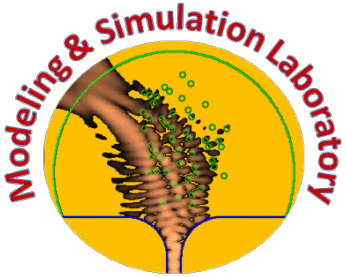
Machine-Learning Approach for Medium Range Forecast of Ionosphere Anomalies

Chunming Wang

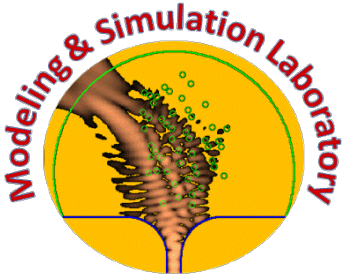
Modeling & Simulation Laboratory
Department of Mathematics
University of Southern California



Reliable Forecast is a Indispensable Demonstration of Understanding

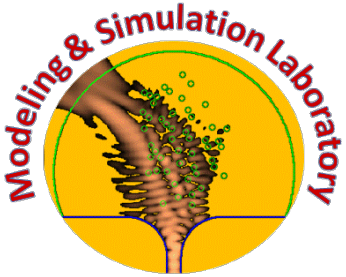
- Physics principle based models represent pinnacle of our understanding of the space weather system.
  - A highly coupled multiscale system involving solar, interplanetary magnetic field, plasma and ionosphere and thermosphere.
  - Extremely small portion of this complex system is observable.
- An effective forecasting system needs to be a combination of model and data driven system.

Development of Forecast Capability Can Help Model Improvement



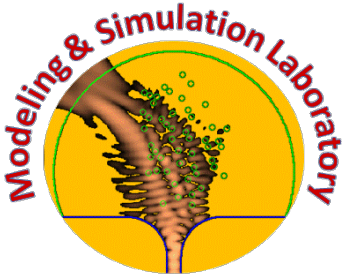
Learning to Forecast Space Weather with Existing Models and Data

- Capability for forecast can be developed with imperfect model and incomplete observation.
  - Measures of success in forecasting can provide insight for model improvements.
- Focus on forecast draws attention to a different set of issues than model development.
  - What, when and how to forecast, how to measure success?
  - Success or failure raise new challenging questions for model developers: why?



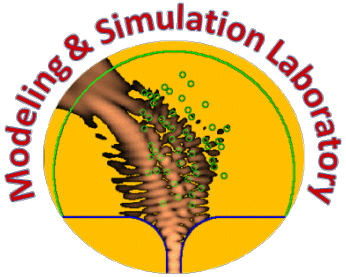
Machine-Learning Combines Statistics, Mathematics and Computer Sciences

- Space weather data have been collected over at least seven decades.
  - Compartmental used of space weather data may have lead to less full exploration of existing data.
- Machine-learning techniques provide general models for connecting different data components.
  - Regression analyses;
  - Random decision tree using uncertainty entropies;
  - Logistic regression and support vector methods.



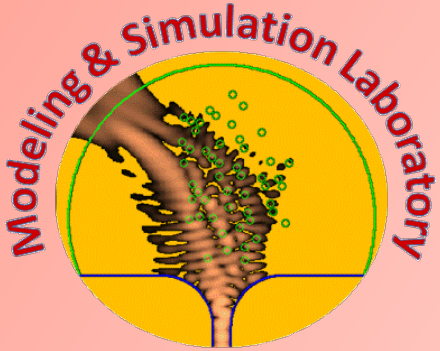
A Space Weather Forecast Test-bed Provides a Incubator for Innovation

- SWFT provides 3 key ingredients for training data driven forecast strategies:
  - A depository of quality controlled, easily accessible space weather data;
  - A collection of data handling, analyses and modeling tools to start with;
  - Computational and data management support at CCMC.
- Young space weather scientists and data science experts can easily collaborate in developing innovative forecast approaches.

