.
- Physics-based proxies for inputs and try using other datasets (HF network?)
- Forecasting other Spread-F features (onset altitude, range coverage, etc).
- Predict Spread-F RTIs.

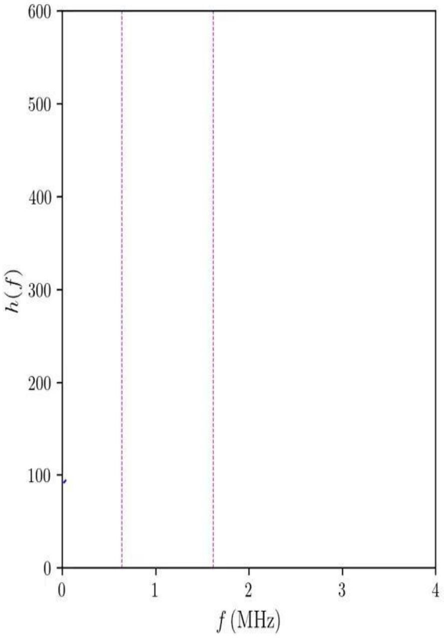
Thank you.

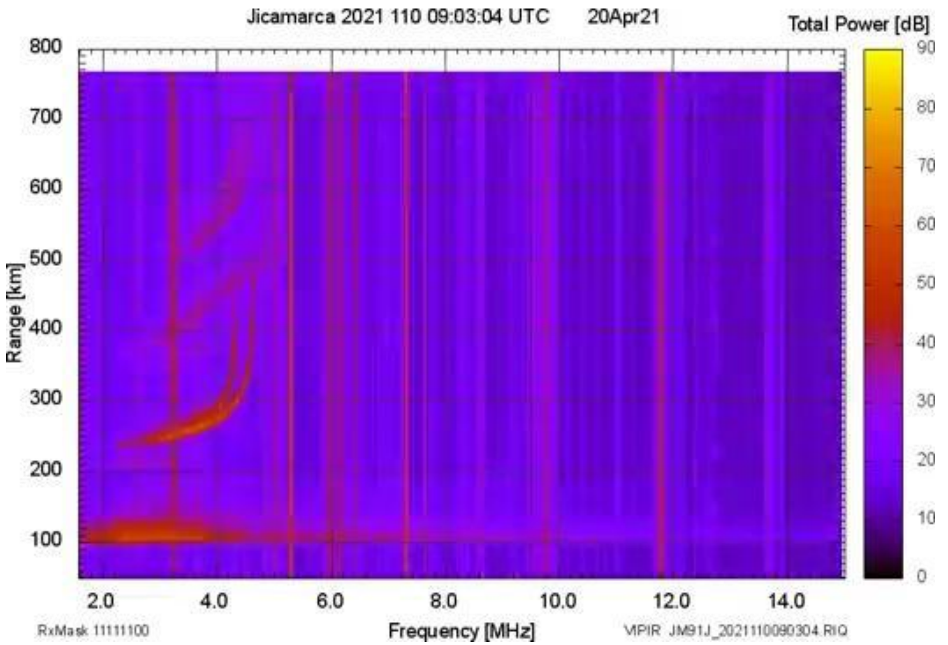
Enrique Rojas
elr96@cornell.edu

EXTRAS

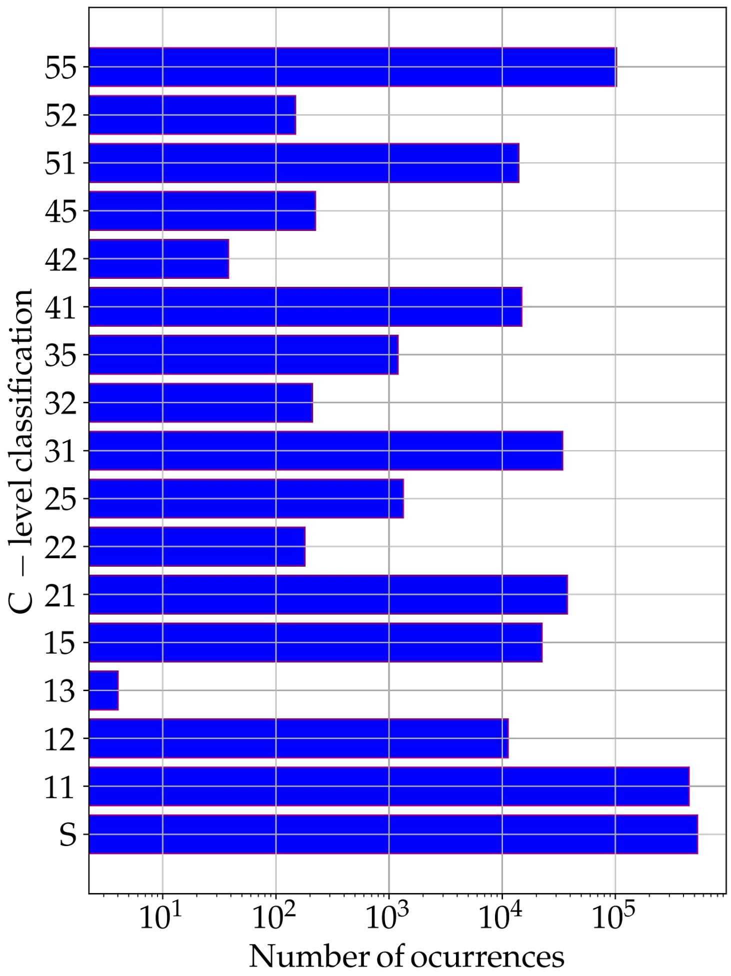
2.1 Learning from ionosondes: Predicting ionograms.

