Mining study of ionospheric data

Baume G.[1,2], Gularte E.[1,3], Carpintero D.D.[1,2]

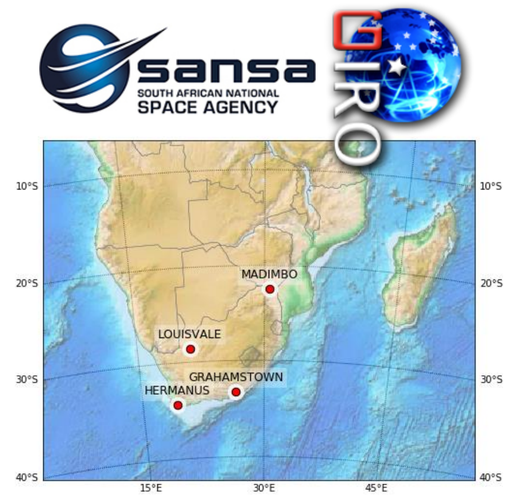
1- Facultad de Ciencias Astronómicas y Geofísicas, UNLP
2- Instituto de Astrofísica de La Plata, UNLP-CONICET
3- Grupo de Geodesia Espacial y Aeronomía, UNLP

Introduction

A preliminary study of different data mining techniques applied to the case of strong temporal dependence, large number of dimensions and irregular sampling is presented. In particular, data corresponding to the characterization of the Earth's ionosphere have been used.

General procedure

- Identification and elimination of outliers.
- Standardization and/or categorization of features.
- Application of **association rules** and **clustering** techniques ("Classical", "K-means", "Agglomerative") to describe the ionospheric behavior.
- Construction of **decision trees** ("Random Forest") to assess the relevance of the features and a possible linkage of ionospheric parameters with the geomagnetic activity.



Data

- **Source:** Global Ionospheric Radio Observatory (GIRO)
- **Period:** 2009-2018
  Solar cycle 24 sampled every 15[m]
- **Place:** South Africa
  Four locations at mid-latitudes

Association Rules

Non supervised method

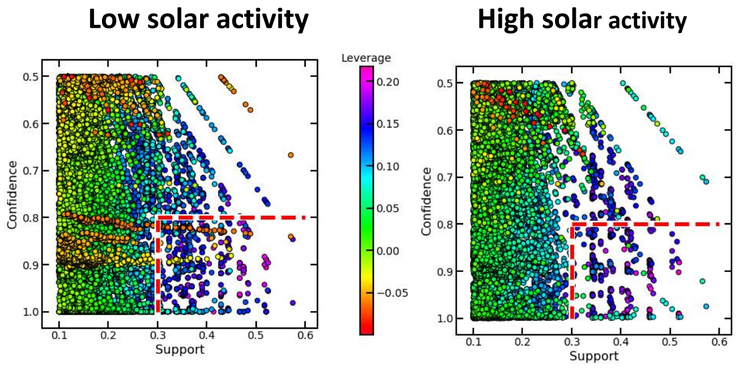
Data selection: Selected by year and by location.

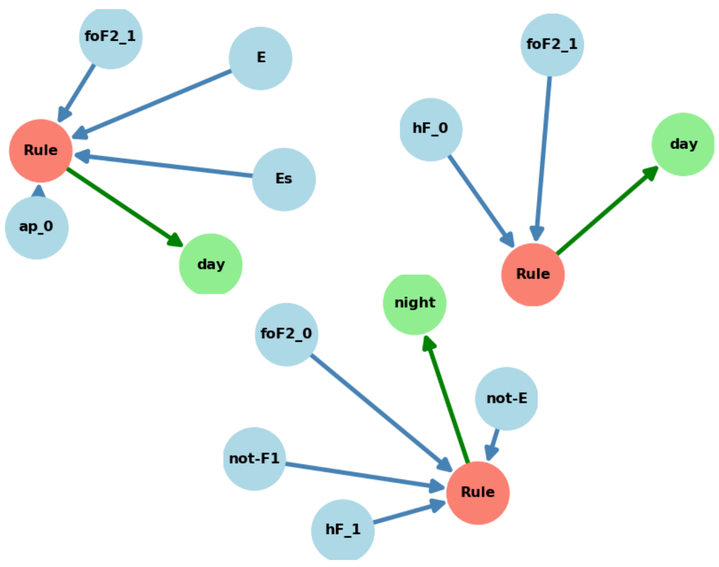
Features (categorical):

- Critical frecuency and thickness of F2 layer (“foF2“, “hmF2-hF“), F layer height (“hF“), and propagation factor of the F2 layer (“MD“). All of them were categorized in two levels (0: low, 1: high);
- E, Es and F1 layers presence (or not);
- Geomagnetic index “ap“ using three categories (0: low, 1: medium, 2: high);
- Sunlight presence (day or night).

