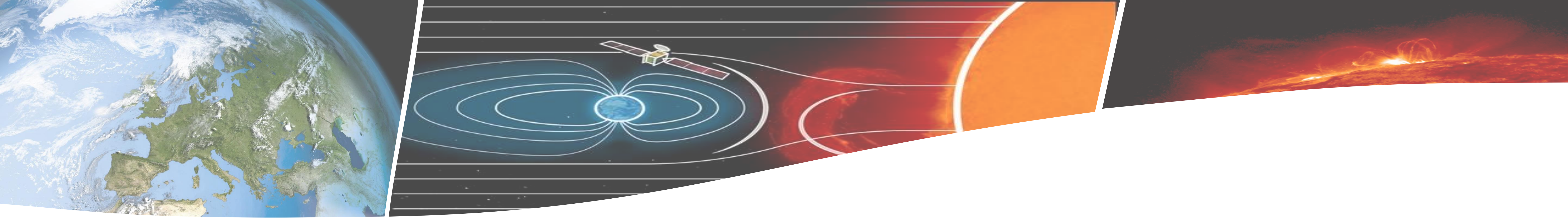


Statistical comparison of different methods for filtering solar activity in foF2 and hmF2 for long-term analysis at Juliusruh station

Lourdes M. A. Victoria [1,2,3], Maria G. Tan Jun Rios [3,4], Claudia Borries [3], and Ana G. Elias [1,2]

1 Department of Physics, Faculty of Exact Science and Technology (FACET), Universidad Nacional de Tucumán (UNT), 4000 Tucumán, Argentina
2 Laboratory of Ionosphere, Neutral Atmosphere y Magnetosphere (LIANM), FACET, UNT, 4000 Tucumán, Argentina
3German Aerospace Center (DLR), Institute for Solar-Terrestrial Physics, Neustrelitz, Germany