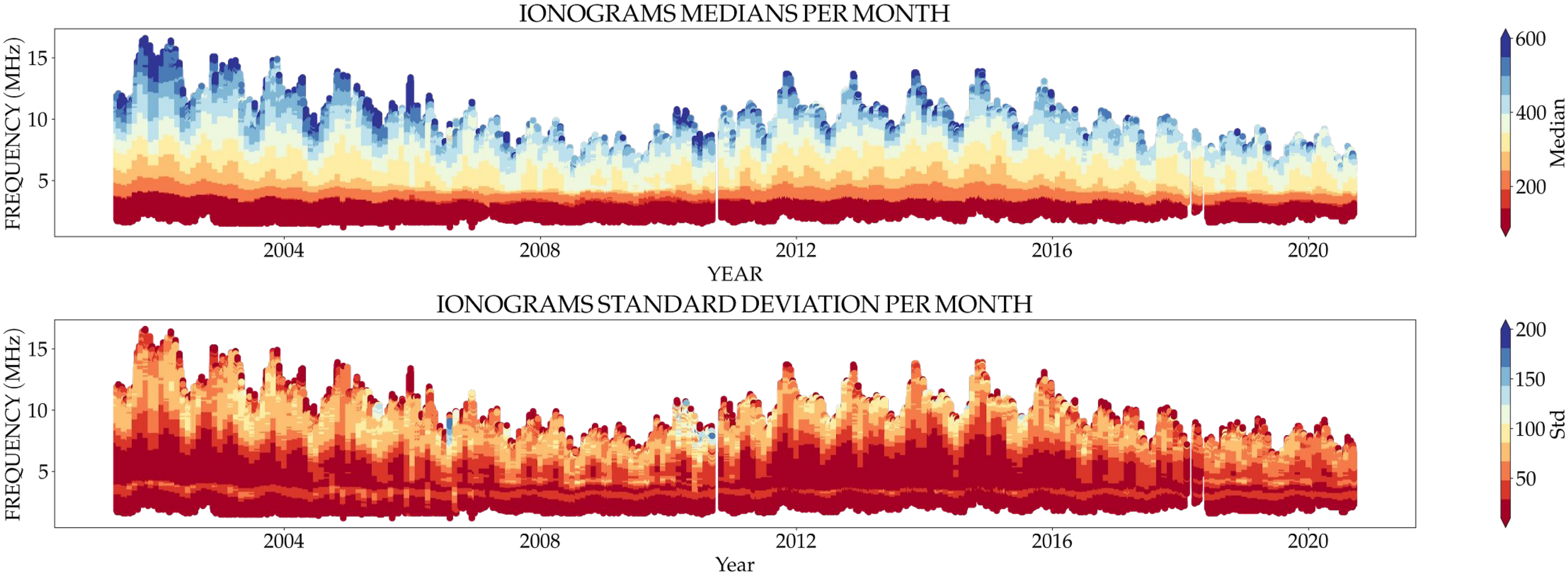


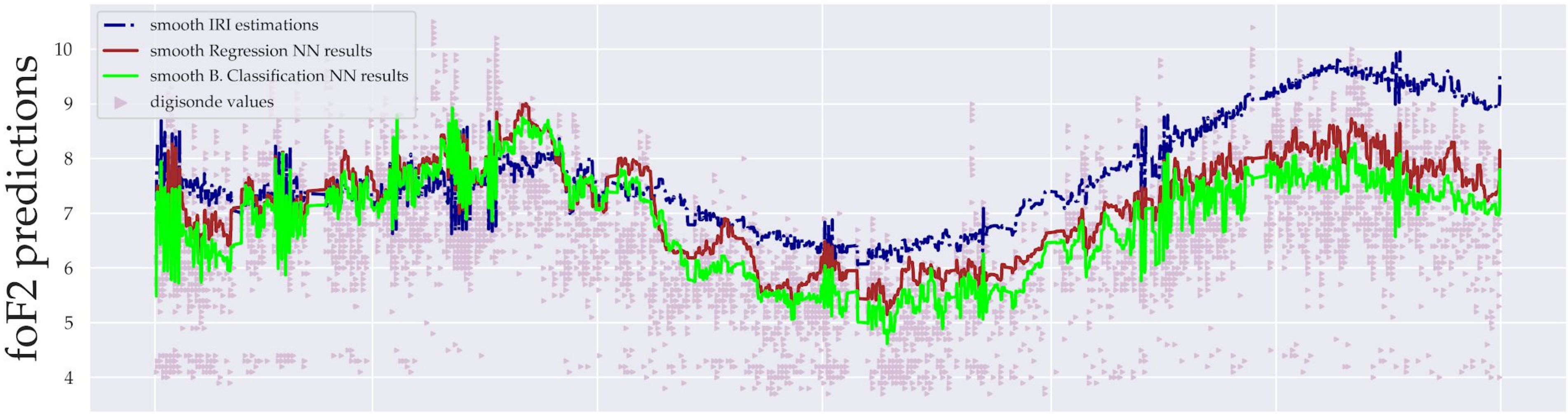




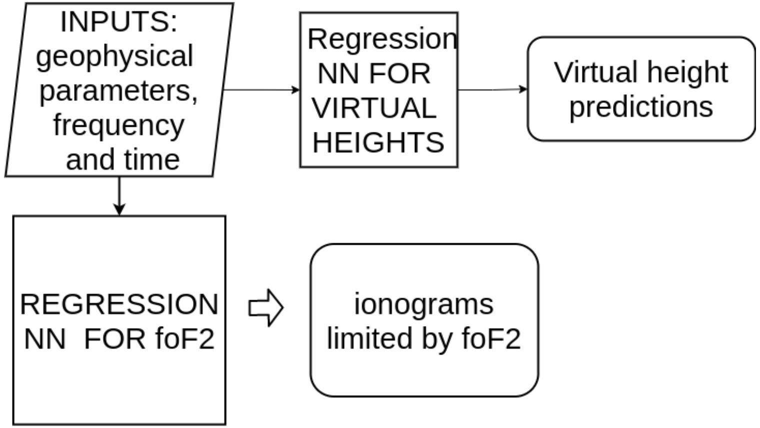
Neural Networks are F: X ---> Y , but we have G(Y).


