# Statistical Modeling of Earth's Plasmasphere

Victoir Veibell

July 21, 2016

Introduction

- System
- Motivation
- Data
- Questions

Methodology

- Linear and ARX
- Neural Network
- Epoch
- Classification and Prediction

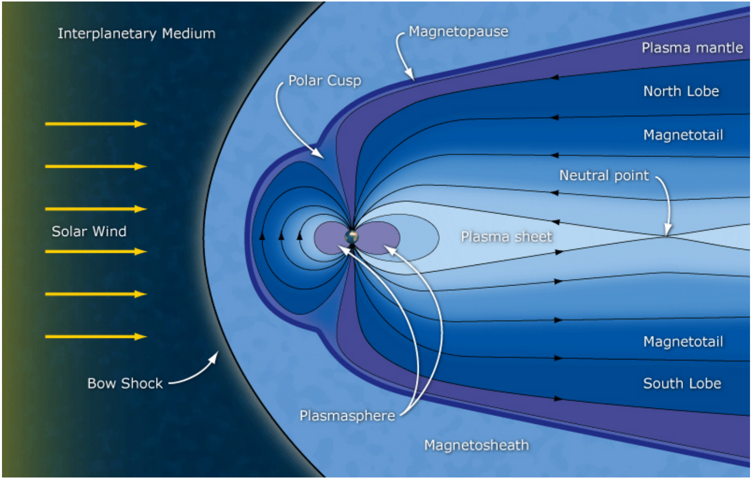
Results

- Traditional Models
- Epoch
- Classification and Forecasting

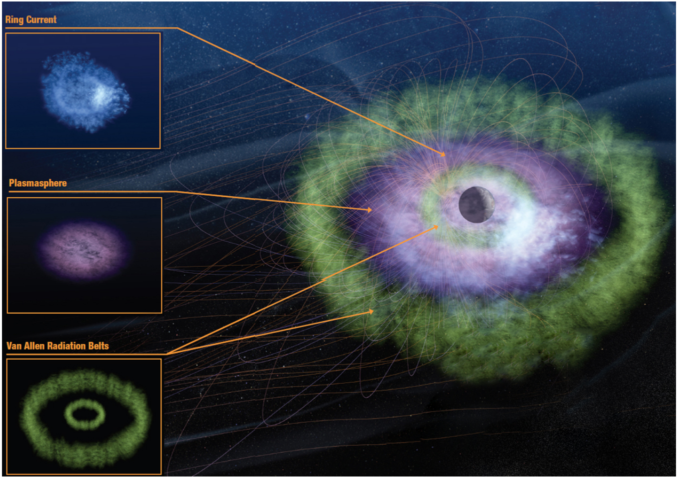
Conclusions

- Conclusions
- Future work

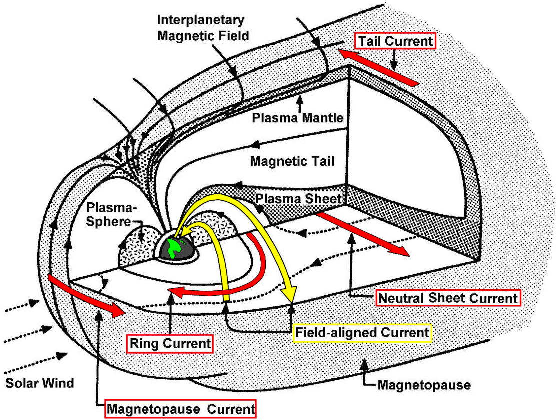
# Introduction



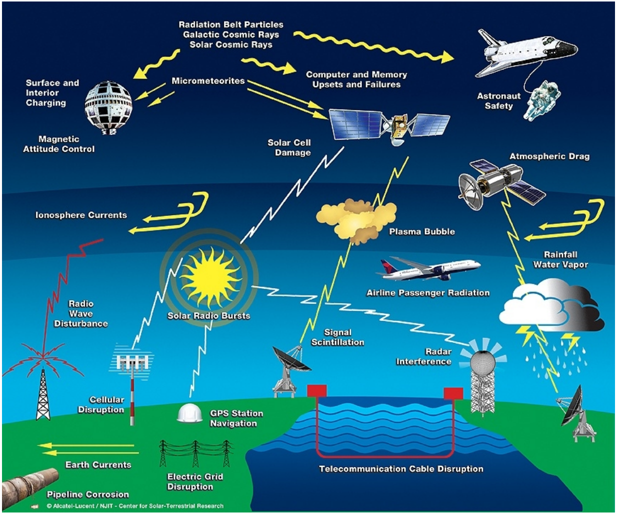

Overview of the magnetosphere and plasmasphere [Russel (2007)].