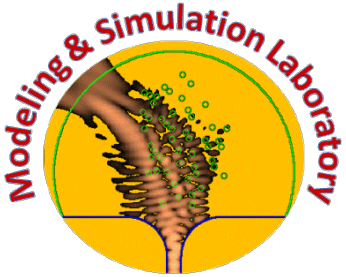


Initial Machine-Learning Based Forecasts Establish Bench-Mark Performance

- Preliminary results obtained in the LWS program for medium range forecast (1-3 days) of ionosphere anomalies showed promises.
  - Identified useful indicators for ionosphere anomalies through unsupervised machine-learning techniques;
  - Established reliable connection between space weather indices and ionosphere anomalies.
  - Preliminary forecast performance for various space weather parameters provides initial bench-mark for future forecast approaches.

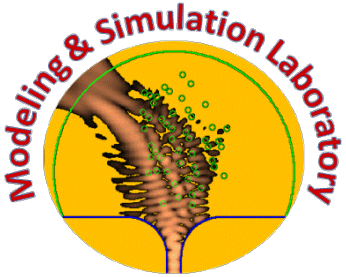


Forecasting Solar Wind



Forecasting Solar-Wind Using Past Observations

- Solar wind velocity is a key parameter used by thermosphere-ionosphere models.
- When these models are used to forecast conditions in the ionosphere, forecasted values of solar wind speed are required.
- Multi-linear regression method can be used to derive a forecast value for the solar wind by examining the correlation between past values of space weather observation to future one.



Multi-linear Regression Requires Computation of Covariance

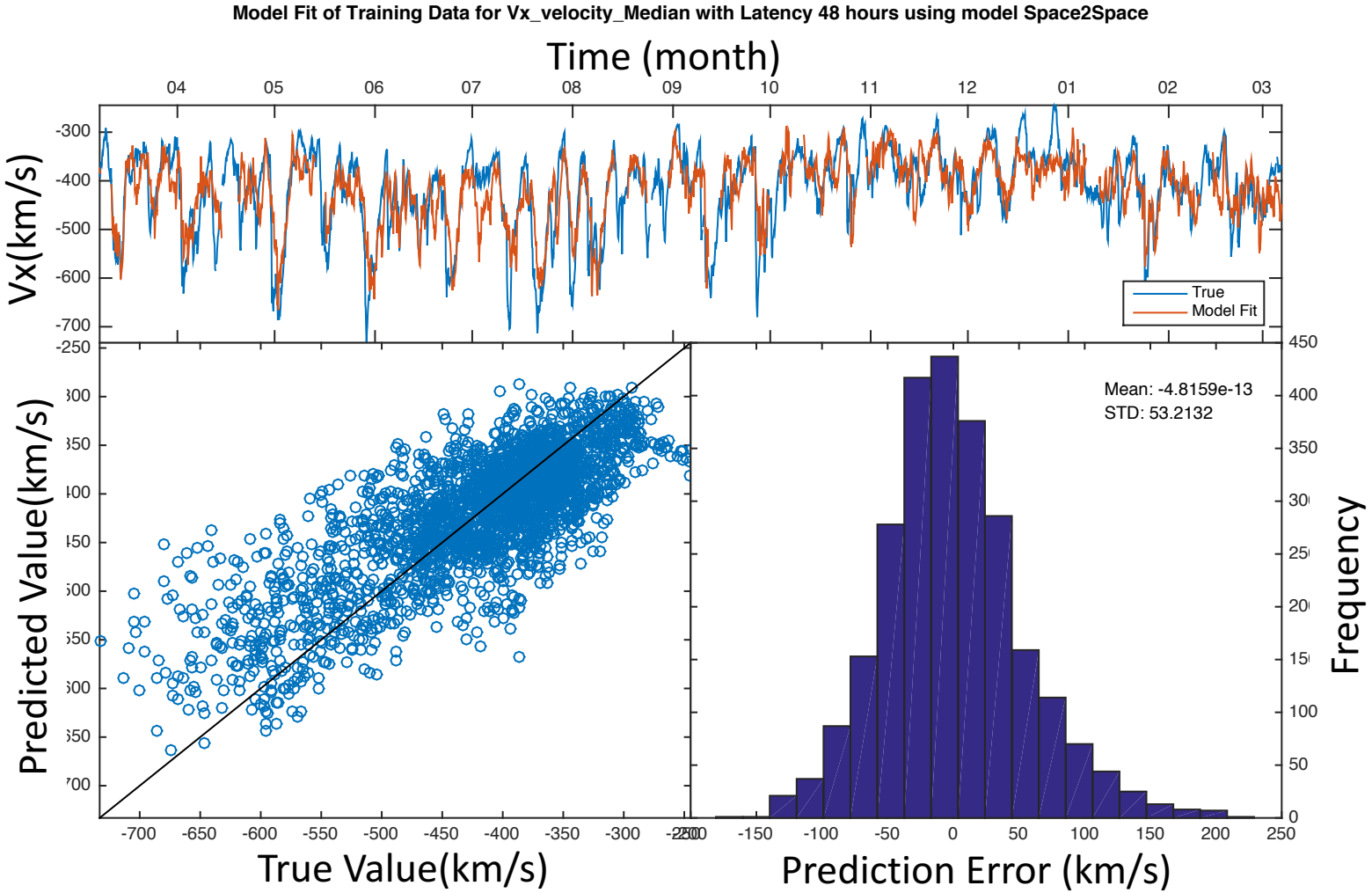
- Regression consists of finding optimal coefficiets for a linear model of the form