55
52
51
45
42
41
35
32
31
25
22
21
15
13
12
11
S
C − level classification
10^1
10^2
10^3
10^4
10^5
Number of ocurrences

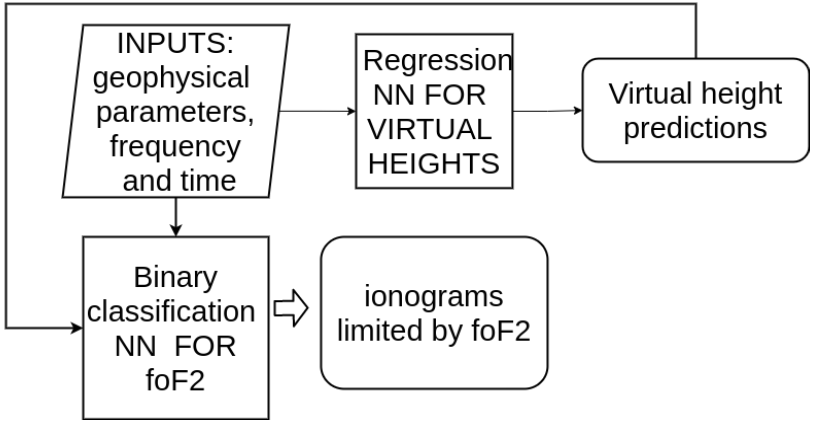

IONOGRAMS MEDIANS PER MONTH
FREQUENCY (MHz)
15
10
5
2004
2008
2012
2016
2020
YEAR
600
400
200
Median
IONOGRAMS STANDARD DEVIATION PER MONTH
FREQUENCY (MHz)
15
10
5
2004
2008
2012
2016
2020
Year
200
150
100
50
Std

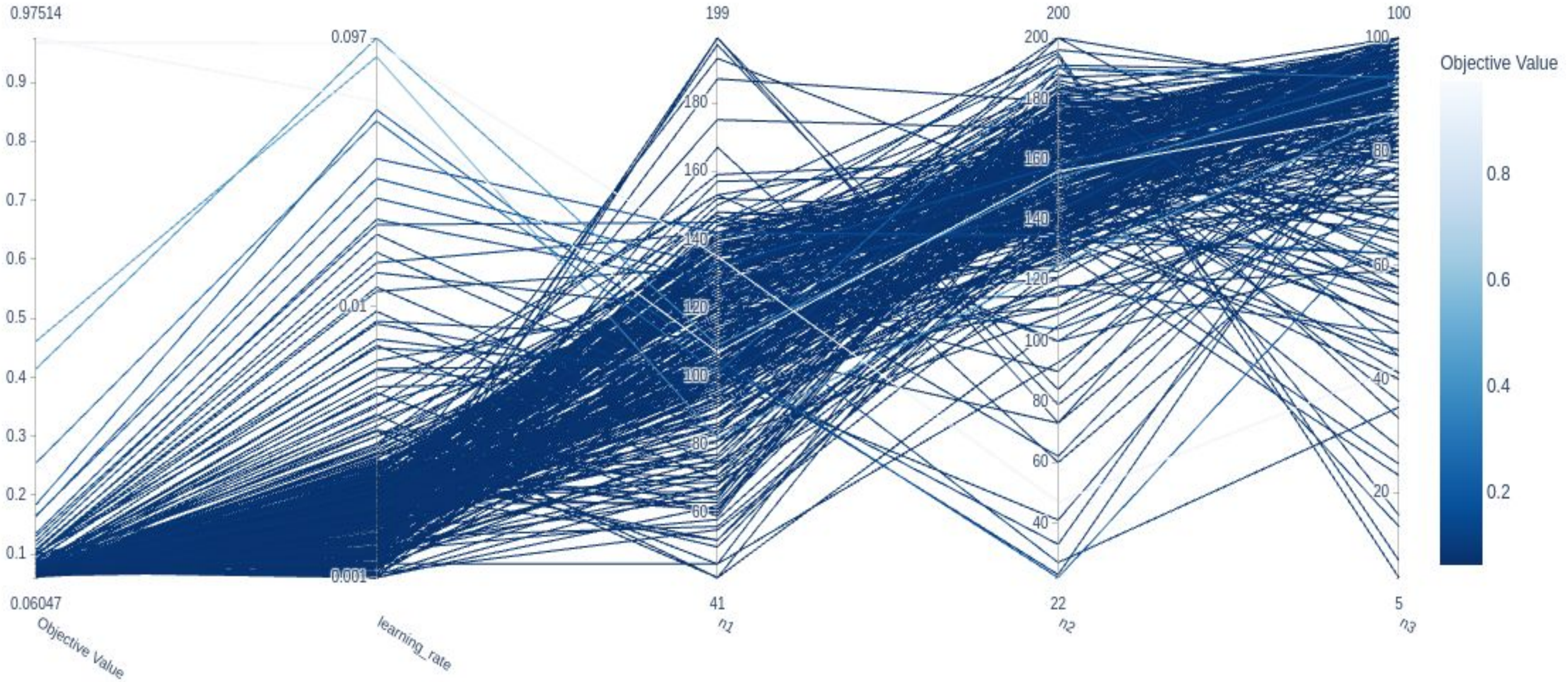

foF2 predictions
10
9
8
7
6
5
4
smooth IRI estimations
smooth Regression NN results
smooth B. Classification NN results
digisonde values

