# Ionospheric scintillation prediction using gradient boosting algorithm

**Xiukuan Zhao, Guozhu Li, Haiyong Xie, Lianhuan Hu, Wenjie Sun, Sipeng Yang, Yi Li, Baiqi Ning**

**Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China ( Email: zxk@mail.iggcas.ac.cn )**

Ionospheric scintillation is the rapid fluctuation of radio signals traversing through ionospheric irregularities. Severe scintillation can cause loss of lock for the systems using Global Navigation Satellite System (GNSS) signals. The dependences of scintillation on seasonal, solar and geomagnetic activities have been widely studied, but its day-to-day variability and prediction still remain a challenge. The relationship between scintillation occurrences and a variety of factors is complex. The machine learning algorithm could handle nonlinear problems and thus uncover the implicit correlations between multiple factors. The gradient boosting, which is a type of machine learning technique, has been demonstrated to be effective in many fields, such as in the field of intrusion detection. Here we employ the gradient boosting algorithm, together with long-term observations in the Brazilian longitude sector to investigate if the day-to-day occurrence of low latitude ionospheric scintillation could be predicted.

## Geographic distributions of data

The ionospheric scintillation data used in this study were collected from eight GPS receivers belonging to the Low-Latitude Ionospheric Sensor Network (LISN). Geographic distributions of the Digisonde (red asterisk) and GPS receivers (blue circles) are shown in Figure 1.

Figure 2a shows an example of S4 (gray dots) and its peak value (red dot) at dip lat. -12° during pre-midnight hours on March 25, 2015. Figure 2b shows the temporal variation of TEC (gray curve) and its moving average (thick gray curve) using a time window of ten minutes. The variability of the F layer critical frequency (foF2) and peak height (hmF2) as a function of local time is shown in Figure 2c-d.