Overview of inner magnetosphere. Adapted from NASA.




Currents in/around the magnetosphere. Adapted from Maus (2010).




Impacts of Space Weather [Lanzerotti]

## Data

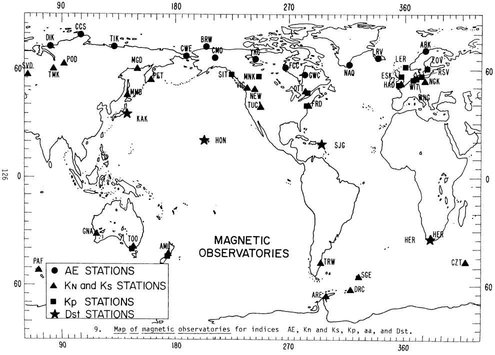
Data used come from three sources:

- Denton (2007) for $\rho_{eq}$, MLT, and $AE$
- King (2005) for $F_{10.7}$
- Kondrashov (2014) for $B_z$, $V_{SW}$, Kp, $\rho_{sw}$, and $D_{st}$

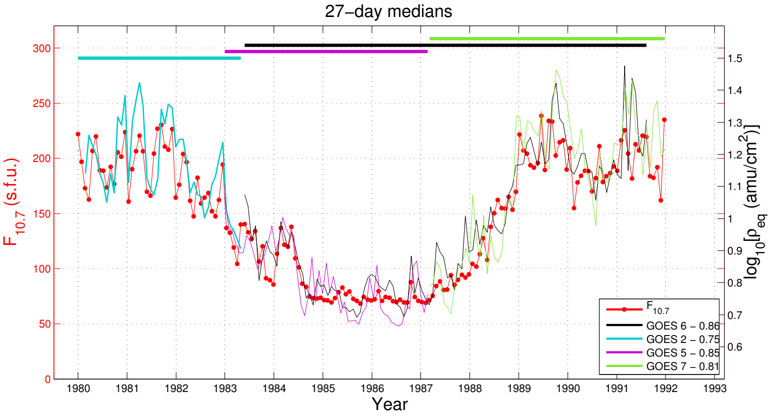
Data coverage:

- ∼ Denton (2007): 10 minute, non-uniform, non-complete, from 1980-1991, GOES 2, 5, 6, and 7
- ∼ King (2005): 1 hour uniform, non-complete, from 1972-2013
- ∼ Kondrashov (2014): 1 hour uniform, complete, from 1972-2013

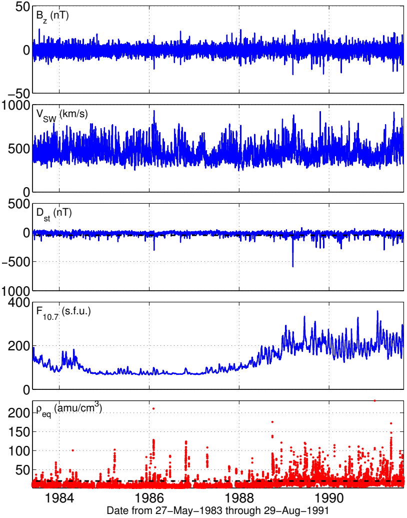
How we measure geomagnetic activity:




Map of ground stations used to measure the $K_p$, $AE$, and $D_{st}$ indices [Allen (1982)].

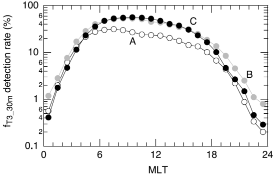


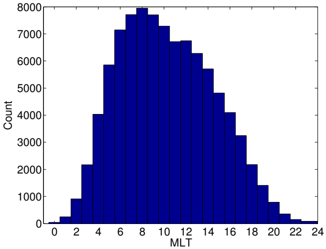

Comparing $F_{10.7\_27d}$ and $\log_{10}(\rho_{eq\_27d})$ using all available satellites.



