Classification of Sources of Ionospheric Scintillation in High Latitudes through Machine Learning

A.-M. Bals *(balsa@my.erau.edu)*, C. Thakrar, K. B. Deshpande
Embry-Riddle Aeronautical University, Daytona Beach, Florida

Intro

Global Navigation Satellite Systems (GNSS) experience a lot of positioning errors in high latitudes due to **ionospheric scintillations ->** carrier signal is distorted by rapid fluctuations in phase and amplitude

Clouds of ionized particles in the ionosphere have different scales and structures that define the signature of the distorted signal

By deciphering the signatures we can trace back to which kind of irregularity caused them

With this information we can tune and extend our models to different irregularities, as well as find out the physical parameters that drive and characterize them

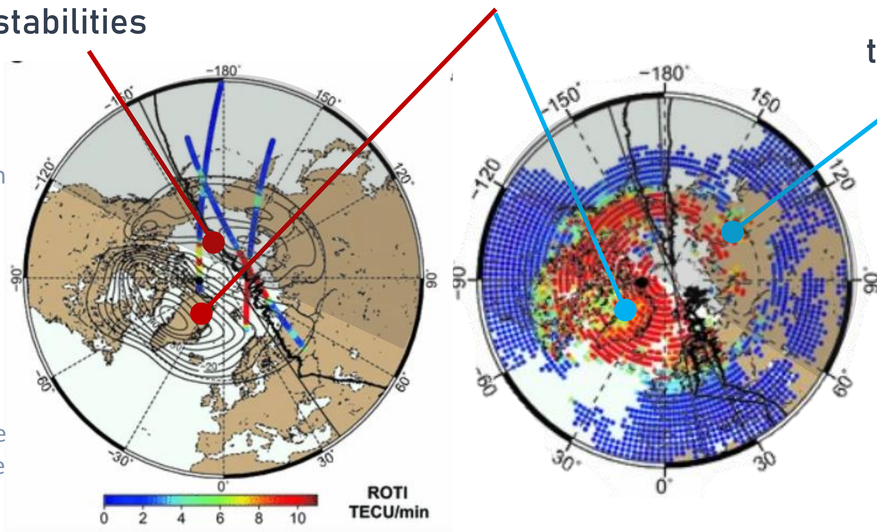
Motivation

3 different source regions for ionospheric scintillation in high latitudes:

Polar cap: (nightside/dayside border) gradient drift instability, Rayleigh-Taylor instabilities

Cusp region: (dayside) shear instabilites

Auroral oval: most frequently phase scintillations and in the dayside
Happening in parallel to auroral emission and therefore precipitation as dominant effect
Also Kelvin Helmholtz instability



SuperDARN polar potential maps for the Northern Hemisphere at 18.0 UT with superimposed low earth orbit (LEO) Rate of TEC (ionospheric total electron content) index (ROTI) (*colored lines*) and in situ (*thick black line*) observations. *Black dot* indicates the position of the magnetic pole. TEC data are the relative slant TEC measurements. *[1]*

Hourly rate of TEC (ionospheric total electron content) index (ROTI) maps over the Northern Hemisphers in geographical projections at 18 UT on March 17th 2015.
Geomagnetic poles are marked by *black dots*. [1]

Years of data from different stations/networks available
-> we need an **automatic** approach of categorizing

Conditions at 3 different regions vary a lot: Can we distinguish what **source region** the scintillation signature originated from? We know which station the signal is from -> can we write an algorithm and train it so it can **distinguish between regions**?

If we can categorize signals from different regions, can we find the criterion/**pattern** by which they can be distinguished or traced back to **different irregularities** and extract a whole database of events?

Are the signatures **distinguishable** at all? Can we trace them back to certain irregularity types?

Different effects in different regions: with at database of events per station, could we categorize them in terms of **temporal and spectral signatures** into the different types of irregularities that appear in that region with Machine Learning and analytics?

Case studies

March 9th, 2012, UTC 3-4: Tromso, Norway (auroral station) 66.7°N geomag.
and Ny-Alesund (polar cap station) 76.4°N geomag.

