Impact of Analytics and Meta-learning on Estimating Geomagnetic Storms

A Two-stage Framework for Prediction

Taylor K. Larkin[1] and Denise J. McManus[2]
Information Systems, Statistics, and Management Science
Culverhouse College of Commerce
The University of Alabama
Tuscaloosa, AL 35487-0226
Email: tklarkin@crimson.ua.edu[1], dmcmanus@cba.ua.edu[2]

***Abstract*—Cataclysmic damage to telecommunication infrastructures, from power grids to satellites, is a global concern. Natural disasters, such as hurricanes, tsunamis, floods, mud slides, and tornadoes have impacted telecommunication services while costing millions of dollars in damages and loss of business. Geomagnetic storms, specifically coronal mass ejections, have the same risk of imposing catastrophic devastation as other natural disasters. With increases in data availability, accurate predictions can be made using sophisticated ensemble modeling schemes. In this work, one such scheme, referred to as stacked generalization, is used to predict a geomagnetic storm index value associated with 2,811 coronal mass ejection events that occurred between 1996 and 2014. To increase lead time, two rounds (stages) of stacked generalization using data relevant to a coronal mass ejection's life span are executed. Results show that for this dataset, stacked generalization performs significantly better than using a single model in both stages for the most important error metrics. In addition, overall variable importance scores for each predictor variable can be calculated from this ensemble strategy. Utilizing these importance scores can help aid telecommunication researchers in studying the significant drivers of geomagnetic storms while also maintaining predictive accuracy.**

Keywords–ensemble modeling; space weather; quantile regression; stacked generalization; telecommunications.

I. INTRODUCTION

Predicting geomagnetic storms is an ever-present problem in today's society, given the increased emphasis on advanced technologies [1]. These storms are fueled by coronal mass ejections (CMEs), which are colossal bursts of magnetic field and plasma from the Sun as displayed in Figure 1. Typically, a CME travels at speeds between 400 and 1,000 kilometers per second [2] resulting in an arrival time of approximately one to four days [3]; however, they can move as slowly as 100 kilometers per second or as quickly as 3,000 kilometers per second (or around 6.7 million miles per hour) [4]. These phenomena can contain a mass of solar material exceeding 10^{13} kilograms (or approximately 22 trillion pounds) [5] and can explode with the force of a billion hydrogen bombs [6]. Naturally, CME events are often associated with solar activity such as sunspots [4]. During the solar minimum of the 11 year solar cycle (the period of time where the Sun has fewer sunspots and, hence, weaker magnetic fields), CME events occur about once a day. During a solar maximum, this daily estimate increases to four or five. One plausible theory for these incidents taking place involves the Sun needing to release energy. As more sunspots develop, more coronal magnetic field structures become entangled; therefore, more energy is required to control the volatility and convolution. Once the energy surpasses a certain level, it becomes beneficial for the Sun to release these complex magnetic structures [2].

When this force approaches Earth, it collides with the magnetosphere. The magnetosphere is the area encompassing Earth's magnetic field and serves as the line of defense against solar winds. The National Oceanic and Atmospheric Administration (NOAA) describes this event as "the appearance of water flowing around a rock in a stream" [7] as shown in Figure 2.

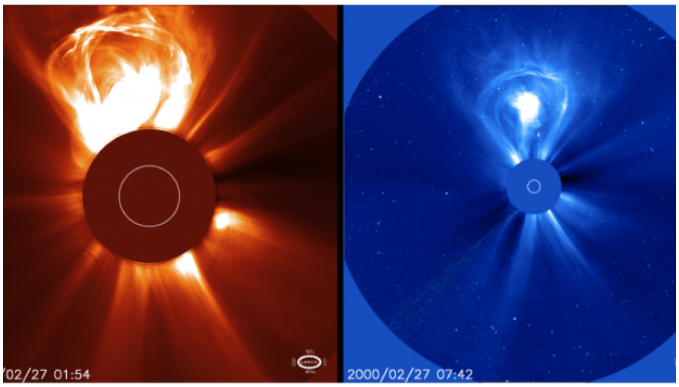


Figure 1. LASCO coronagraph images [4], courtesy of the NASA/ESA SOHO mission.