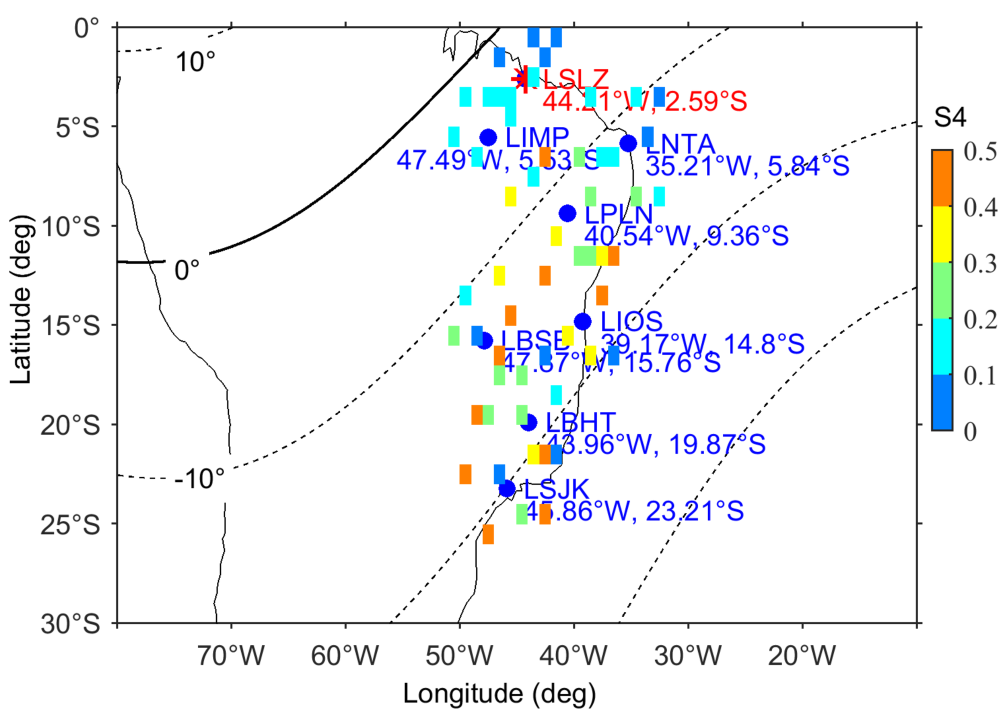


Figure 1

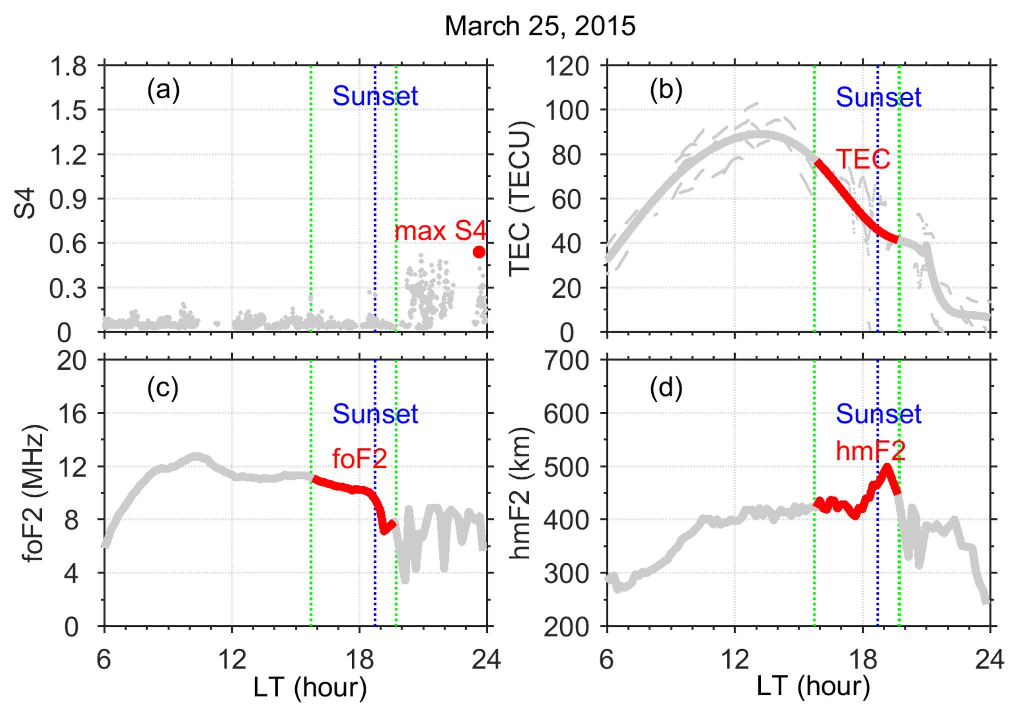


Figure 2

## Construction of prediction model

In general, the construction process of the prediction model is shown in Figure 3. Firstly, the dataset is built with 82 values of the input parameters and 1 output. Secondly, the data obtained are separated as training set and test set. Thirdly, a five-fold cross validation (CV) approach is utilized to tune the model hyperparameters. Fourthly, using the optimal hyperparameters configuration, the optimal prediction model is fitted based on the training set. Finally, the test set is adopted to evaluate model performance according to the prediction results.