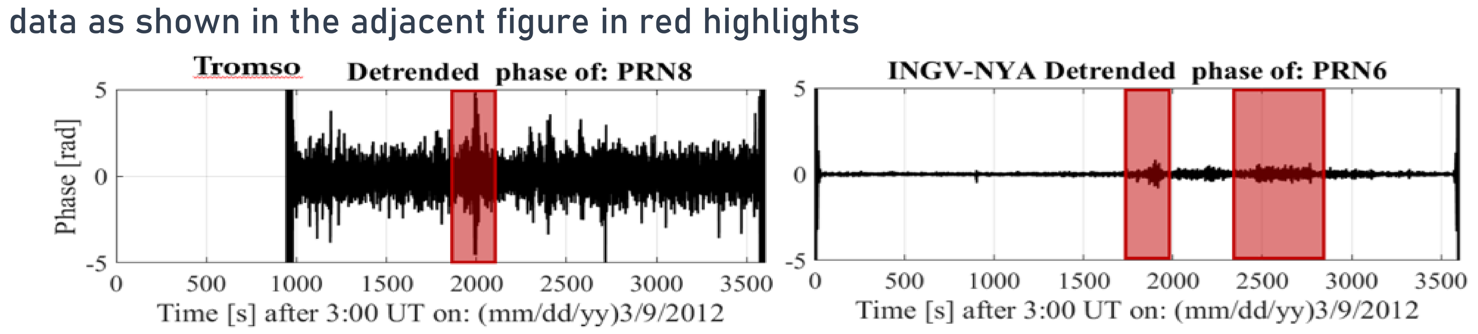
- Series of very active geomagnetic storms and substorms during 7 to 17 March 2012 [3]
- On March 9 Kp index was 6 -> indicates the beginning of a geomagnetic storm
- a lot of information available from the auxiliary observations
- March 9 has good high-rate scintillation measurements over several geographic regions
- Final selection for UTC 3-4 due to S4 and σϕ fluctuations, SuperDARN and ISR observations
- Continuous periods of scintillation > 30 [s] (requirement for inverse modeling and spectral analysis) [4]

March 17th, 2015, UTC 13-14: PokerFlat, Alaska (auroral station) 65.4°N geomag.
Resolute Bay, Arctic (polar cap station) 74.7°N geomag.

- Due to a solar event on March 15th, 2015 and labeled as a super storm

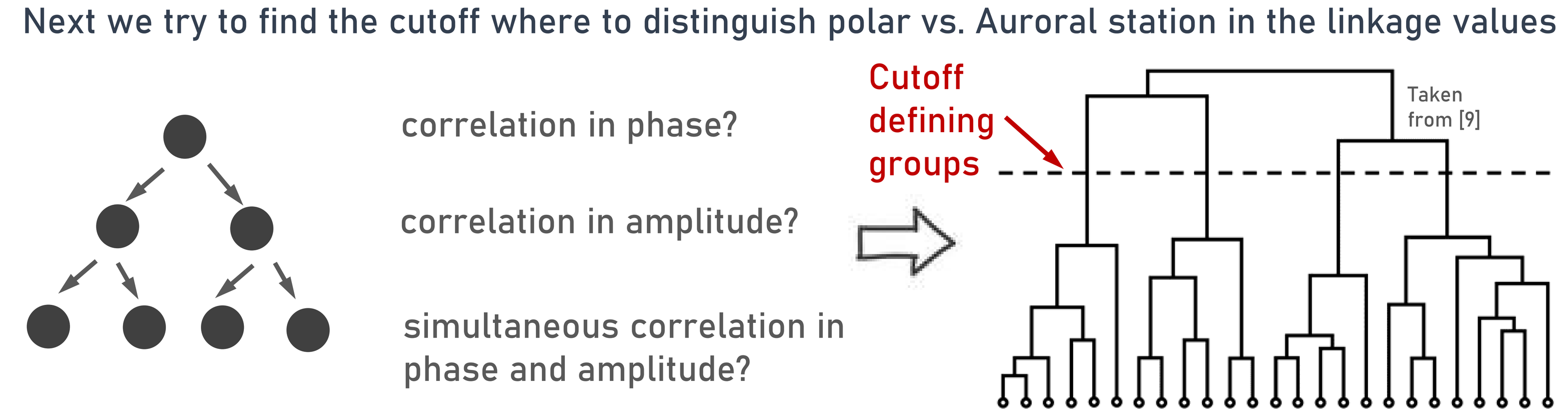
Decision Tree and hierarchical clustering

- IDEA: In different geomagnetic latitudes we expect different types and scales of irregularities -> the scintillation signatures in different regions will therefore look different, especially polar cap vs. auroral oval stations
- We will use a decision tree hierarchical clustering approach to find the events that look similar and cluster them into two classes: polar cap station vs. auroral oval station
- Scintillation events are manually selected and extracted from the amplitude and phase data as shown in the adjacent figure in red highlights