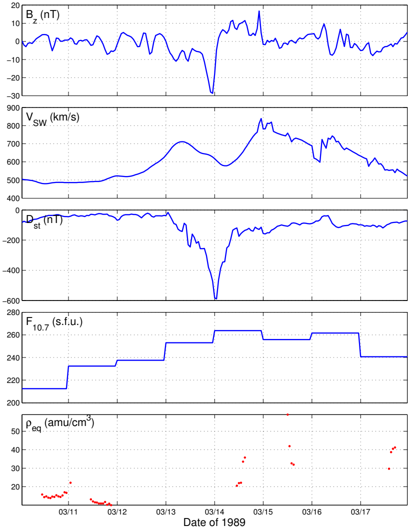

Data coverage with dashed lines indicating default event thresholds.

$\rho_{eq}$ is derived from toroidal harmonic frequencies in the plasmatrough. Harmonics are not always detectable.

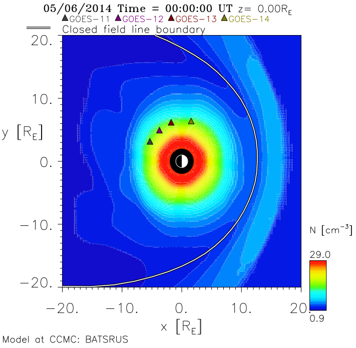


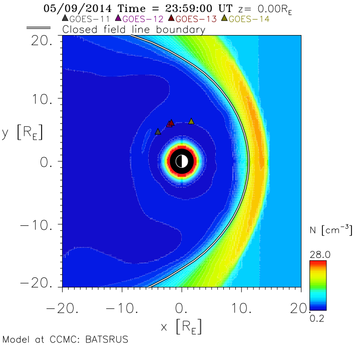




Left: Detection rate of $f_{T3\_30m}$ for magnetic latitudes of 5, 9, and 11 degrees (curves A, B, and C respectively) [Takahashi (2010)]. Right: MLT of all available $\rho_{eq}$ data.

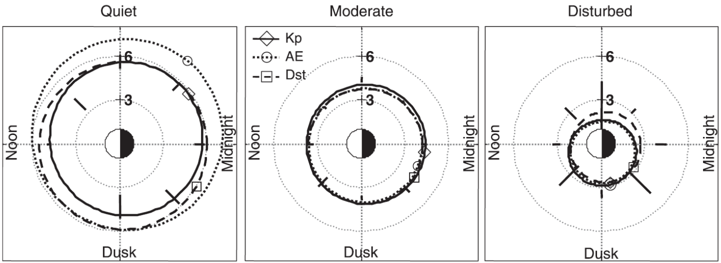



Data from GOES 6 around March 1989 geomagnetic storm.

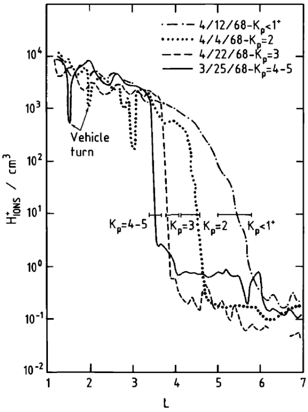



Model of magnetopause/plasmasphere before and after geomagnetic activity, showing location of GOES satellites in geosynchronous orbit [CCMC].




Model of plasmapause location as it varies with geomagnetic activity where the symbols indicate the local time of maximum plasmapause location [O'Brien and Moldwin (2003).




Plasmapause position varying with $K_p$ as represented by several particular plasmapause crossings made on outbound passes between local times of midnight and 0400 [Lemaire (1998)].

## Questions

Questions addressed by this dissertation:

- What enhances $\rho_{eq}$ in the plasmatrough? Solar wind (via geomagnetic storms) or internal processes (e.g. ionospheric outflow)?
- Does a $\rho_{eq}$ enhancement depend on current IMF conditions?
- Can a $\rho_{eq}$ enhancement be classified or forecasted?

# Methodology

Four main methods of analysis used:

1. Linear/Auto-Regressive with exogenous inputs (ARX)
2. Nonlinear Neural Network
3. Epoch
4. Nonlinear Classification and Prediction

# Linear and ARX

Linear uses Ordinary Least Squares on model of form $y = AX + c + \varepsilon$ with median of four hours before onset as X.