Regression NN



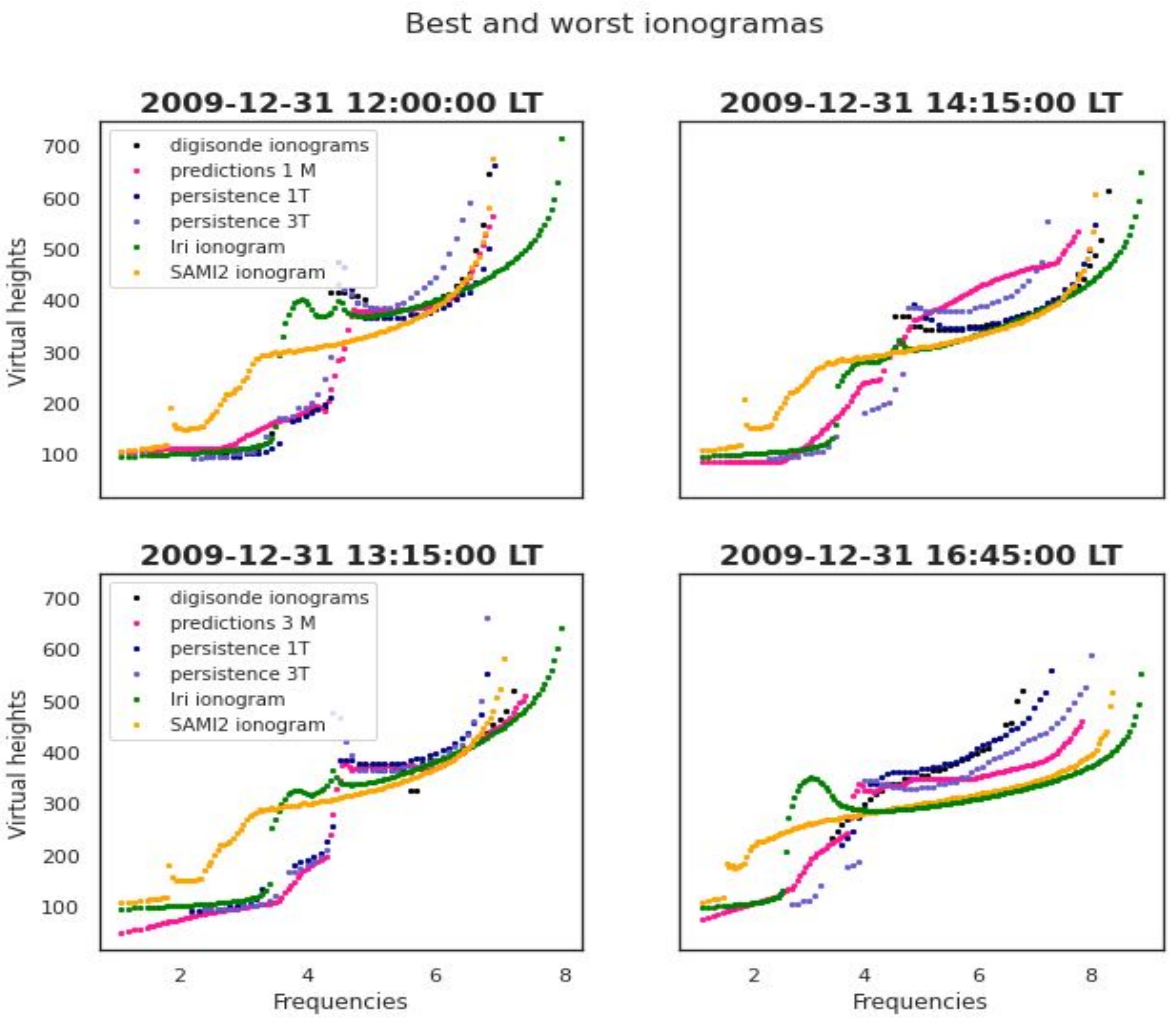
Classification NN





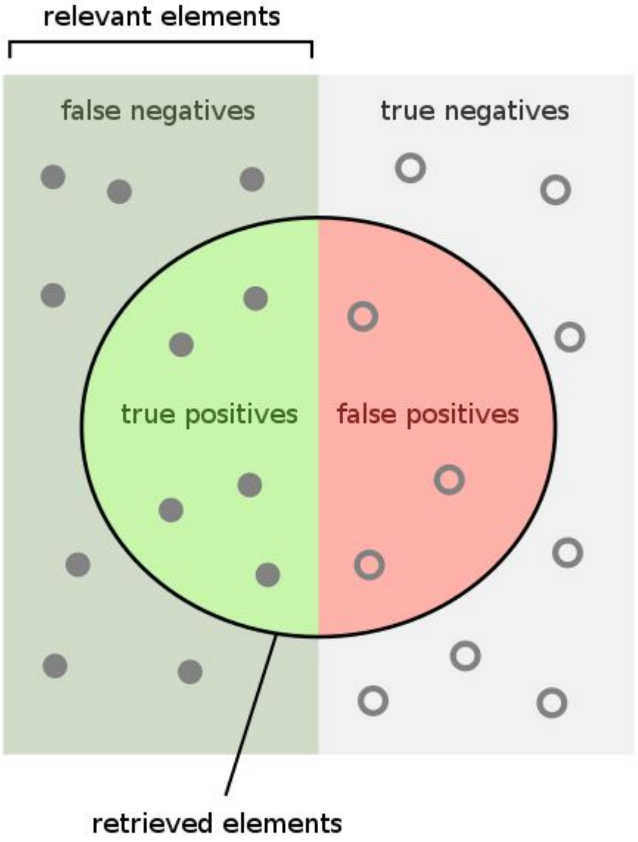
{'learning_rate': 0.0020807556518588114, 'n1': 116, 'n2': 162, 'n3': 91}

