Forecasting Spread F at Jicamarca

Reynaldo Rojas Zelaya[1], Enrique Rojas Villalba[2], Jhassmin Aricoché Del Campo[3], Marco Milla Bravo[4]
1. Universidad de Ingeniería y Tecnología, Lima, Perú
2. Cornell University, New York, USA
3. Jicamarca Radio Observatory, Lima, Perú
4. Sección Electricidad y Electrónica, Pontificia Universidad Católica del Perú, Lima, Perú
reynaldo.rojas@utec.edu.pe

Introduction

Highly non-linear processes take place in the Ionosphere. Equatorial spread F (ESF) is one of these phenomena, and it is characterized by plasma depletions. Fortunately, neural networks can learn from high dimensional, non-linear data.
In this work, we **propose a day to day prediction model**. We compared the proposed model's predictions to FIRST (Forecasting Ionospheric Real-time Scintillation Tool) [1, 9].

Measurements

ESF occurrences were **characterized from JULIA** (Jicamarca Unattended Long-term Investigations of the Ionosphere and Atmosphere) radar measurements. Jicamarca's digisonde's measurements, along with other parameters and ESF occurrences, were merged resulting in a dataset that spanned the years 2002 - 2020.

Data processing

We use a **binning-binarization algorithm** to characterize ESF [10]. Since we were only concerned with occurrences along the time axis, we collapsed the heights as shown in figure 1. We processed the SNR (signal to noise ratio) data in order to produce a ground truth dataset (of ESF occurrences) necessary for the supervised learning algorithm. The characterization procedure relies on thresholds so we compared different values in figure 2.

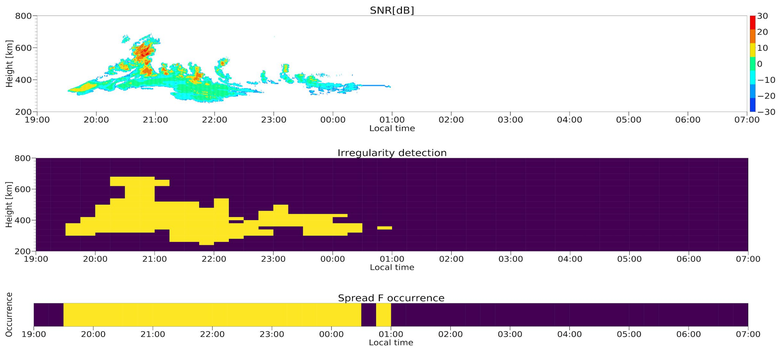

Figure 1. Characterization of ESF occurrence for January 2nd, 2000.

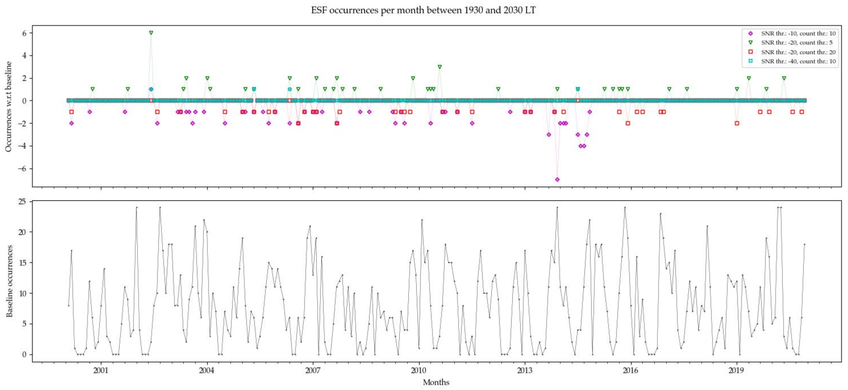

Figure 2. ESF occurrences per month (bottom) and differences obtained by varying the thresholds (top).

Climatology

We extend climatology plots [3] for ESF's onset altitude with our dataset (years 2002-2020), with the difference that we use median and interquartile ranges as statistics.