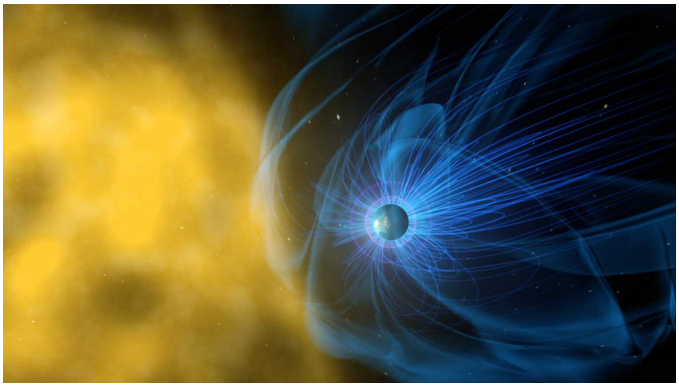


Figure 2. Rendering of Earth's magnetosphere interacting with the solar wind from the Sun [8], courtesy of the NASA.

After the solar winds compress Earth's magnetic field on the day side (the side facing the Sun), they travel along the elongated magnetosphere into Earth's dark side (the side

opposite of the Sun). The electrons are accelerated and energized in the tails of the magnetosphere, filtering down to the Polar Regions and clashing with atmospheric gases causing geomagnetic storms. This energy transfer emits the brilliance known as the *Aurora Borealis*, or Northern Lights, and the *Aurora Australis*, or Southern Lights, which can be seen near the respective poles.

While mainly responsible for the illustrious Northern Lights, geomagnetic storms have the potential to cause cataclysmic damage to Earth. Normally, the magnetic field is able to deflect most of the incoming plasma particles from the Sun. However, when a CME contains a strong southward-directed magnetic field component (B_z), energy is transferred from the CME's magnetic field to Earth's through a process called magnetic reconnection [9][10][11] (as cited in [12]). Magnetic reconnection leads to an injection of plasma particles in Earth's geomagnetic field and a reduction of the magnetosphere towards the equator [2]. Consequently, more energy is amassed in the upper atmosphere, particularly at the poles. Moreover, this energy is impressed upon power transformers causing an acute over-saturation and inducing black-outs via geomagnetically induced currents (GICs) [13]. Some other residuals of this over-accumulation of energy include the corrosion of pipelines, deteriorations of radio and GPS communications, radiation hazards in higher latitudes, damages to spacecrafts, and deficiencies in solar arrays [14]. These ramifications pose a significant threat to global telecommunications and electrical power infrastructures as CMEs continue to be launched towards Earth [15] and remain the primary source of major geomagnetic disturbances [16][17][18] (as cited in [19]). From a business perspective, risk factor mitigation is an absolute necessity within the global business environment [20]. This can be accomplished using advanced analytical techniques on data collected about these phenomena.

The subsequent sections of this work read as follows. Section II introduces previous studies on predicting geomagnetic storms. Section III provides detail about the basics of the methodology used, the dataset studied, and the experimental strategy. Section IV displays and discusses the results as well as postulates areas for future work. Section V concludes with a summary.

II. Literature Review

A. Predicting Dangerous CMEs

CMEs present an ever-increasing threat to Earth as society becomes more dependent on technology, such as satellites and telecommunication operations. Nevertheless, because of this increase in technology, more data has been collected about these acts and the solar wind condition in general. This, in turn, has allowed for empirical models to be developed. Burton, McPherron, and Russell [21] presented an algorithm to predict the disturbance storm time index (DST) value [22] based on solar wind and interplanetary magnetic field parameters. The DST value is a popular metric to assess geomagnetic activity. Expressed in nanoteslas (nT) and recorded every hour from observatories around the world, it measures the depression of the equatorial geomagnetic field, or horizontal component of the magnetic field; thus, the smaller the value of the DST, the more significant the disturbance of the magnetic field [2]. Many researchers have used this information for building forecasting models to predict geomagnetic storms [23][24].