$$v_X(t) = T\begin{pmatrix} \vec{u}(t-d) \\ \vdots \\ \vec{u}(t-d-m) \end{pmatrix}, \vec{u}(t) = \begin{pmatrix} v_X(t) \\ \vdots \\ B_X(t) \\ \vdots \\ n_{SunSPot}(t) \end{pmatrix}.$$

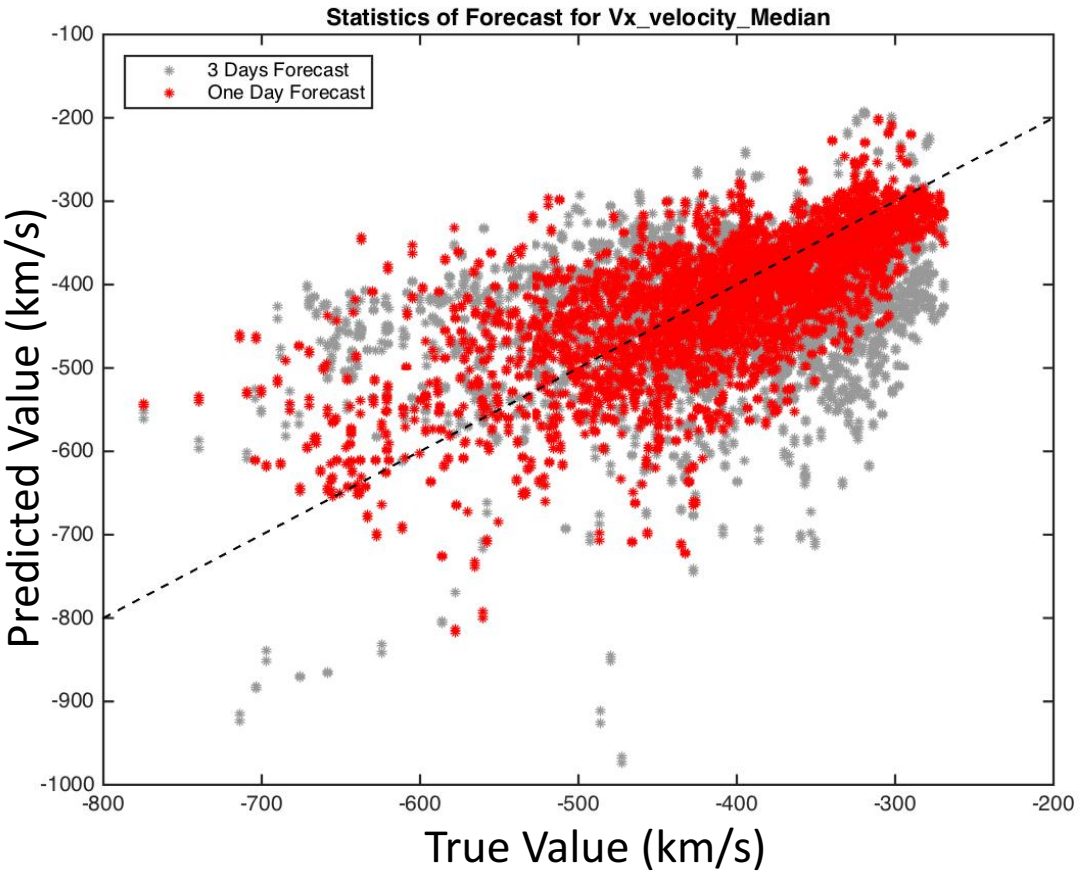
- Optimization for T consists of solving the least square minimization problem using historic data