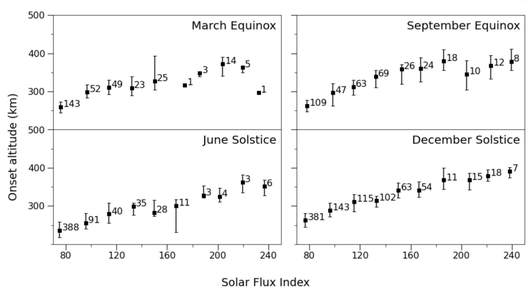

Figure 3. Median onset altitude versus mean solar flux for the dates from January 2002 to August 2020.

As depicted in figure 3, on June Solstice the altitudes corresponding to the lowest solar flux index values are lower than on the other seasons. As expected [1, 3, 9], onset altitudes increase with solar flux in all seasons. This might be due to the inverse dependence of the growth rate (γ) on F10.7 and its direct dependence on height, both through the ion-neutral collision frequency (ν_{in}) [6]. In the next section, we explore other possible input features.

Neural Network Inputs

Rationale

h'F (1930 LT): "The height of the nighttime F layer is the single most important parameter controlling the generation of spread F" [4]. We use the measurements at 1930 LT for two reasons: The onset time of ESF is usually around 1920-1945 LT for equinox and December solstice [3] and because we compare our model with FIRST which also makes predictions at the same LT.
h'F (prev. 30 min): A previous h'F value up to 30 mins before 1930 LT.
Δh'F/Δt: The rate of change of h'F might have an effect on ESF.
F10.7, **F10.7 (90 days):** Solar flux and its average. This should provide some information about the solar cycle as well as the current conditions.
ap, ap (24 h): Geomagnetic activity and its average. Depending on the local time, season, and solar cycle, it has an effect on the occurrence of irregularities [5].
Day of the year: Figure 3 depicted the part of the seasonality of ESF.

Feature exploration

We compared the accuracy (ratio of correct predictions to total number of predictions) obtained for the validation dataset with different subsets of geophysical parameters when passed as inputs to the neural network shown in figure 4.

DNS, DNC 79%	261	62	h'F, F10.7 81%	268	56	h'F, h'F(prev. 30 min), F10.7 82%	282	66
	42	140		35	146		21	136
h'F, F10.7, Ap 82%	276	63	h'F, h'F (prev. 30 min), F10.7, F10.7 (90 d), AP, AP (24 h) 83%	275	55	h'F, F10.7, DNS, DNC 83%	274	56
	27	139		28	147		29	146
h'F, h'F (prev. 30 min), F10.7, F10.7 (90 days), Ap, Ap (24 h), DNS, DNC 84%	280	54	h'F, h'F (prev. 30 min), F10.7, F10.7 (90 days) 83%	280	60	h'F, h'F (prev. 30 min), F10.7, F10.7 (90 days), DNS, DNC 84%	278	55
	23	148		23	142		25	147
h'F, h'F_prev, Δh'F/Δt, F10.7, F10.7 (90 d), AP, AP (24 h), DNS, DNC 85%	277	48	h'F, h'F_prev, Δh'F/Δt, F10.7, F10.7 (90 d), AP, AP (24 h), foF2, DNS, DNC 84%	275	48	h'F, foF2, F10.7 83%	270	51
	26	154		28	154		33	151

Table 2. Accuracy comparison for different feature sets. For the colormap reference, see Figure 9.

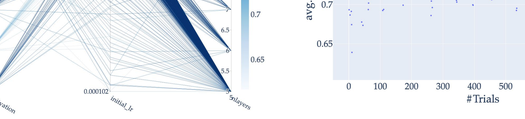

Figure 4. Neural network used during feature exploration.

Input pre-processing

All variables except the day of the year, are scaled between 0 and 1, using Scikit-learn's `preprocessing.MinMaxScaler`. A time series of some features before pre-processing is shown in figure 5. The day of the year (D) is a periodic variable and it was processed as such [2] .