However, many of these systems only use *in-situ* data, or data that can only be measured close to Earth. To improve prediction, studies have included data gathered at the onset of a CME and the near-Earth interplanetary information (IPI) regarding the solar wind condition as the CME approaches Earth [25][26][27]. These have ranged from using logistic regression [26] to neural networks [28] to make predictions based on this combination of data. Further improvements have been made by using multi-step frameworks. To narrow the scope, this work will focus on reviewing two recent two-step procedures that predict geomagnetic storms using both near-Earth IPI and CME properties taken near the Sun.

Valach, Bochníček, Hejda, and Revallo [29] reinforced one of the primary issues facing geomagnetic storm prediction: the inability to estimate the orientation of the interplanetary magnetic field from an incoming CME more than a few hours out. It is well-known that one of the largest predictor variables is the magnitude of the aforementioned magnetic field component B_z [21][26][2]; however, this is difficult to predict prior to reaching the L1 Lagrangian point (the position close to Earth where much of the IPI is collected) due to complexities in a CME's magnetic topology [30]. Hence, under the assumption that the direction of the magnetic field component is unpredictable, the authors first study the behavior of B_z for 2,882 days between 1997 to 2007 before implementing any predictive construct. Based on their analysis, they determined that for the majority of the days with a high-level of geomagnetic activity, B_z was negative for at least 16 hours during the course of the day (behavior exhibited by roughly 31% of the days studied). Then, after building a neural network using these observations, they forecasted the daily level of geomagnetic activity with initial CME and solar X-ray information. The benefits to their approach are that the predictions are timely (absence of IPI in the second step enable forecasts at least a day out) and are well-suited for the strongest of storms (since the training observations are composed of days where B_z is negative for more than 16 hours). However, as noted by the authors, it does not do as well differentiating moderate and weak geomagnetic storms. In addition, the time scale of the prediction is in days, which is not as granular as hours.

Kim, Moon, Gopalswamy, Park, and Kim [27] argued that only using information based on urgent warning IPI for prediction does not provide a practical lead time for preparations to be made on Earth, even though the forecasts are more accurate. At the same time, strictly employing initial CME data becomes frivolous as each CME experiences changes in composition as they propagate through the interplanetary medium, thereby, making prediction difficult. Therefore, the authors constructed a two-step forecasting system using both urgent warning IPI and initial CME data. At the first stage, they applied multiple linear regression models to predict the strength of geomagnetic activity for northward and southward events at the onset of a CME using its location, speed, and direction parameter (estimated from the magnetic orientation angle of the related active region on the Sun). The estimation of the direction (north or south) is based on the assumption that these rarely deviate from that of the associated active region [31]. Next, they administered a set of rules based on the IPI to update the forecast and classify the impending CME as causing a moderate or intense storm. This method contributes

a medium-term to short-term forecast from the first observance of a CME to its approach to Earth. While this method yields accurate and interpretable results, only 55 CMEs from 1997-2003 were studied. Moreover, the absence of using a validation scheme when creating the rules can lead to over-fitting when predicting on future data [32].

Interestingly enough, the former work assumes the direction of the magnetic field component in a CME is unpredictable while the latter estimates this in their step one models. In this work, the direction is not considered in any of the steps. Instead, the dataset captures values of B_z prior to the climax of a given geomagnetic storm [33]. Thus, if this value is high in magnitude, then this reflects the southward behavior.

Aside from the work by Dryer et al. [25], which used an ensemble of four physics-based models to predict shock arrival times, the idea of using ensembles of models has not been very prevalent in the literature. Stacked generalization [34] is a type of ensemble that uses the individual predictions from a set of base models as inputs for another model to make a final prediction. This strategy has been the backbone of successful schemes in areas such as predicting financial fraud [35], bankruptcy [36], and user ratings in the famous Netflix Prize competition [37]. Therefore, leveraging more advanced ensemble frameworks for predictive modeling has the opportunity to increase accuracy in this field.