$$\min_T \sum_{k=1}^{N} \left| v_X(k) - T\begin{pmatrix} \vec{u}(k-d) \\ \vdots \\ \vec{u}(k-d-m) \end{pmatrix} \right|^2 .$$

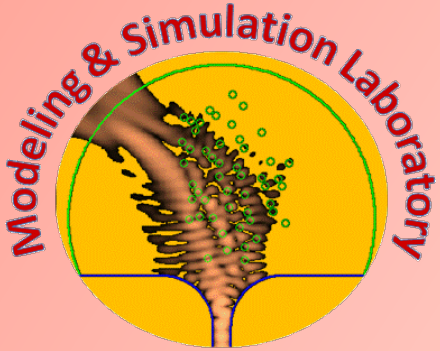
Data for Model Training, Verification and Validation Can be Reviewed



User Defined New Forecast Model Can be Tested on SWFT

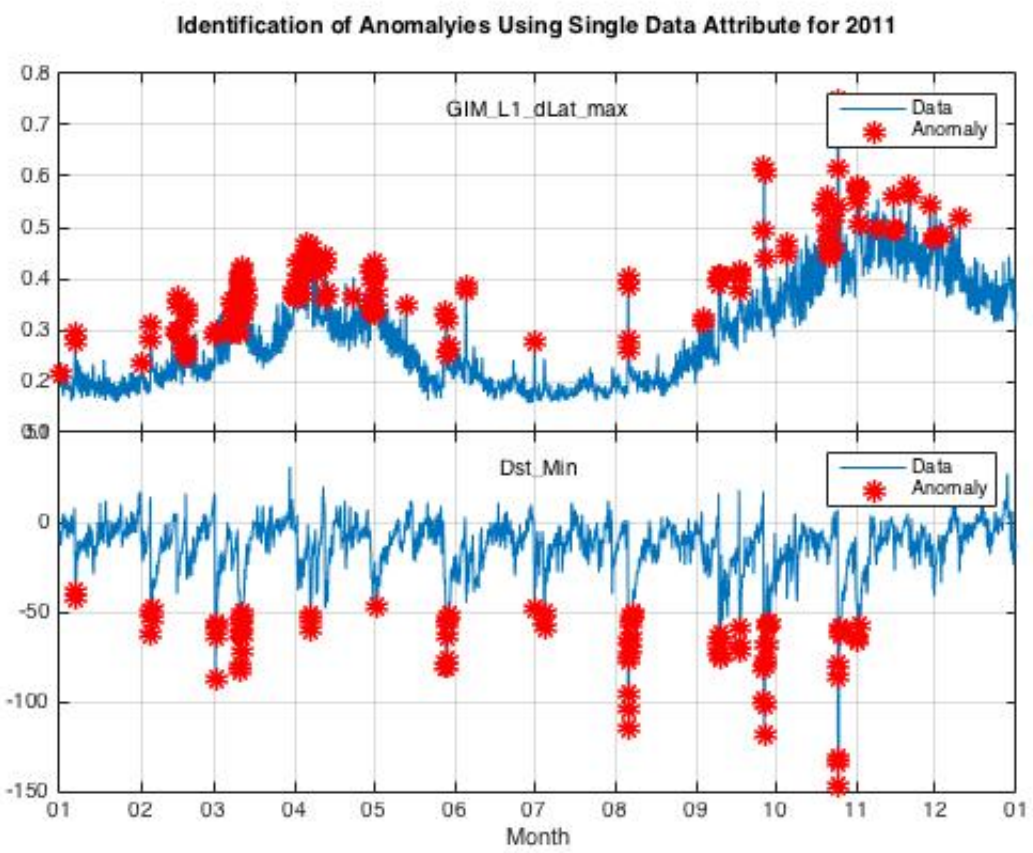


- SWFT script can be used to evaluate a new forecast strategy using extended time periods for evaluation.
- Scoring of the new forecast strategy can be made using repositories of model output.