Location: Grahamstown (GR13L)

Year 2009 **Low solar activity** — **Year 2014** **High solar activity**





Preliminary results

- The rules found are consistent with the behavior of the already known ionosphere.
- At high solar activity, the number of rules is smaller, indicating less number of links among the considered features.
- Metrics reveal the presence of rule groups linked linearly to each other.

Clustering techniques

Non supervised method

Data selection: Selected by year, by location and by month.

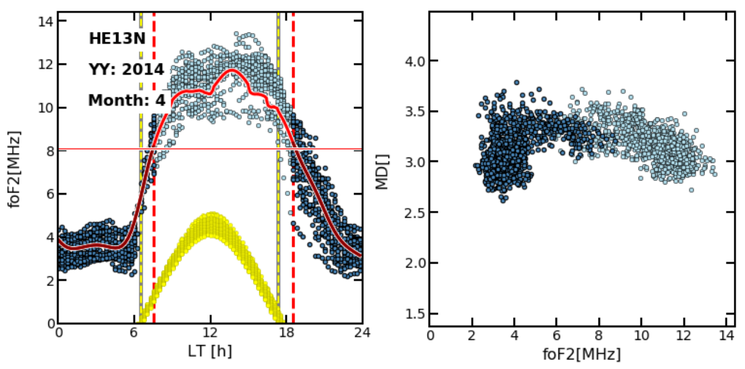
Features (numerical) :

- Critical frecuency of F2 layer (“foF2“);
- F layer height (“hF“);
- Propagation factor of F2 layer (“MD“);
- Local time (“LT“).

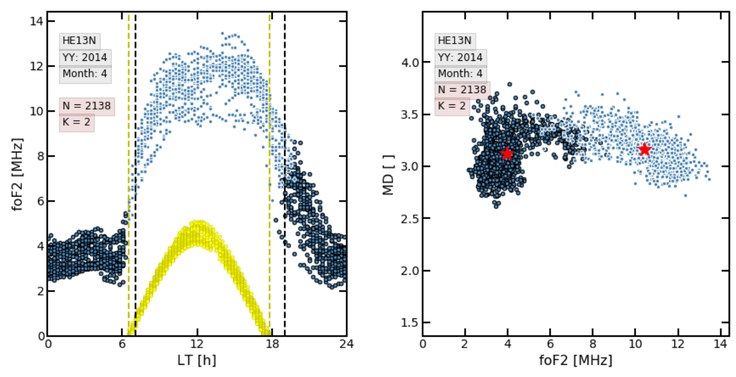
Location: Hermanus (HE13N)

Year 2014 (High solar activity)

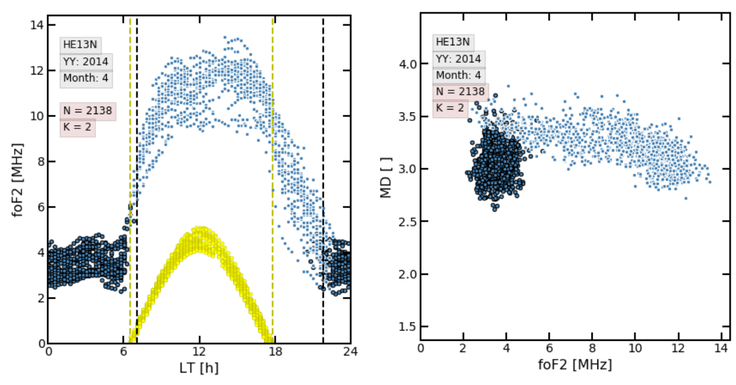
Classic method



K-means method



Agglomerative method



Preliminary results

- The groups found by the K-means method are similar, although somewhat better, than those obtained with the Classical method. In addition, the K-means method provides all the coordinates of the centers of the groups in the space of the features used.
- The Agglomerative method provides the best results, since it is based on the distance between the data of each group and not on the distance to a common center.
- The behavior of the clusters found is similar in all locations. This is consistent with always working at mid-latitude.