ARX starts with Box-Jenkins model:

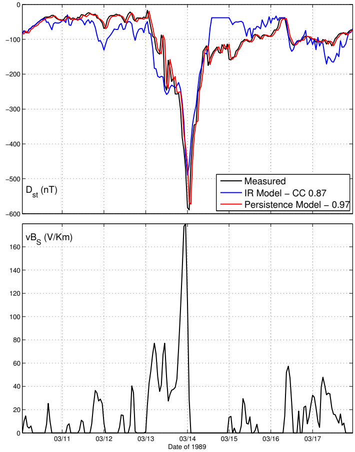
$$x(t) = \sum_{j=1}^{m} b_j \cdot f(t - j\Delta t) + c + \varepsilon_t$$

Modified to be an autoregressive model with exogenous inputs (ARX):

$$\hat{x}(t) = \sum_{i=1}^{l} a_i \cdot x(t - i\Delta t) + \sum_{j=1}^{m} b_j \cdot f(t - j\Delta t) + c + \varepsilon_t$$

Resultant matrix to solve:

$$\begin{pmatrix} x_0 & \dots & x_{\tau-1} & f_0 & \dots & f_{\tau-1} & 1 \\ x_1 & & x_\tau & f_\tau & & f_\tau & 1 \\ \dots & & & & & & \\ x_{N-\tau} & \dots & x_{N-1} & f_{N-\tau} & \dots & f_{N-1} & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ \dots \\ a_{\tau-1} \\ b_0 \\ \dots \\ b_{\tau-1} \\ c \end{pmatrix} = \begin{pmatrix} x_\tau \\ x_{\tau+1} \\ \dots \\ x_N \end{pmatrix}$$




Top: $D_{st}$ (black), persistence (red), 12-hour impulse response model (blue). Bottom: $vB_S$ input.

# Neural Network

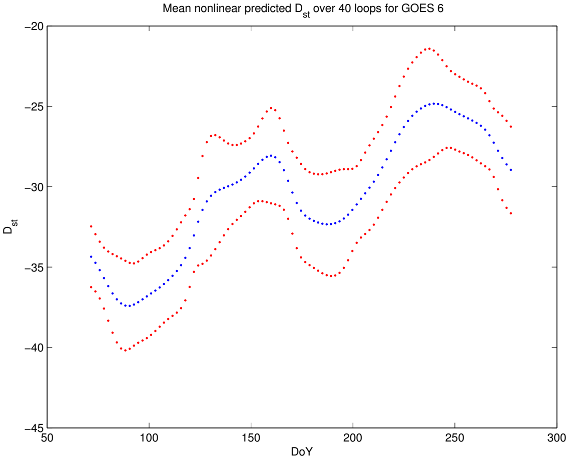
Neural Network model benefits:

1. Can model nonlinear effects
2. Easily adaptable to binary classification

Neural Network model disadvantages:

1. Susceptible to overfitting
2. More complex to create and analyze
3. No closed-form optimum solution

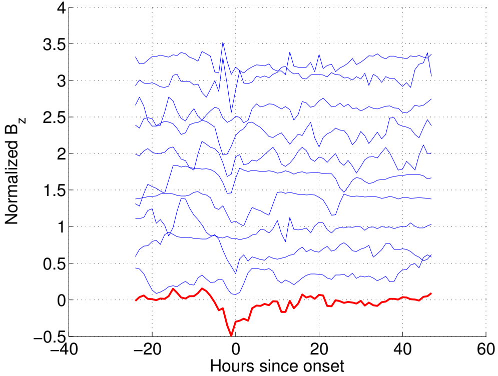
Example nonlinear effect:




$D_{st}$ predicted by nonlinear model of day of year.

# Epoch

Epoch Analysis averages multiple events to find patterns.




Sample epoch of first ten events and median

# Classification and Prediction

Is an onset distinguishable using just the variables provided?

| 0 | 0 | 0 | 1 |
|---|---|---|---|
| $B_{z_{\tau-3}}$ | $B_{z_{\tau-2}}$ | $B_{z_{\tau-1}}$ | $B_{z_\tau}$ |
| $F_{10.7_{\tau-3}}$ | $F_{10.7_{\tau-2}}$ | $F_{10.7_{\tau-1}}$ | $F_{10.7_\tau}$ |
| $K_{p_{\tau-3}}$ | $K_{p_{\tau-2}}$ | $K_{p_{\tau-1}}$ | $K_{p_\tau}$ |
| $\rho_{eq_{\tau-3}}$ | $\rho_{eq_{\tau-2}}$ | $\rho_{eq_{\tau-1}}$ | $\rho_{eq_\tau}$ |

| 0 | 0 | 0 | 1 |
|---|---|---|---|
| $B_{z_{\tau-3}}$ | $B_{z_{\tau-2}}$ | $B_{z_{\tau-1}}$ | $B_{z_\tau}$ |
| $F_{10.7_{\tau-3}}$ | $F_{10.7_{\tau-2}}$ | $F_{10.7_{\tau-1}}$ | $F_{10.7_\tau}$ |
| $K_{p_{\tau-3}}$ | $K_{p_{\tau-2}}$ | $K_{p_{\tau-1}}$ | $K_{p_\tau}$ |
| $\rho_{eq_{\tau-3}}$ | $\rho_{eq_{\tau-2}}$ | $\rho_{eq_{\tau-1}}$ | $\rho_{eq_\tau}$ |

Diagram of classification method.