4Department of Earth and Planetary Science, Kyushu University, Fukuoka, Japan

Motivation

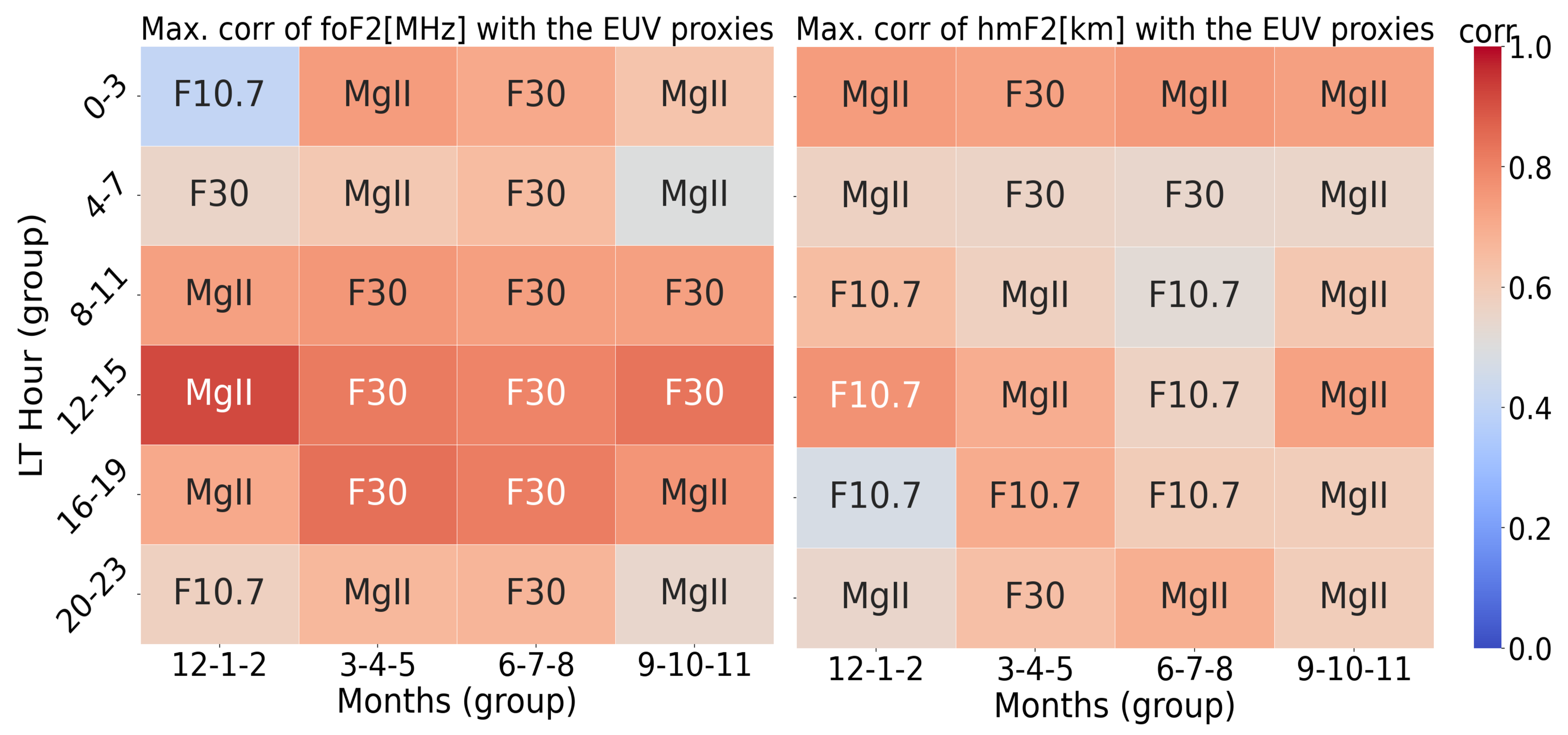
The ionospheric F2 layer plays a key role in radio wave propagation. It is characterized by the critical frequency (foF2) and the peak electron density height (hmF2), both of which are strongly influenced by solar activity. Accurately modeling their response to solar variability is essential for filtering out solar-driven effects and studying long-term trends. This work compares five models describing the response of foF2 and hmF2 to solar activity and evaluates different EUV proxies, selecting the most suitable one for subsequent long-term trend analysis.