DNS = sin(2πD/365), DNC = cos(2πD/365), D: Day of year (1-365)

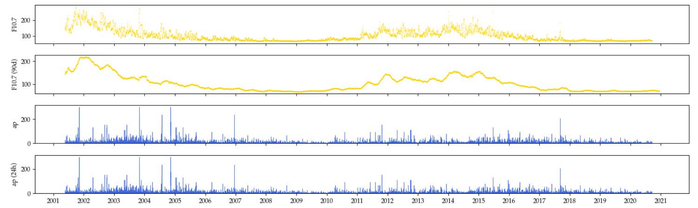

Figure 5. Time series of some geophysical parameters and their derived quantities.

Neural Network

Hyper-parameter optimization

Optimization was conducted using Optuna's TPE implementation. The algorithm's goal was to maximize the average accuracy across a number of folds. These folds were chosen using the sliding window heuristic with training and validation windows of size 2 and 1 years respectively. The hyper-parameters optimized were the learning rate, the number of layers, the number of units for each layer, the probability of dropout for each layer, and the activation function. The optimization history and a sub-configuration of each hyper-parameter configuration tried are shown in figure 6.

Figure 6. Hyper-parameter sub-configurations (left) and optimization history (right).

Architecture

It is a Multilayer Perceptron with SELUs as activation functions. It outputs a real number, which is then passed to a sigmoid function (a function which maps a real number to a number between 0 and 1). Finally, this value is interpreted as a probability for Spread F occurrence.

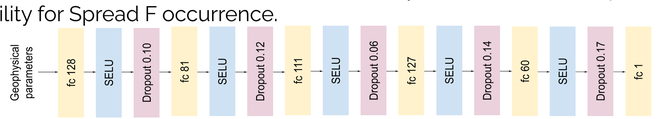

Figure 7. Neural network architecture.

In figure 7, fc d_2 is a fully-connected layer which, when applied to a vector of a given dimension, d_1, outputs a vector of dimension d_2. SELU, is an activation function and it adds non-linearity to the network. Dropout p, is a layer that, with probability p, drops (sets to zero) a given element of its input. This helps the previous layer to evenly distribute the weights of its neurons. We trained the network for 30 epochs with `nn.BCEWithLogitsLoss` [7] and batch size 16.

Input Sensitivity

In this work it is fundamental to understand how the model makes predictions based on different feature values. Figure 8 shows the ranked feature contributions along with their SHAP values. These contributions can be interpreted as the effect that each feature has on the predictions made by the model. We used 300 data points taken from the dataset to make these estimates.

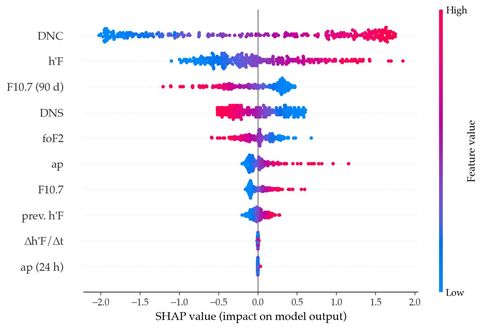

Figure 8. Ranked input features and their corresponding SHAP values.

Results

These results correspond to the evaluation and comparison of our model and FIRST on the testing dataset. It is important to point out that, in this work, FIRST is evaluated with occurrences obtained from the characterization presented earlier as opposed to the evaluation in the original paper [1], which apparently used occurrences from manually labeled ionograms. There were a total of 45 days for which FIRST could not make a prediction.