Two PRNs in phase for Tromso (left) and Ny-Alesund (right)

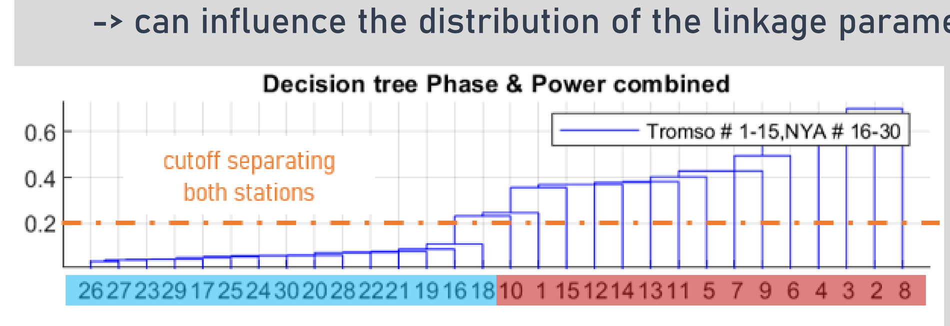
- We will use a decision tree hierarchical clustering approach to find the events that look similar and cluster them into two classes: polar cap station vs. auroral oval station
- Events will be split up into 50 [s] time segments to be comparable in analysis
- Matlab *linkage* calculates the correlation matrix for power and phase segments (15 per station per run) and determines how similar they are to each other (linkage parameter). It compares amplitude, phase and combined amplitude and phase time series.
- Next we try to find the cutoff where to distinguish polar vs. Auroral station in the linkage values

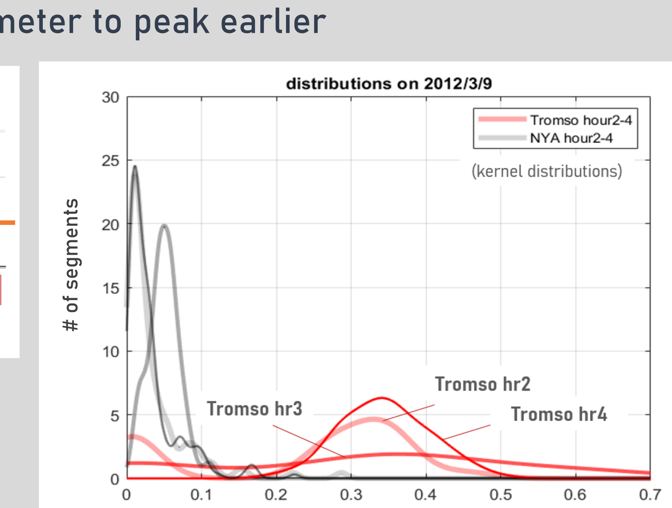


- Matlab ´linkage' calculates the correlation between the data and forms groups of signatures that are most alike. The linkage value describes how far the Euclidean distance in this plane is from the other signatures in its groups and as well as those of the other groups

Preliminary results

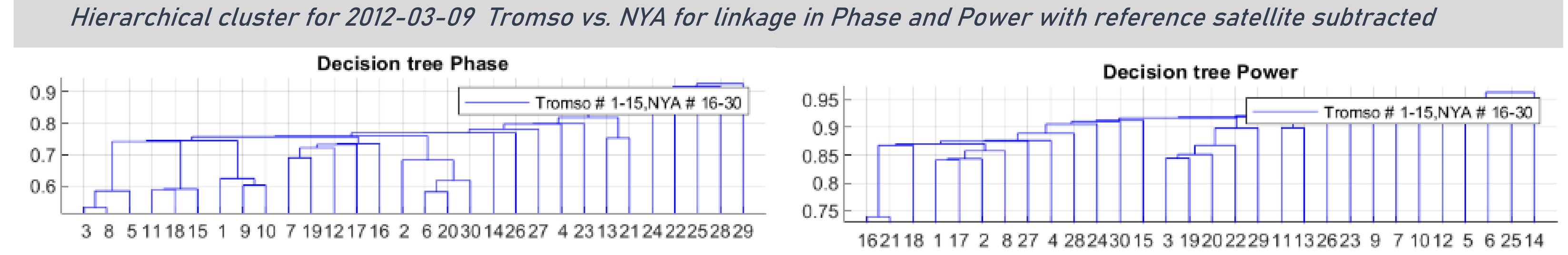
- For the northern hemisphere, the polar cap station (NYA) can be very well separated from the auroral oval station (Tromso) in 2012. This is due to very similar/alike events happening in Tromso leading to a high correlation, vs. Different scintillation signatures in NYA.
- Depending on how well the actual phase of activity containing scintillation was extracted from the background signal of the PRN, very low linkage values can appear for all stations that are just segments that contain mostly background
  -> can influence the distribution of the linkage parameter to peak earlier