Decision Trees

Supervised method

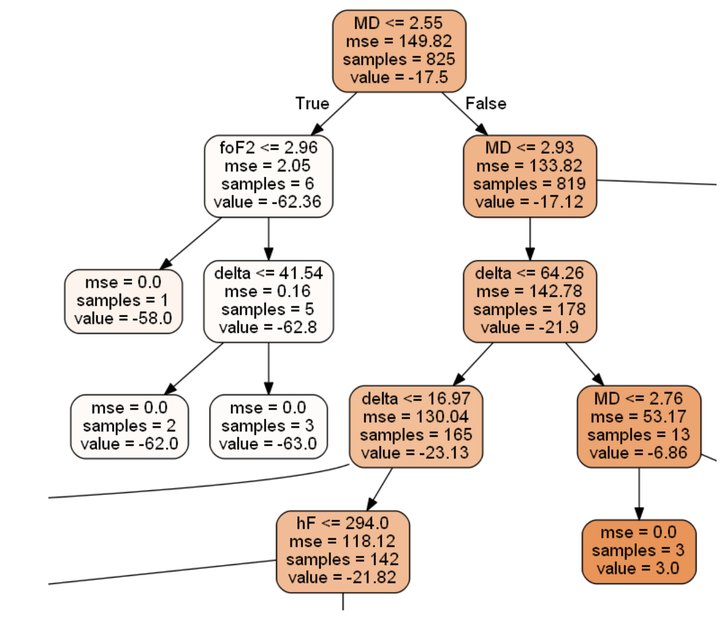
Data selection: Selected by year, by month and by location. Grouped by sunlight (day - night) and by F1 layer presence (or not) along the day.

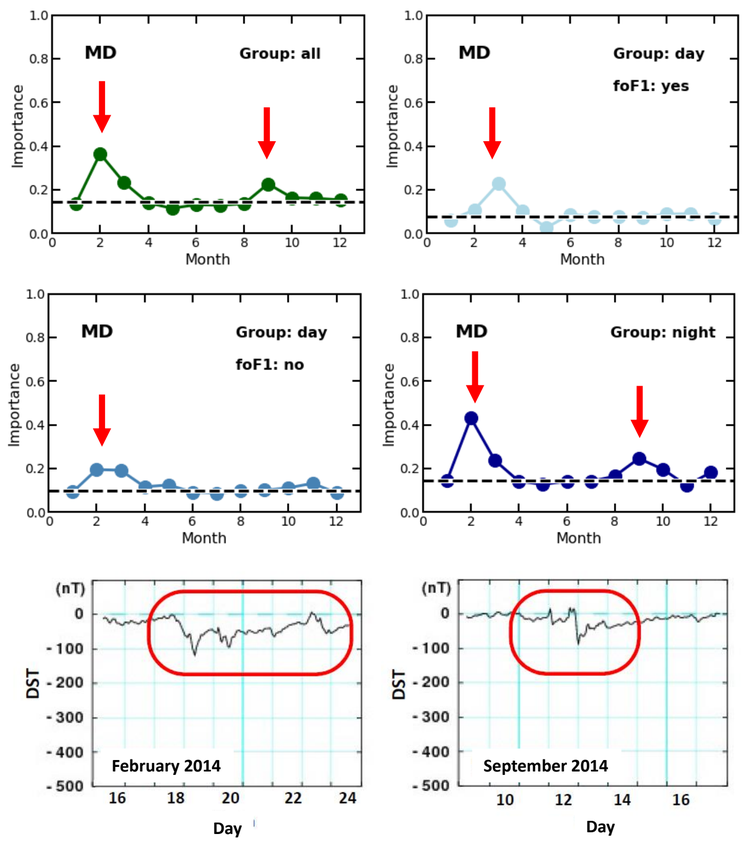
Input features: Frecuency, thicknesses and heights of the ionospheric layers, together with the local time (“LT“).

Output target: We tested several geomagnetic indices (both numerical and categorical). Best results were obtained with the disturbance storm-time index (“DST“) in numerical format.

Location: Grahamstown (GR13L)

Year 2014 (High solar activity)





Preliminary results

- "MD" and "foF2" features were always important in all the studied cases.
- Features associated with the heights of the layers (e.g. “MD”, “hmF2-hF”) seem to correlate with the geomagnetic storms.

Conclusions

- **Association rules** have the potential to find novel relations between features (or attributes).
- **Decision trees** allow to define the “importances” of the features and to associate the geomagnetic and/or solar activity with the behavior of the ionosphere.
- **K-means and Agglomerative methods** allow both to use all the features simultaneously.
- **Agglomerative method** provides the time interval indicating the action of the Sun over the ionosphere.

Contact: gbaume@fcaglp.unlp.edu.ar - erika@fcaglp.unlp.edu.ar - ddc@fcaglp.unlp.edu.ar