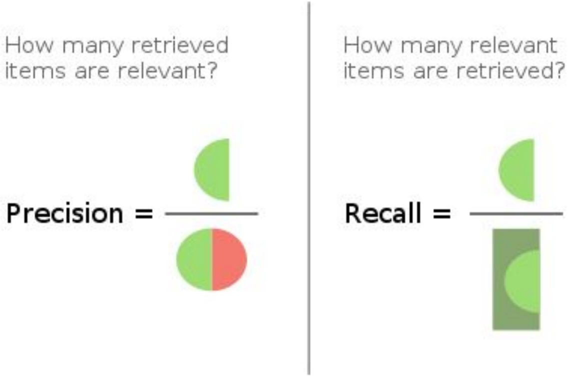
. Some examples of both how accurate and bad the estimations can be.



Evaluation of neural network models to estimate foF2 (RMSE-MHz)				
Metrics	Model 2 (1 month of data)	Model 2 (3 months of data)	IRI ESTIMATIONS	SAMI2 ESTIMATIONS
Solstice of a Solar Minimum (December 2009)	0.44	0.47	1.12	0.59
Equinox of a solar minimum (March 2009)	0.58	0.51	1	0.75
Solstice of a solar maximum (June 2014)	0.62	0.82	0.67	1.47
Equinox of a solar maximum (March 2013)	1.81	1.53	1.25	0.70

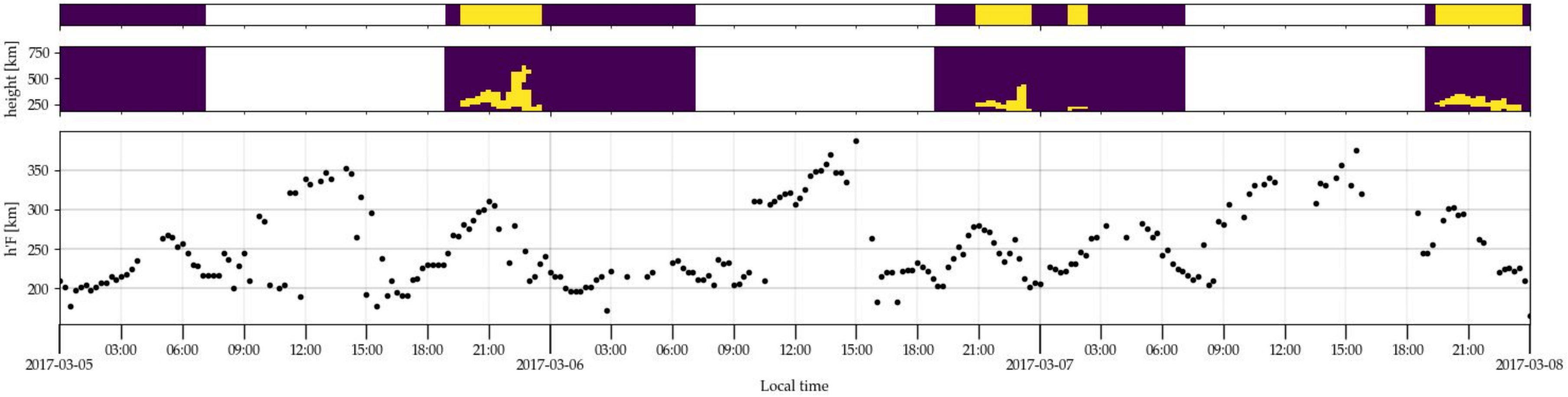
Evaluation of neural network models to estimate foF2					
Season	Metrics	Binary classification network (1 month of data)	Binary classification network (3 months of data)	Regression network of only foF2 (1 month of data)	Regression network of only foF2 (3 months of data)
Solstice of a Solar Minimum (December 2009)	F1	0.97	0.97	does not apply	does not apply
Solstice of a Solar Minimum (December 2009)	RMSE	0.44	0.47	0.48	0.78
Equinox of a solar minimum (March 2009)	F1	0.9566	0.9665	does not apply	does not apply
Equinox of a solar minimum (March 2009)	RMSE	0.58	0.51	0.75	0.48
Solstice of a solar maximum (June 2014)	F1	0.967	0.95	does not apply	does not apply
Solstice of a solar maximum (June 2014)	RMSE	0.62	0.82	0.66	0.38
Equinox of a solar maximum (March 2013)	F1	0.89	0.9	does not apply	does not apply
Equinox of a solar maximum (March 2013)	RMSE	1.81	1.53	0.78	0.96





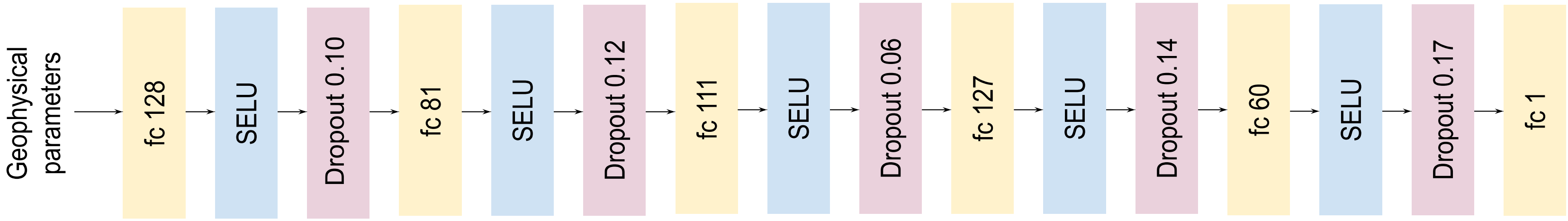
$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}} = \left(\frac{\sum\limits_{i=1}^{n} x_i^{-1}}{n} \right)^{-1}.$$