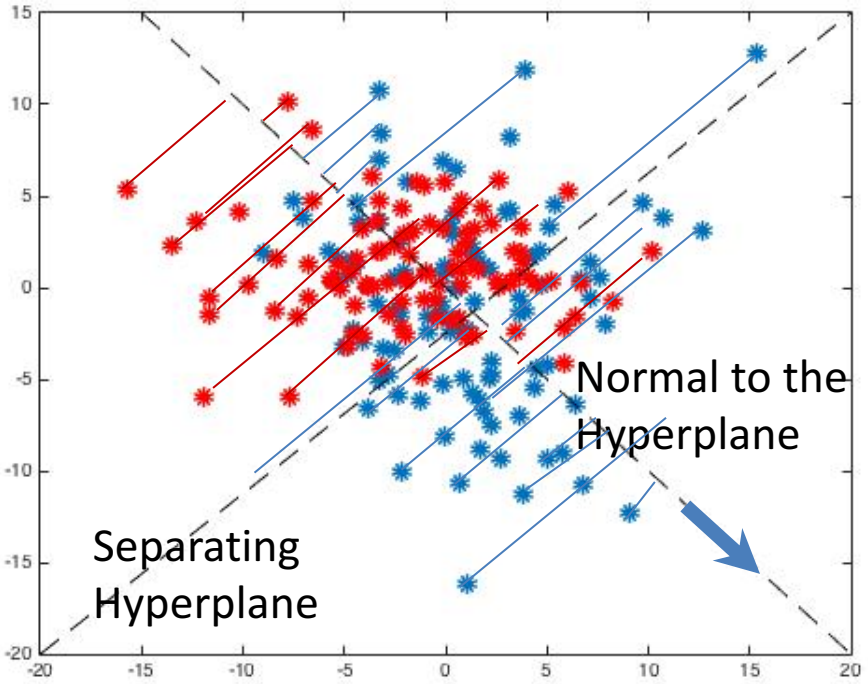
Forecasting Ionosphere Anomalies

Anomalies Can be Defined Using a Single or Multiple Metrics



- Instead of forecasting "observations", *anomalies* can be defined and forecast.
- Machine learning algorithms exist for making categorical forecasts.