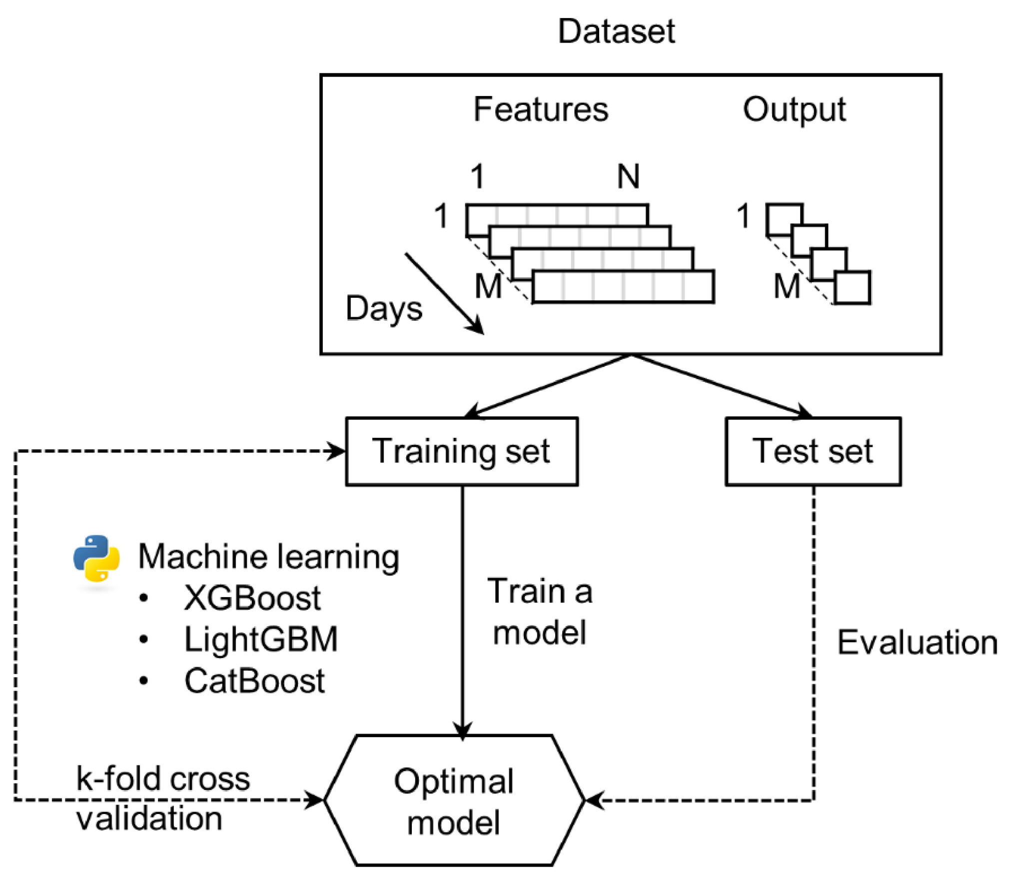


Figure 3

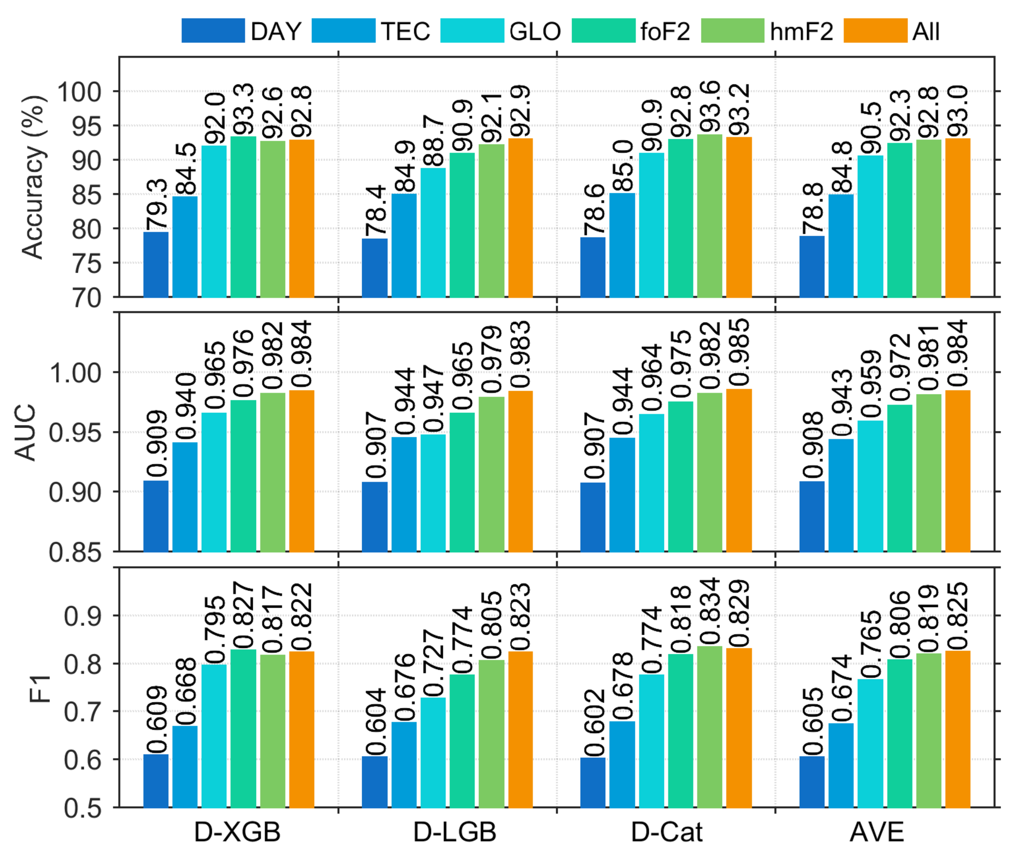


Figure 4

## Comparison of different groups of parameters

The accuracy, AUC and F1 of prediction models using different groups of parameters are shown in Figure 4.