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

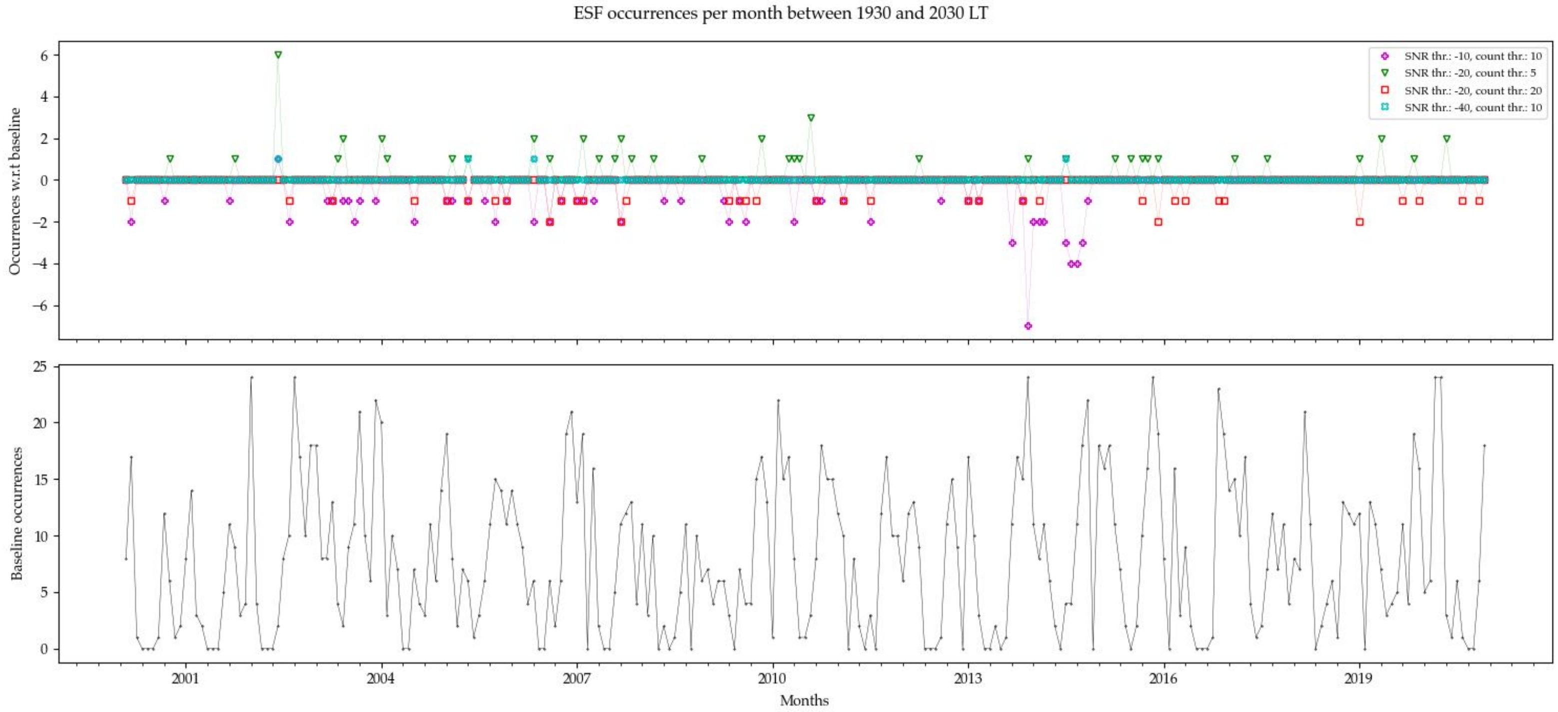

height [km]
750
500
250
h'F [km]
350
300
250
200
03:00
06:00
09:00
12:00
15:00
18:00
21:00
2017-03-05
2017-03-06
2017-03-07
2017-03-08
Local time

2.2 Spread-F forecasting: Predicting occurrence.

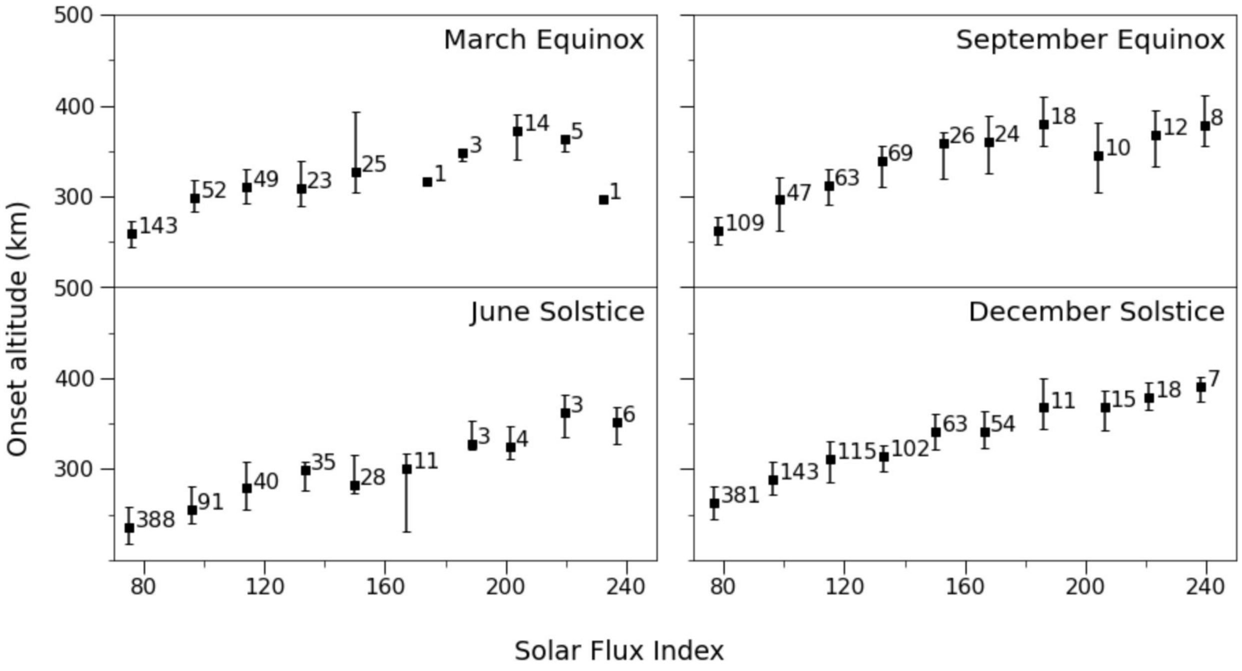
. Model architecture proposed after hyper-parameter optimization:



. Training strategy:



identified based on number of echoes above a certain SNR within a time versus height interval. The observations for each day were grouped in a height versus time matrix. The matrix has resolutions of 15 min and 20 km in height. Irregularity detection is assigned to a matrix bin when more than 10 echoes with SNR greater than −20 dB were detected. The SNR threshold (−20 dB) and number of echoes (at least 10) were chosen as to minimize the misidentification of interference and clutter as echoes caused by ESF irregularities. Despite our

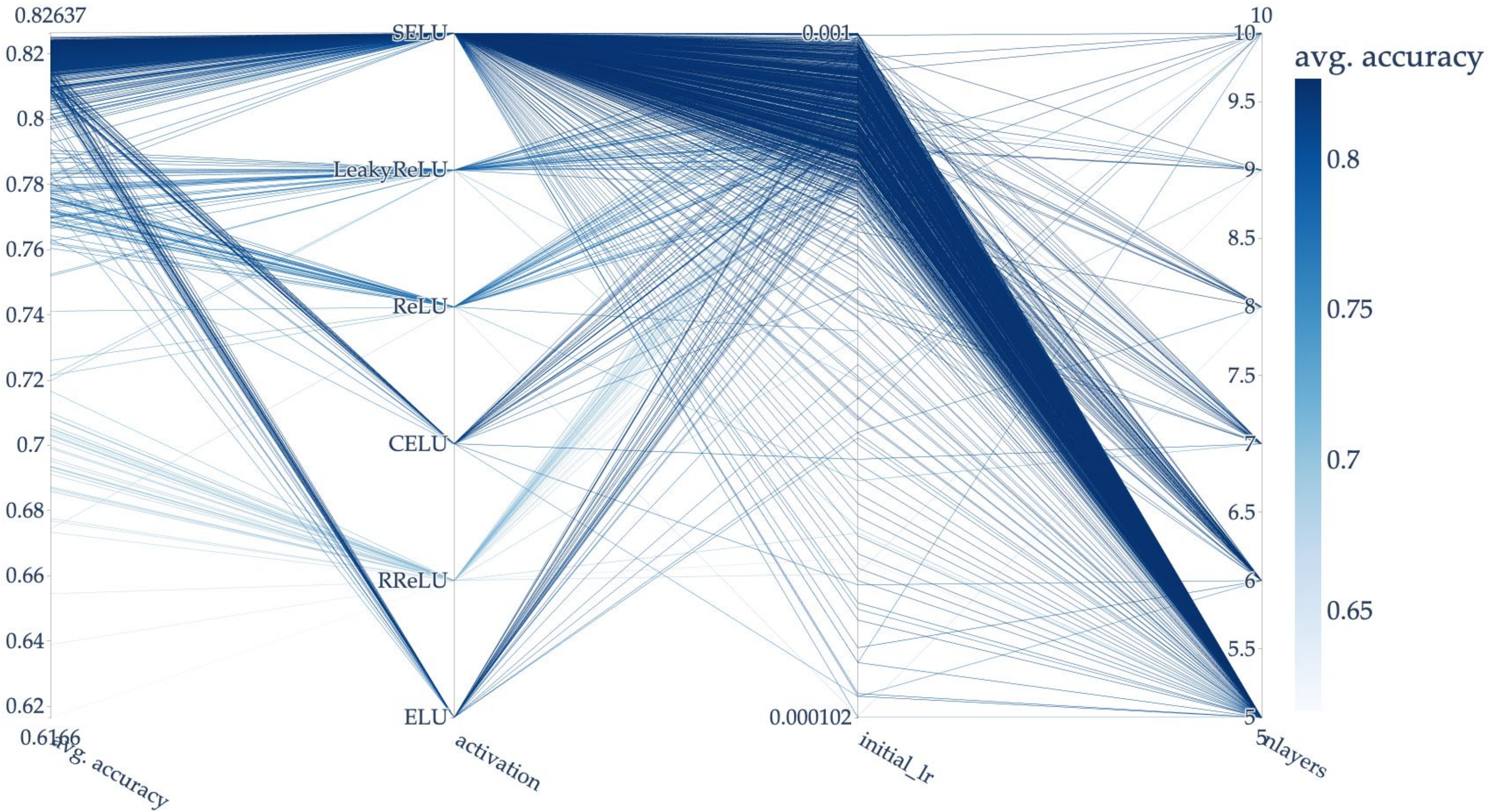


As we can see, on June Solstice, the altitudes corresponding to the lowest solar flux index values are lower than on the other seasons; on September Equinox and December Solstice the altitudes corresponding to the highest solar flux index values are higher than on the other seasons. The altitudes also increase with solar flux for every season as expected (Chapagain et al, 2009; Redmon et al 2013, 2017).