Can an onset be forecasted using the previous four hours or days?

| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
|---|---|---|---|---|---|---|---|
| $B_{z_{\tau-7}}$ | $B_{z_{\tau-6}}$ | $B_{z_{\tau-5}}$ | $B_{z_{\tau-4}}$ | $B_{z_{\tau-3}}$ | $B_{z_{\tau-2}}$ | $B_{z_{\tau-1}}$ | $B_{z_{\tau}}$ |
| $F_{10.7_{\tau-7}}$ | $F_{10.7_{\tau-6}}$ | $F_{10.7_{\tau-5}}$ | $F_{10.7_{\tau-4}}$ | $F_{10.7_{\tau-3}}$ | $F_{10.7_{\tau-2}}$ | $F_{10.7_{\tau-1}}$ | $F_{10.7_{\tau}}$ |
| $K_{p_{\tau-7}}$ | $K_{p_{\tau-6}}$ | $K_{p_{\tau-5}}$ | $K_{p_{\tau-4}}$ | $K_{p_{\tau-3}}$ | $K_{p_{\tau-2}}$ | $K_{p_{\tau-1}}$ | $K_{p_{\tau}}$ |
| $\rho_{eq_{\tau-7}}$ | $\rho_{eq_{\tau-6}}$ | $\rho_{eq_{\tau-5}}$ | $\rho_{eq_{\tau-4}}$ | $\rho_{eq_{\tau-3}}$ | $\rho_{eq_{\tau-2}}$ | $\rho_{eq_{\tau-1}}$ | $\rho_{eq_{\tau}}$ |

Diagram of prediction method.

# Results

# Traditional Models

| | *Linear* | *Nonlinear* | $ARX_{24}$ |
|---|---|---|---|
| $DoY$ | $-0.06 \pm 0.06$ | $+0.32 \pm 0.22$ | $+0.04$ |
| $MLT$ | $+0.01 \pm 0.23$ | $+0.40 \pm 0.32$ | $+0.27$ |
| $B_z$ | $+0.08 \pm 0.14$ | $+0.17 \pm 0.19$ | $+0.22$ |
| $V_{sw}$ | $+0.06 \pm 0.11$ | $+0.19 \pm 0.24$ | $+0.21$ |
| $D_{st}$ | $+0.06 \pm 0.13$ | $+0.02 \pm 0.17$ | $+0.05$ |
| $\rho_{sw}$ | $+0.12 \pm 0.19$ | $+0.20 \pm 0.22$ | $+0.08$ |
| $F_{10.7}$ | $+0.51 \pm 0.06$ | $+0.48 \pm 0.25$ | $+0.45$ |
| $B_z + V_{sw}$ | $+0.12 \pm 0.10$ | $+0.15 \pm 0.21$ | $+0.29$ |
| $B_z + F_{10.7}$ | $+0.56 \pm 0.16$ | $+0.46 \pm 0.21$ | $+0.49$ |
| $D_{st} + F_{10.7}$ | $+0.54 \pm 0.07$ | $+0.47 \pm 0.15$ | $+0.46$ |
| $All$ | $+0.61 \pm 0.11$ | $+0.60 \pm 0.35$ | $+0.61$ |

Table of linear model test-set correlations showing the median of 100 random samples. Each sample trained on half of the data (via randomly selected rows of the least squares matrix) and tested on the other half.

- Nonlinear has more variation in samples, and between test and training sets, but does well for nonlinear variables such as DoY and MLT
- Can see that $F_{10.7}$ is the dominant source of correlation for all models
- Other variables are highly collinear
- Even using a model of all variables barely accounts for half of $\rho_{eq}$.
- Previous studies make more complex models or focus on fewer events; this dissertation attempts epoch analysis