Hierarchical clustering of the ´linkage´ comparing compared correlation of phase and amplitude of Tromso and Ny-Alesund with the cutoff at 0.2 linkage (left) and distribution of combined linkage over 3 hours on 2012-03-09 for Tromso and NYA (right)

Hierarchical cluster for 2012-03-09 Tromso vs. NYA for linkage in Phase and Power with reference satellite subtracted

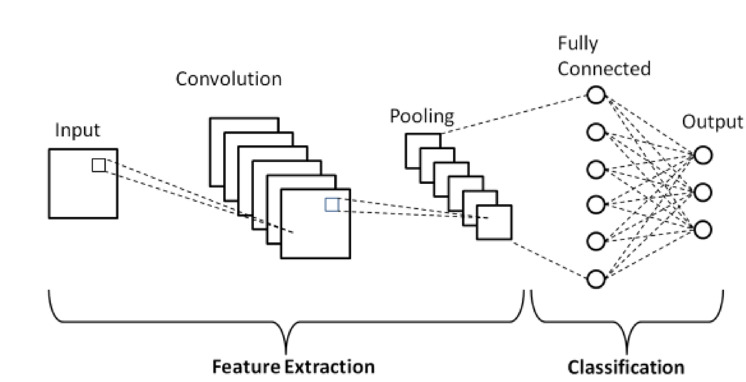


Machine Learning

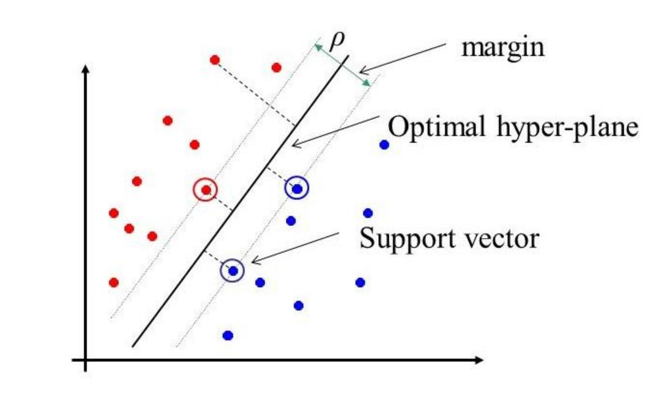
Supervised: re-finding patterns from a training set of data in a new set of data
vs
Unsupervised: finding patterns in a dataset and learning to distinguish based on automatic iterations

The task of Machine Learning is the prediction, clustering and classification – but not the explanation

- **Mc Granaghan et al:** binary scintillation detection with **support vector machine** and set a benchmark for prediction performance – detecting if scintillation is present or not using a lot of auxiliary channels including solar wind, geomagnetic and interplanetary features
- **Linty et al:** Detection of scintillation events by use of a **decision tree clustering**. Much less computationally demanding. Instead of S4 and Sigma Phi, it is much more efficient and better performance using the high rate data and averaging it directly, basically decomposing the components of S4 and Sigma Phi
- **Maimaiti et al:** Detection of substorm onsets with a Convolutional Neural Network. 5 channels of years of prelabeled onset geomagnetic data.



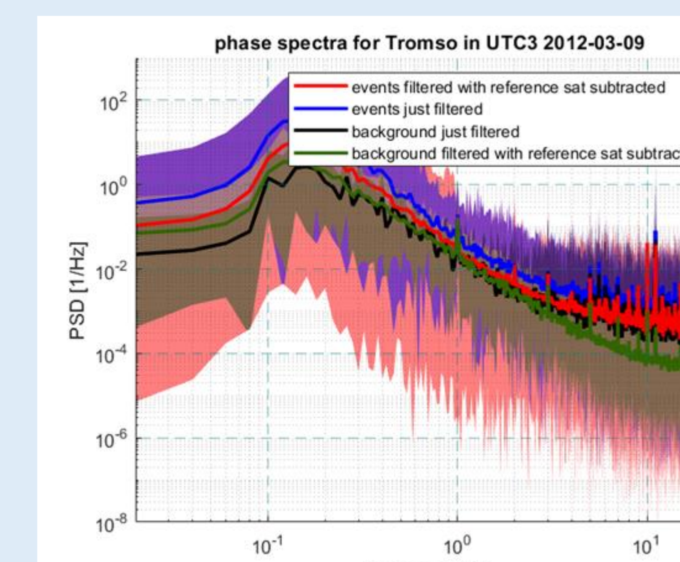
Schematic of the layers of a convolutional neural network [7]