Objective

This work aims to perform a statistical comparison of different methods and EUV proxies in order to model the response of foF2 and hmF2 to solar activity

EUV proxy comparison

An analysis of extreme ultraviolet (EUV) proxies, including MgII, F30, and F10.7, is carried out to identify the proxy with the highest correlation with the ionospheric parameters.



Heatmap of the maximum correlations between the EUV proxies and the ionospheric parameters (foF2 on the left and hmF2 on the right). The EUV proxy achieving the highest correlation is indicated in each panel,.

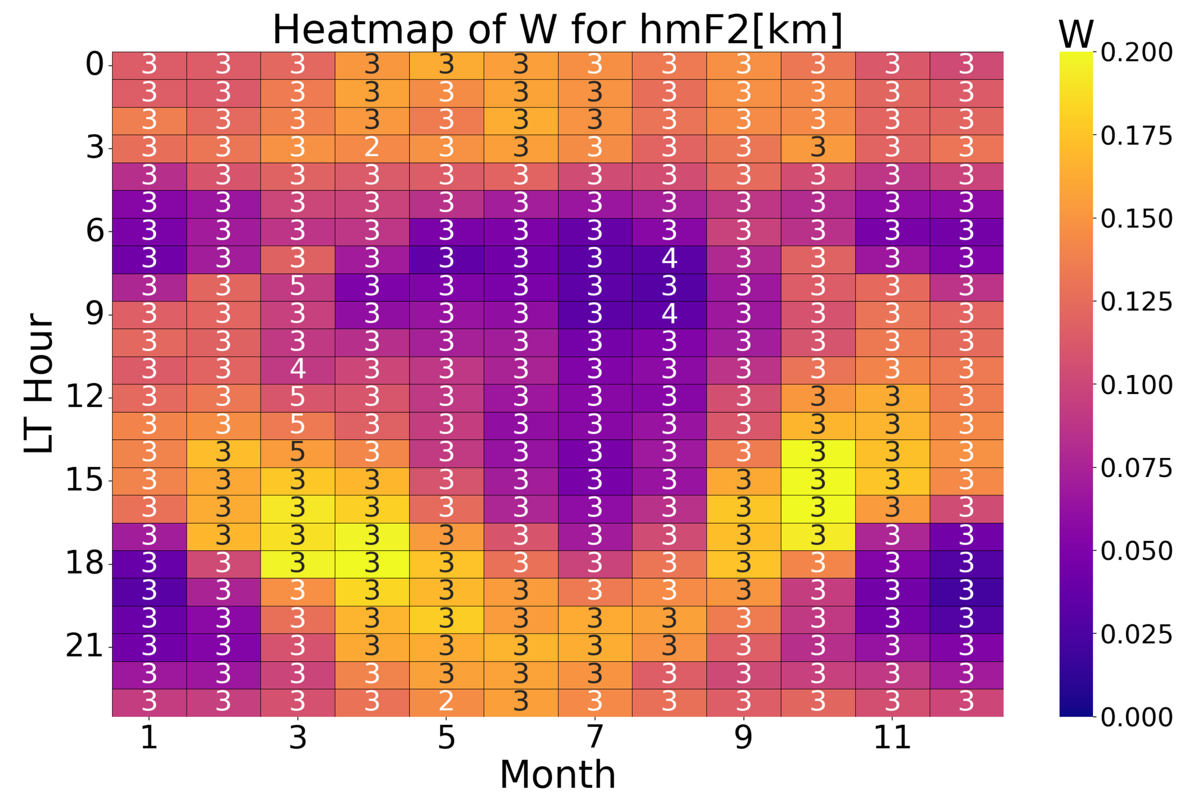
Model comparison

A comparison of five methods for modeling the impact of solar activity on foF2 and hmF2 is conducted using hourly data from Juliusruh. The models considered are three polynomial regressions based on data clustered by solar cycle, month and hour of the day, and two Fourier Series applied to data clustered by hour of the day.

Models	Method (X=EUV PROXY)
Lineal Regression (1)	$y = aX + b$
Quadratic Regression (2)	$y = aX^2 + bX + c$
Cubic Regression (3)	$y = aX^3 + bX^2 + cX + d$
2nd order Fourier Series (4)	$y = \sum_{n=0}^{2}\left[\begin{array}{l}(a_nX^3 + b_nX^2 + c_nX + d_n)\cos\left(\frac{2\pi n}{365}\,DoY\right)\\ +(e_nX^3 + f_nX^2 + g_nX + h_n)\sin\left(\frac{2\pi n}{365}\,DoY\right)\end{array}\right]$
3rd order Fourier Series (5)	$y = \sum_{n=0}^{3}\left[\begin{array}{l}(a_nX^3 + b_nX^2 + c_nX + d_n)\cos\left(\frac{2\pi n}{365}\,DoY\right)\\ +(e_nX^3 + f_nX^2 + g_nX + h_n)\sin\left(\frac{2\pi n}{365}\,DoY\right)\end{array}\right]$

The relation between two statistical tools such as the coefficient of determination (R^2) and mean absolute percentage error (MAPE) is employed for a statistical comparison of the five methods.