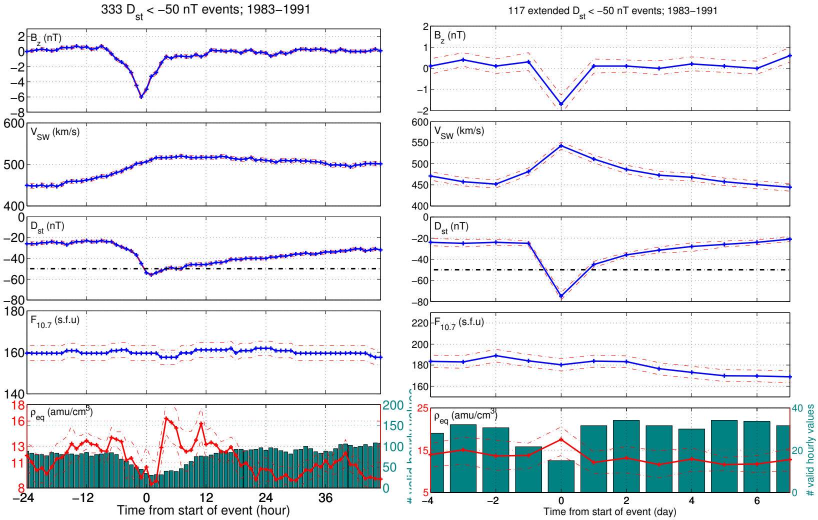
# Epoch

Epoch Analysis performed on two types of events:

- $\rho_{eq} > 20$ amu/cm$^3$
- $D_{st} <$ -50 nT

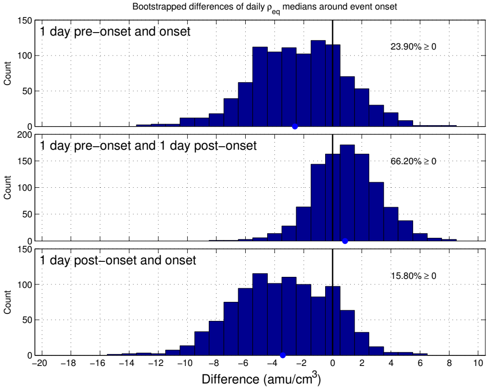
Threshold crossings for events considered at an hourly timescale.

Linear interpolation done on $\rho_{eq}$ to approximate event onsets when lacking data, but left as fill-values for analysis.

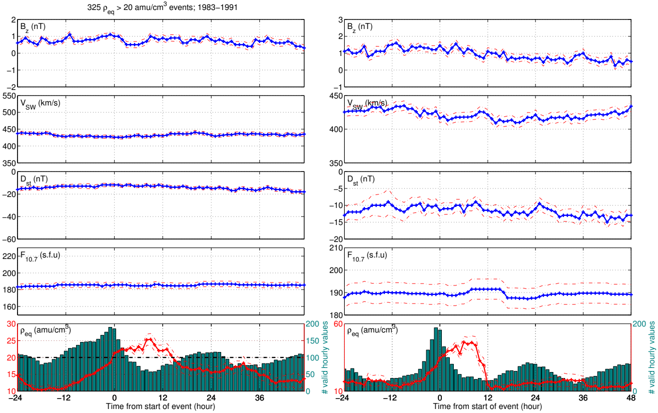



$D_{st}$ events on an hourly (left) and daily (right) timescale.

To verify whether events were significantly different between day of onset and the surrounding days, a bootstrap test of differences was performed:

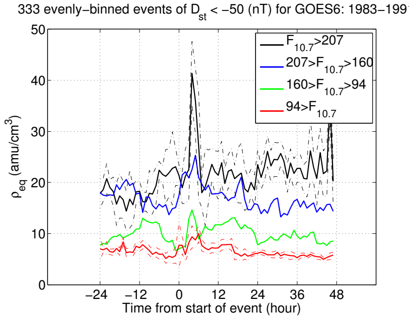


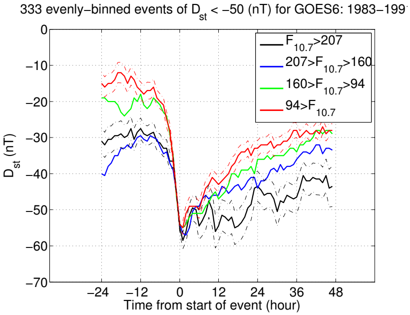

Bootstrap differences between median daily value of events using the years of 1983-1991.




Left: $\rho_{eq}$ events on hourly timescale. Right: $\rho_{eq}$ events where $\rho_{eq}$ increased by at least 5 amu during onset hour. Daily averages showed similar lack of significance and median northward $B_z$.