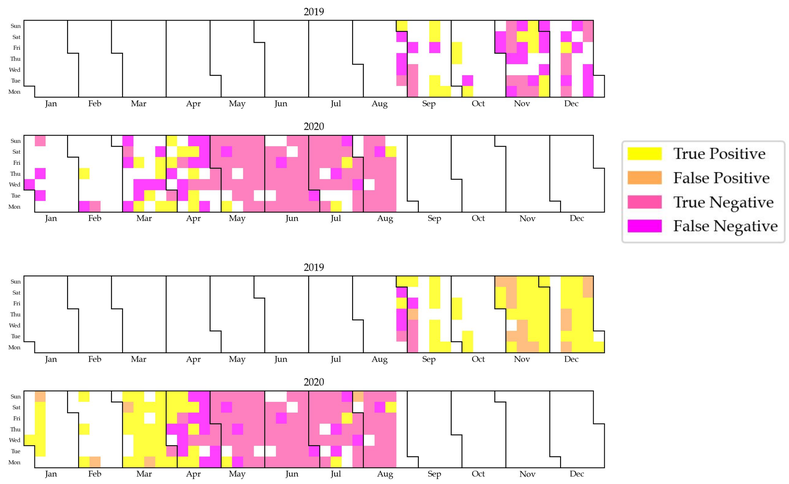

Figure 9. Evaluation of FIRST (top) and our model. (bottom)

Accuracy: 73%

43	0
58	120

113	23
27	103

Accuracy: 81%

Figure 10. Confusion matrices for FIRST (left) and our model (right).

Conclusions

The onset altitude climatology is consistent with previous studies. The day of the year hugely contributes to the occurrence predictions. Moreover, they can be greatly improved by adding the virtual height and solar flux index as input features. Our preliminary results suggest that the predictive power of our model is superior to FIRST but further analysis is required to validate this claim.

Acknowledgements

Special thanks to Alejandro Palacios for the assistance provided with the datasets. We also thank the support with the Jicamarca Radio Observatory staff. The Jicamarca Radio Observatory is a facility of the Instituto Geofisico del Peru operated with support from the NSF AGS-1732209 through Cornell University.

References

[1] Anderson, D. N., and R. J. Redmon (2017), Forecasting scintillation activity and equatorial spread F, *Space Weather, 15*, 495–502, doi:10.1002/ 2016SW001554.
[2] Bishop, C. M. (2006), *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer.
[3] Chapagain, N. P., B. G. Fejer, and J. L. Chau (2009), Climatology of postsunset equatorial spread F over Jicamarca, *J. Geophys. Res., 114*, A07307, doi:10.1029/2008JA013911.
[4] Fejer, B. G., Scherliess, L., and de Paula, E. R. (1999), Effects of the vertical plasma drift velocity on the generation and evolution of equatorial spread *F, J. Geophys. Res.*, 104(A9), 19859– 19869, doi:10.1029/1999JA900271.
[5] Hysell, D. L., & Burcham, J. D. (2002). Long term studies of equatorial spread F using the JULIA radar at Jicamarca. *Journal of Atmospheric and Solar-Terrestrial Physics, 64* (12-14), 1531-1543. https://doi.org/10.1016/S1364-6826(02)00091-3
[6] Kelley, M. C. (2009). *The earth's ionosphere: Plasma physics and electrodynamics* (2nd ed.). Academic Press.
[7] Paszke, A. et al. (2019) "PyTorch: An Imperative Style, High-Performance Deep Learning Library", in Wallach, H. et al. (reds) *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., bll 8024–8035. Available at: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
[8] Pedregosa, F. et al. (2011) "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, 12, bll 2825–2830.
[9] Redmon, R. J., D. Anderson, R. Caton, and T. Bullett (2010), A Forecasting Ionospheric Real-time Scintillation Tool (FIRST), *Space Weather, 8*, S12003, doi:10.1029/2010SW000582
[10] Zhan, W., Rodrigues, F., & Milla, M. (2018). On the genesis of postmidnight equatorial spread F: Results for the American/Peruvian sector. *Geophysical Research Letters, 45*, 7354–7361. https://doi.org/10.1029/2018GL078822