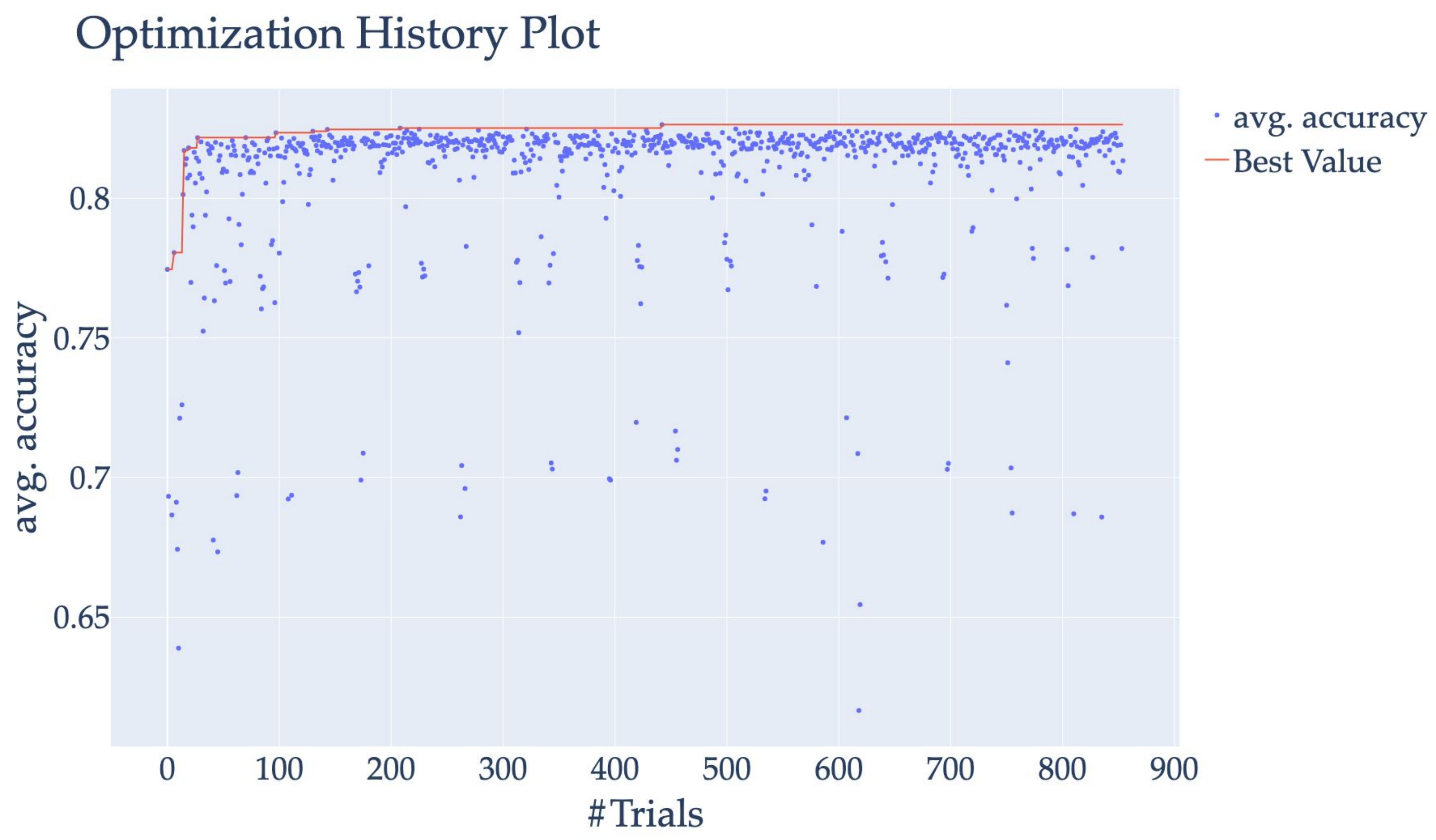
h’F (1930 LT): We use the time 1930 LT for two reasons: The onset time of spread F is usually around 1920 LT and 1945 LT for equinox and December solstice (Chapagain et al, 2009) and also because we compare our model with the FIRST.
h’F (prev. 30 min): This is the first value of h’F for which we have available data between 1900 LT and 1930 LT. This might indicate how fast the F layer has risen in the past 30 minutes.
F10.7: Correlates with onset altitude.
F10.7 (90 days): This is an average value of solar flux index in the last 90 days and it provides some information about the solar cycle.
Ap, Kp, Ap (24 h): Geomagnetic activity, depending on the local time, season and solar cycle, affects the occurrence of irregularities (Hysell and Burcham, 2002).
Day of the year: This is relevant because of the season-to-season variability shown in Figure 2.

$$DNS = \sin(2\pi D/365),\ DNC = \cos(2\pi D/365),\ D\text{: Day of year (1-365)}$$

The network was trained with the **Adam** optimization algorithm for **20 epochs**. The loss function used was **binary cross-entropy** loss with **numerical stability**. The **batch size** chosen was **16**. We did **not** conduct **hyperparameter optimization**.



0.82637
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0.68
0.66
0.64
0.62
0.6166
avg. accuracy
SELU
LeakyReLU
ReLU
CELU
RReLU
ELU
activation
0.001
0.000102
initial_lr
10
9.5
9
8.5
8
7.5
7
6.5
6
5.5
5
nlayers
avg. accuracy
0.8
0.75
0.7
0.65


Optimization History Plot
avg. accuracy
Best Value
0.8
0.75
0.7
0.65
avg. accuracy
0
100
200
300
400
500
600
700
800
900
#Trials