Logistic Regression Algorithms Helps Classify Data



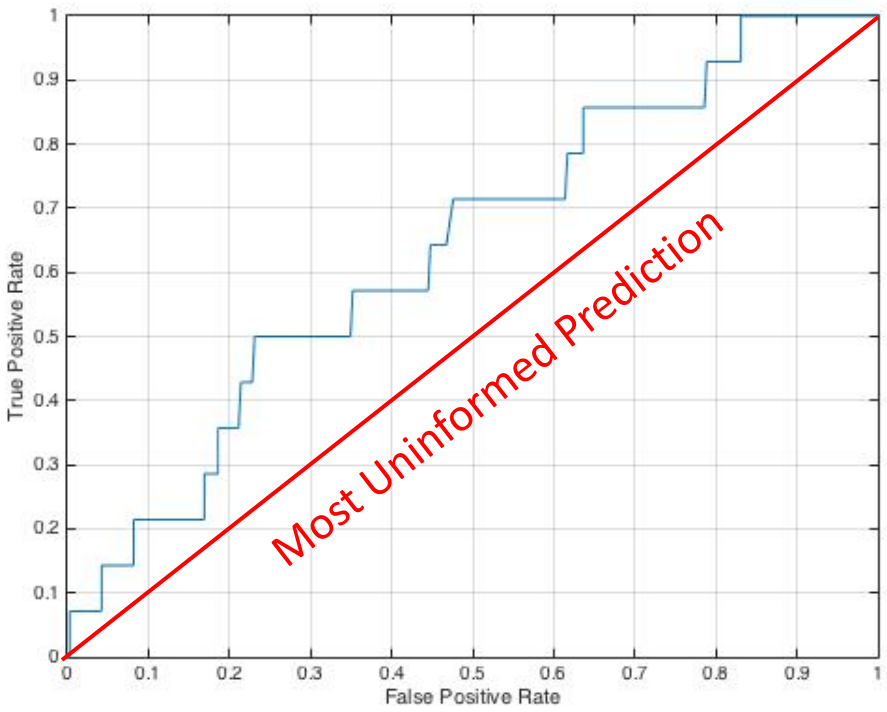
Data Score

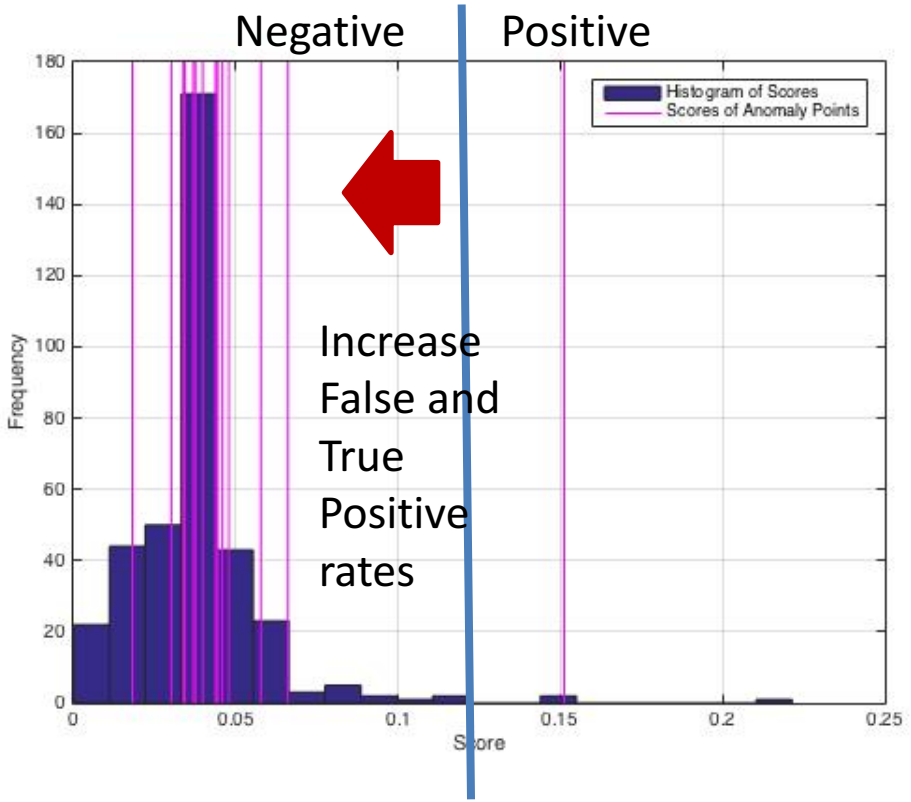
$$p(\vec{x}) = \frac{e^{\vec{a}^T\vec{x}+b}}{1+e^{\vec{a}^T\vec{x}+b}}$$

- Seeking hyperplane in multiple-dimensional space of forecast variables to best separate two categories of data.
- Find vector $\vec{a}$ and offset b to maximize the likelihood function

$$\prod_{k=1}^{N}\left(p(\vec{x}_k)\right)^{y_k}\left(1-p(\vec{x}_k)\right)^{1-y_k}$$