The prediction results of D-XGB using different groups of parameters along with the observation are shown in Figure 5. In Figure 5a, the observational results clearly show the seasonal, latitudinal and solar activity dependences of scintillation. In Figure 5b, using the DAY group parameters, the model captures the seasonal and latitudinal variations. However, the solar activity dependence is not presented. From Figure 5c-g, the models capture the seasonal, latitudinal and solar activity dependences, and the consistency between the modeled and observational results is getting much better. Comparing Figure 5a with Figure 5g, it is clear that the model built with all the parameters shows a good agreement with the observation.

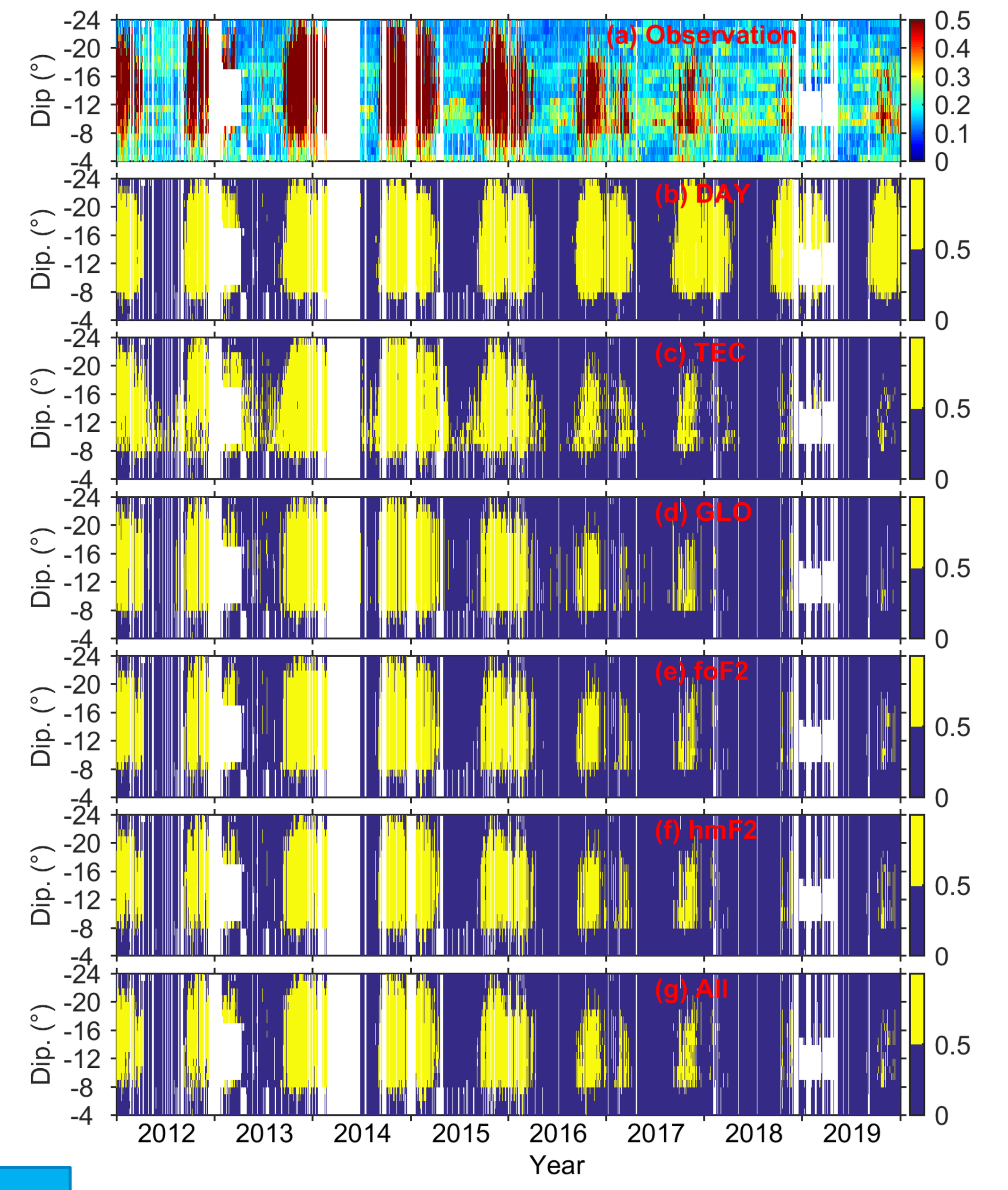


Figure 5

## Comparison of observations and prediction results

Figure 6a-b shows the observations in 2014 and the prediction results. In general, the prediction results show a good agreement with the observations in the seasonal and latitudinal dependences. Figure 6c shows the false positive (blue) and false negative (red) events as functions of month/day and latitude. The ROC curve for the T-Cat in 2014 is shown in Figure 6d. It can be noted that out of a total of 1514 strong scintillation events, 1247 (267) events were successfully (not) predicted. For the events without strong scintillations, 1740 events were identified, while 275 events were misclassified. The corresponding accuracy is 84.64%.

