3rd URSI AT-AP-RASC, Gran Canaria, 29 May – 3 June 2022

Machine learning guided by ray-tracing for mid-latitude SuperDARN backscatter classification

Bharat S. R. Kunduri[(1)], Joseph B. H. Baker [(1)], J. Michael Ruohoniemi [(1)], Simon G. Shepherd[(2)], and Evan G. Thomas[(2)]
(1) Virginia Tech, Blacksburg, VA, USA, e-mail: bharatr@vt.edu; bakerjb@v.edu; mikeruo@vt.edu
(2) Dartmouth College, Hanover, NH, USA; e-mail: simon.g.shepherd@dartmouth.edu; evan.g.thomas@dartmouth.edu

The Super Dual Auroral Radar Network (SuperDARN) is a network of High Frequency (HF) radars that are used for monitoring plasma convection in the Earth's ionosphere. A majority of SuperDARN backscatter can broadly be divided into three categories: 1) ionospheric scatter due to reflections from plasma irregularities in the E and F regions of the ionosphere, 2) ground scatter due to reflections from the ground, and 3) backscatter from meteor trails left by meteoroids as they enter the Earth's atmosphere. While all these scatter types are useful for scientific analysis, it is however, not straightforward to distinguish between these different categories, especially at mid-latitudes. The traditional method used to distinguish between ionospheric and ground backscatter relies on thresholds for the magnitude of line-of-sight velocities and spectral width of the radar measurements. At high-latitudes the magnitude of velocities associated with ionospheric backscatter is usually much higher (hundreds of m/s) compared to the ground (a few tens of m/s), and the traditional approach works well under such conditions. However, at mid-latitudes, the traditional approach misclassifies much of the ionospheric scatter as ground, because both types of scatter exhibit very similar characteristics with low velocities and spectral width.

In this study, we present a new machine learning based technique for the classification of mid-latitude SuperDARN data into E- and F-region ionospheric backscatter, ground scatter, meteor scatter and potential regions of mixed scatter. Our algorithm takes a two-step approach. First, we train a neural network with three intermediate layers on a "synthetic" dataset generated using ray-tracing simulations mimicking the radar. The network has two sets of outputs: 1) a classification layer with a sigmoid activation function, and 2) a regression layer with a tanh activation. The classification layer predicts the probabilities of different types of scatter (E-/F-region ionospheric, ground etc.), and the regression layer predicts the refractive index in the scattering volume, reflection altitude and elevation angle associated with the scatter. The predicted refractive index from the regression layer can be useful in determining the state of the ionosphere, whereas the predicted elevation angles are useful for validating the model predictions. In the second step, we use the Density-based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to cluster the outputs of the classification layer (from the neural network) and actual measurements of line-of-sight velocities and spectral width from the radar. Each of these clusters are be classified into different categories based on the characteristics of the majority data points of the cluster. The new algorithm can distinguish between backscatter from meteor trails and E-region in the near ranges, and the amount of F-region ionospheric scatter identified increases by more than twice.