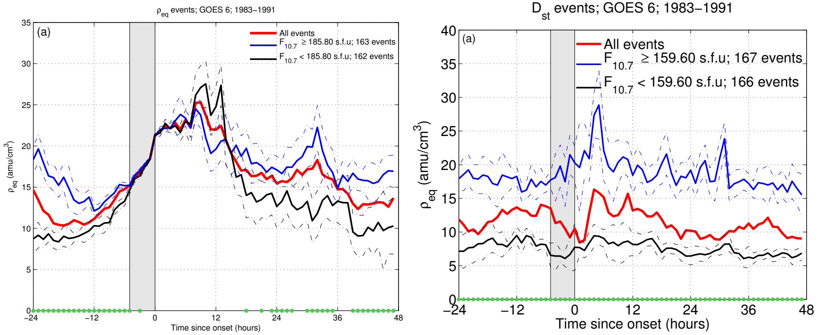
Want to investigate nonlinear dependencies by binning events:



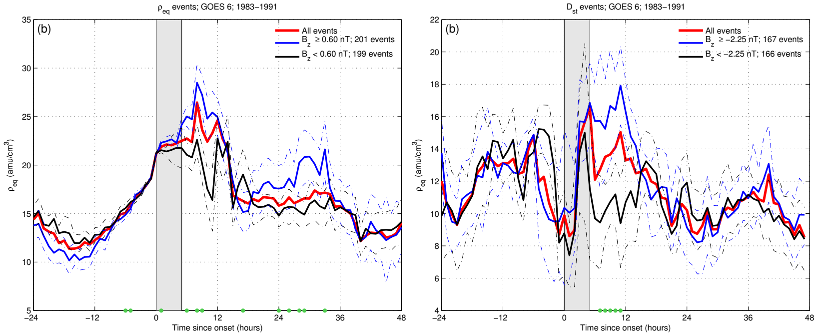

$\rho_{eq}$ (left) and $D_{st}$ (right) of $D_{st}$ events binned by median $F_{10.7}$ values.

Verifying that distribution of $\rho_{eq}$ is significantly different per $F_{10.7}$ bin before $\rho_{eq}$ or $D_{st}$ event onset.




$\rho_{eq}$ events (left) and $D_{st}$ events (right) binned by median $F_{10.7}$ before event onset.

Attempting to bin by IMF following event:




$\rho_{eq}$ events (left) and $D_{st}$ events (right) binned by median $B_z$ after event onset.

Can conclude:

- $\rho_{eq}$ decreases significantly between the day of a $D_{st}$ event and the day following
- Elevated $F_{10.7}$ coincides with elevated $\rho_{eq}$ before and after $\rho_{eq}$ events
- Elevated $F_{10.7}$ coincides with a short, multi-hour spike in $\rho_{eq}$ following a $D_{st}$ event
- $\rho_{eq}$ largely unaffected by north-south component of IMF around events.

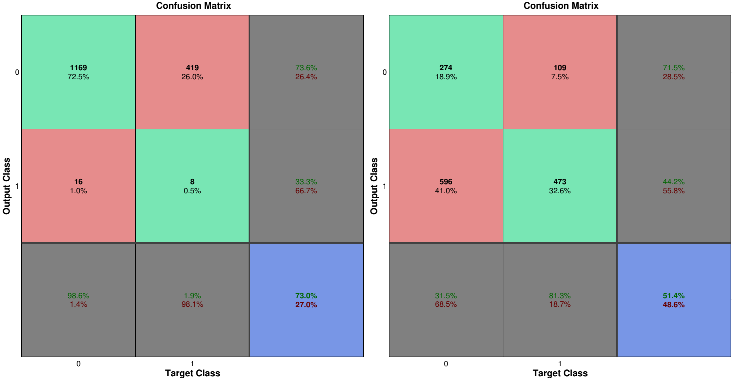
# Classification and Forecasting

Two main questions:

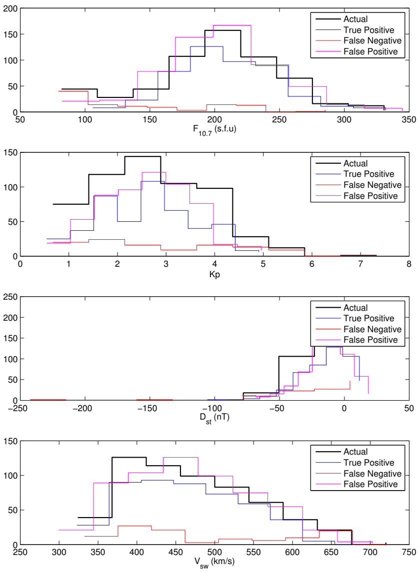
- ∼ Can event onsets be distinguished from the hours or days before onset?
- ∼ Can event onsets be forecasted using the hours or days before?