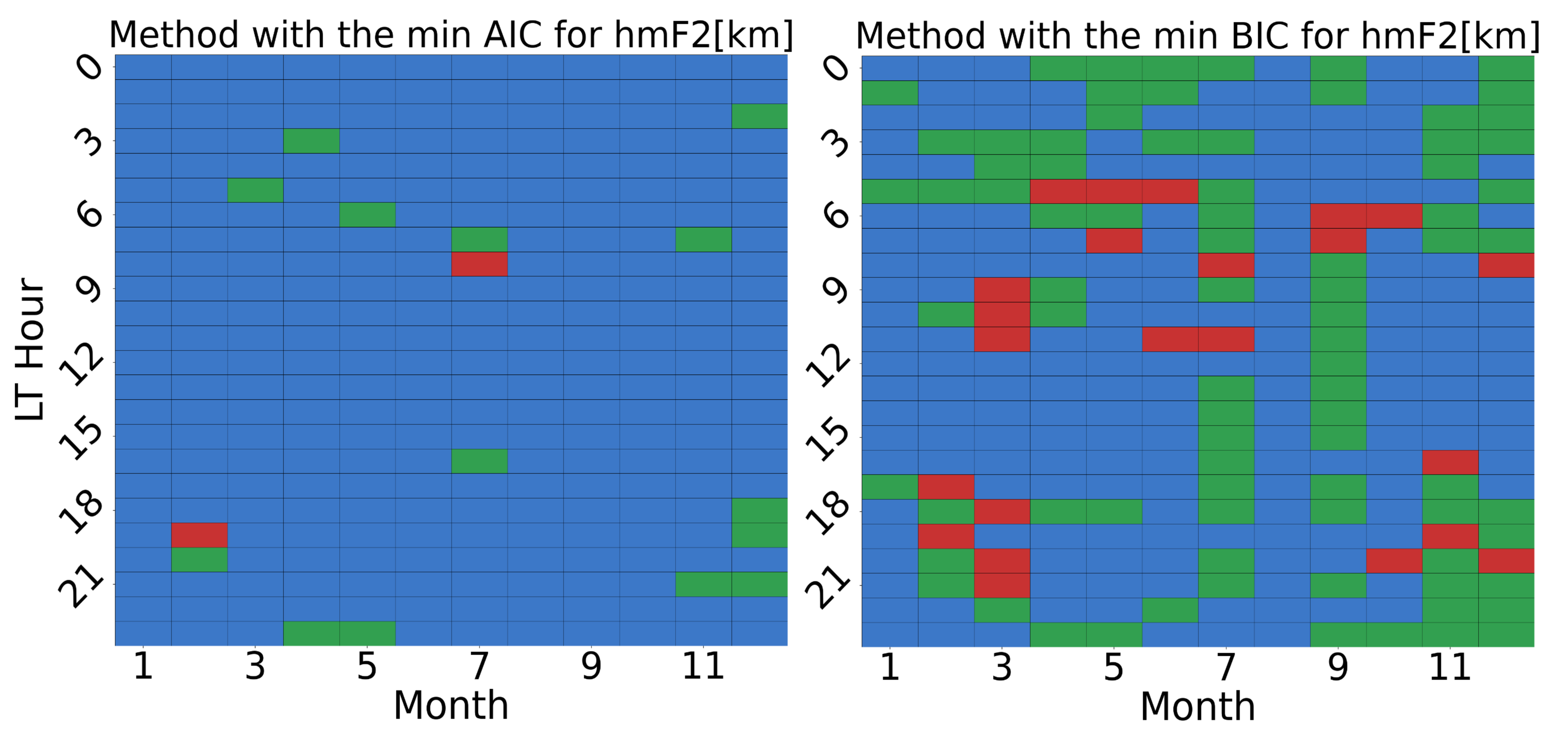
$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}; \qquad MAPE = \frac{100}{n} * \sum_{i=1}^{n}\frac{|(\hat{y}_i - y_i)|}{|y_i|}; \qquad W = \frac{R^2}{MAPE}$$



Heatmap of the highest W values for each monthly-hourly group for hmF2. The model with the highest W value for each group is indicated, in each panel, using a number indicated in the table

Also the Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are employed to study the balance between model complexity and explanatory power; and select the most appropriate method in these terms.

$$AIC = -2n * \ln\left(\frac{SSE}{n}\right) + 2k \qquad BIC = -2n * \ln\left(\frac{SSE}{n}\right) + k * \ln(n)$$



Colormaps for the comparison between polynomial regressions for AIC (left) and BIC (right) for hmF2. The color shown in each panel corresponds to the model with which the minimum value of AIC and BIC is achieved for that monthly-hourly group (red: linear regression, green: quadratic regression, blue: cubic regression)

Conclusion

The results indicate that F30 has the highest correlation with foF2 as an EUV proxy, while MgII shows the strongest correlation with hmF2. The third-degree polynomial regression is found to be the most effective method for modeling the ionospheric response to solar activity, based on a statistical analysis of the balance between model complexity and explanatory power.

References:
Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6), 716-723. https://doi.org/10.1109/TAC.1974.1100705
Chen, Y., Liu, L., Le, H., Zhang, H., & Zhang, R. (2022). Responding trends of ionospheric F2-layer to weaker geomagnetic activities. Journal of Space Weather and Space Climate, 12, 6. https://doi.org/10.1051/swsc/2022005
Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6 (2), 461-464. https://doi.org/10.1214/aos/1176344136
Tan Jun Rios, M. G., Borries, C., Liu, H., & Mielich, J. (2025). Long-term changes in the dependence of NmF2 on solar flux at Juliusruh. Annales Geophysicae, 43, 73-89. https://doi.org/10.5194/angeo-43-73-2025

Deutsches Zentrum für Luft- und Raumfahrt

Acknowledgments:
DLR-DAAD Research Fellowships – Graduates, 2024
Becas Estímulo a las Vocaciones Científicas (EVC-CIN).

Contact: L. Victoria
DLR, Institute for Solar-Terrestrial Physics
Email: lourdevictoria@dlr.de