Schematic of the classification in a Support Vector Machine [8]

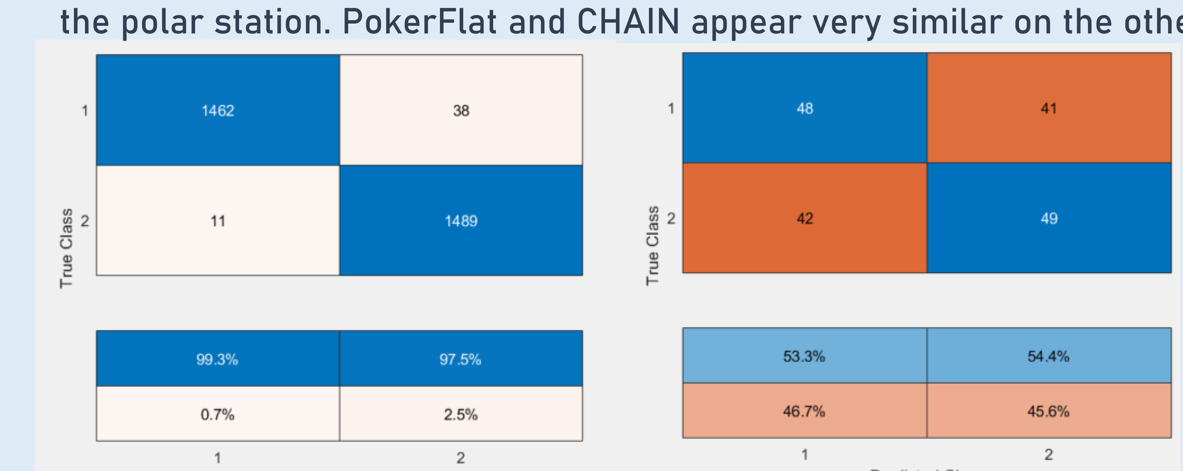
Challenges

- Every receiver and even every PRN has a different noise level
- If we are comparing time series from different stations we will need to try and reduce the noise as far as we can
- If there is enough data available, the background noise level can be determined from PRNs that are not in the line of sight of the irregularity with a so called reference satellite: cutting out segments of background data that do not contain any signals and subtract them from the pieces that have signal in them to get rid of the receiver noise
- If there is too much noise in the data, the *´linkage'* will pick up an artificial correlation that is not actually there and will make it look like there is a correlation



Spectral analysis of Tromso (auroral station) high rate data on March 9th 2012 with and without noise elimination through a reference satellite.

- This is most likely the reason for the big difference in the confusion matrices below displaying the detection performance. For the 2012 case, there was a lot of oscillating receiver noise in the Tromso data and therefore its correlation values are much higher and easily to distinguish from the polar station. PokerFlat and CHAIN appear very similar on the other hand in 2015.



Confusion matrix for 2012-03-09 Tromso vs. NYA (left) and for 2015-03-17 CHAIN vs. PokerFlat (right)
Left: Class 1 = Tromso (auroral station), Class 2 = NYA (polar station)
Right: Class 1 = CHAIN (polar station), Class 2 = PokerFlat (auroral station)

REFERENCES

[1] Cherniak, Zakharenkova: https://doi.org/10.1186/s40623-016-0506-1
[2] Wu et al. : https://doi.org/10.1186/s40623-016-0525-y
[3]: Prikryl et al., 2015
[4]: Deshpande, Bust, Clauer, et al. https://doi.org/10.1002/2016JA022943
[5]: Linty et al., DOI: 10.1109
[6]: Mc Granaghan et al., https://doi.org/10.1029/2018SW002018
[7]: Phung, Rhee: DOI: 10.3390/app9214500
[8]: https://www.mitosistech.com/support-vector-machine/
[9]: Telesford et al: DOI: 10.1089/brain.2011.0055