Utilizing MATLAB's `patternnet` neural network toolbox, classifying onsets as "1" and all other times as "0". Weighting "1"s more heavily based on timescale since events are much less frequent than non-events.

Classifying $\rho_{eq}$ events using $F_{10.7}$, Kp, $D_{st}$, and $V_{SW}$:

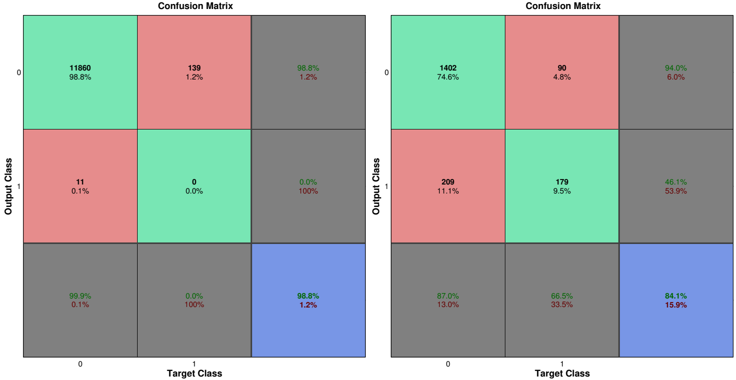



Classification confusion matrix for hourly (left) and daily (right) $\rho_{eq}$ onset.

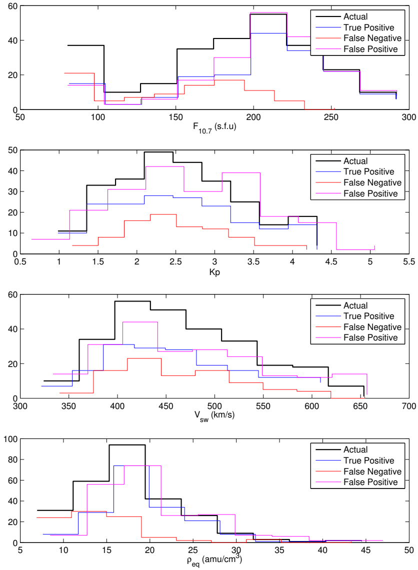



Histogram of onset conditions for daily-averaged $\rho_{eq}$ events binned by correctness of prediction. Appears that model failures are distributed equally to model successes.

Predicting $\rho_{eq}$ onsets using $F_{10.7}$, Kp, $D_{st}$, $V_{SW}$, and $\rho_{eq}$:




Prediction confusion matrix for hourly (left) and daily (right) $\rho_{eq}$ onset events.




Histogram of onset conditions for daily $\rho_{eq}$ events including $\rho_{eq}$ as an input, binned by correctness of prediction. Appears low $F_{10.7}$ leads to false negatives suggesting high $F_{10.7}$ is an indicator of a likely event.

# Conclusions

To reitterate the initial questions:

- What enhances $\rho_{eq}$ in the plasmatrough? Solar wind (via geomagnetic storms) or internal processes (e.g. ionospheric outflow)?

  **Geomagnetic storms not nearly as impactful as internal processes, such as outflow driven by $F_{10.7}$.**
- Does a $\rho_{eq}$ enhancement depend on current IMF conditions?

  **Weakly. Northward IMF supports higher $\rho_{eq}$ after either type of event.**
- Can a $\rho_{eq}$ enhancement be classified or forecasted?

  **Not well using most common solar wind and geomagnetic variables. Suggests something else needed.**

Further conclusions:

- ∼ Drop in $D_{st}$ does not significantly influence $\rho_{eq}$ on an hourly timescale, but does over the following day.
- ∼ $\rho_{eq}$ increases do not correlate, on average, with significant changes in $D_{st}$, $F_{10.7}$, or $B_z$.
- ∼ Classifying or forecasting the day of an event onset are possible but both prone to false positives or false negatives, depending on weighting.

# Future work

Many things could be done to improve:

- $\sim$ Exploring remaining signal after all collinear variables removed
- $\sim$ Adding model complexity and extra information
- $\sim$ Use new satellites and/or new Alfvn wave detection methods
- $\sim$ Extend time period of analysis

Thank you for your time.

Questions?