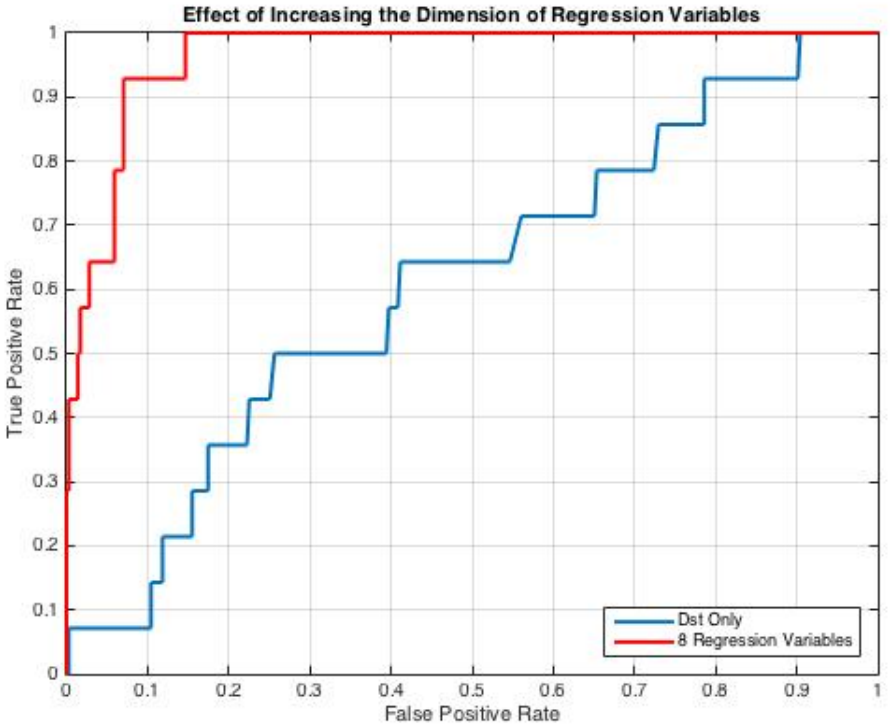
Success in Regression Training is Measured by True/False Positive Rates

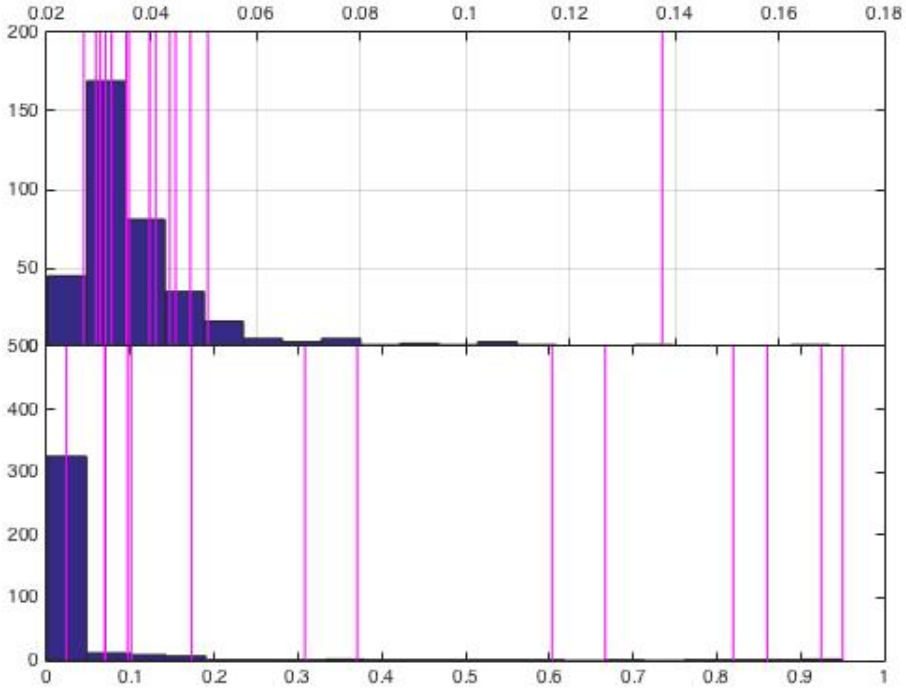




- ROC curve represents the trade-off between true and false positive rates.

Increasing the Dimension of Regression Variables Improves Fit





- Adding regression variables in general improves classification of training data but not necessarily forecast.

Conclusion

- Machine learning or data sciences in general have the potential to bring revolutionary changes in scientific research.
- Forecasting space weather with current imperfect models and data can benefit from machine learning techniques.
- Focusing on forecast requires us to addressing challenging issues in a collaborative way.