Figure 7 shows the results in 2020 which is similar to 2014.

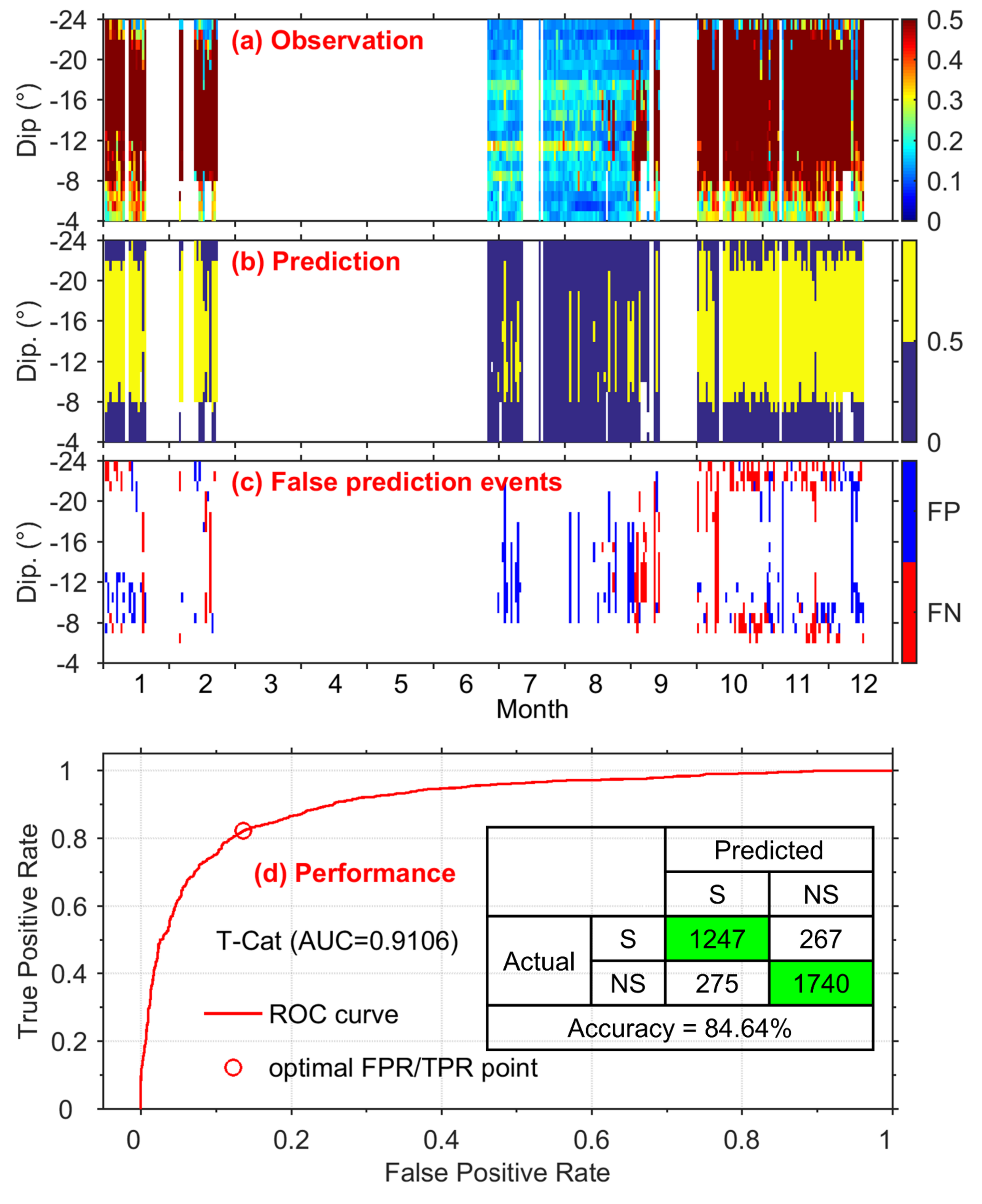


Figure 6

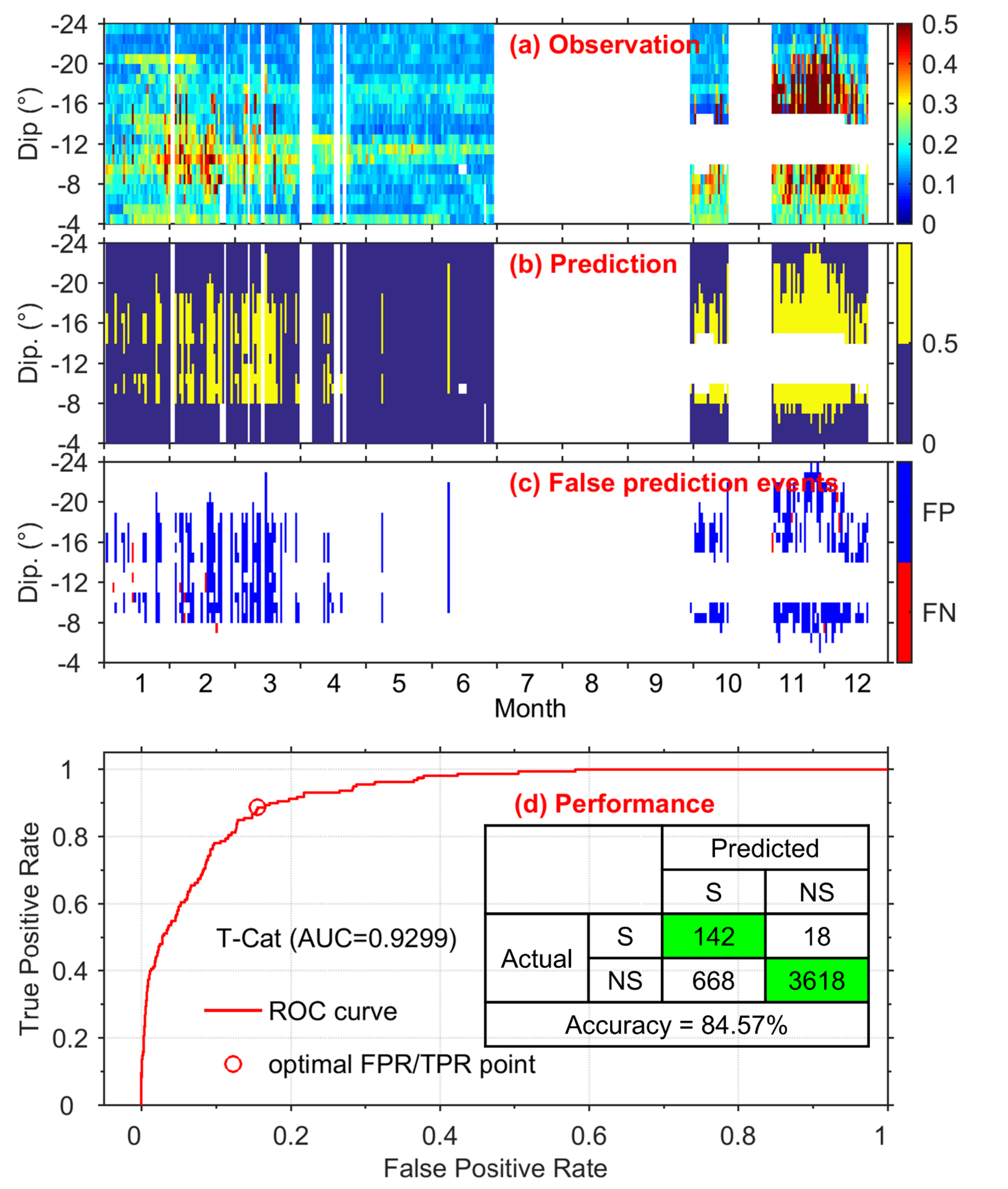


Figure 7

## Conlusion

The results show that with limited input parameters, the prediction accuracy for scintillation occurrence on a daily basis reach ~85%, suggesting that the gradient boosting algorithms are effective for predicting strong scintillations over low latitude. This opens a possibility for scintillation forecasting with acceptable accuracy under the conditions without physical model and powerful computing capability.

**This work is published in Space Weather ( https://doi.org/10.1029/2021SW002884 )**