B. Stacked Generalization

The idea of stacked generalization can be simplified in the following way:

- Construct a dataset consisting of predictions from a set of level 0 (or base) learners using a training and a test set. Refer to this as the metadata.
- Generate a level 1 (or meta) learner that utilizes the predictions made at the previous level as inputs. That is, train the meta-learner on the metadata as opposed to the original training data.

Often times, the predictions from the base-learners are determined via k-fold cross-validation [38]. Define the dataset $S = \{(y_i, \boldsymbol{x}_i), i = 1, ..., n\}$ where $\boldsymbol{x}_i$ is a vector of predictor variables and y_i is the corresponding response value for the i^{th} observation. Specifically, split the dataset S into k near equal and disjoint sets such that $S_1, S_2, ..., S_k$. Let $S^{-k} = S - S_k$ and S_k be the training and test sets, respectively. Execute the base-learner on the first S^{-k} parts and produce a prediction for the held-out part S_k. Repeat this procedure until each subset of S has been used as a test set exactly once. Extract all the hold-out predictions to create the metadata. Because generating the metadata is an independent process across each base-learner, it can be parallelized for faster computation. That is, each base-learner can be trained at the same time. This is key as time plays a pivotal role in geomagnetic storm prediction [27].

The meta-learner's purpose is to gain information about the generalization behavior of each learner trained at the base-level. Popular choices for meta-learners have been linear models [39]. While this ensemble strategy leverages the strengths and weaknesses of the base-learners, it can be prone to over-fitting [40]. Therefore, in order to combat this issue, employing regularized linear methods can perform better than their non-regularized counterparts [41][38][42]. Reid and Grudic [42] experimented with three regularization penalties: ridge [43], lasso [44], and the elastic net [45]. The authors showed that imposing these penalties perform well on multi-class datasets. They commented that using the lasso and elastic net penalties can promote sparse solutions that can reduce the size of the ensemble at the meta-level. Pruning the size of an ensemble model has been explored in other works [46][47][48]. It can lead to better generalization and promote the necessary diversity in the base-learner predictions [34].

Based upon the results in previous studies, it seems advantageous to implement a regularized meta-learner to have the best potential for success in stacked generalization. By using various types of penalty functions, a learning system can effectively make predictions and provide sparse solutions, even in situations with severe multicollinearity since all of the base-learners are trying to predict the same outcome [38]. However, none of the studies mentioned above discuss how to choose a meta-learner when the outlier values are important for regression tasks. Specifically for predicting geomagnetic storms, outliers are important because strong CMEs do not occur often; hence, a meta-learner cannot downplay the effect of these for prediction. If anything, the meta-learner should treat these values with more emphasis. In addition, subsetting the data to only include these outliers for model construction inhibits meta-knowledge to be gained for all CME events. Therefore, for this study, a regularized quantile regression model is chosen for the meta-learner in order to more adequately deal with outliers, improve accuracy, and promote sparse solutions.

C. Regularized Quantile Regression

Recall the ordinary least squares (OLS) solution for the coefficients in linear regression:

$$\hat{\beta}^{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{1}$$

where $\boldsymbol{X}$ is the predictor matrix of dimension $n \times (p+1)$ and $\boldsymbol{Y}$ is the vector of outcomes of dimension $n \times 1$ for n observations and p predictor variables. Specifically,

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Alternatively, Eq. (1) can be written as the following optimization problem:

$$\underset{\boldsymbol{\beta}}{\text{argmin}} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \tag{2}$$

To apply regularization to the estimated coefficients, a penalty term can be added such that [49]

$$\underset{\boldsymbol{\beta}}{\text{argmin}} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \tag{3}$$

where $p_\lambda(\cdot)$ dictates the type of penalty function with a non-negative constant λ to determine the amount of regularization. Utilizing constrained regression approaches enables the ability to perform variable selection or improve prediction in particular environments. However, the main goal in these methods is to estimate the conditional mean of some response given a set of predictor variables. Situations may arise where it is more advantageous to investigate a certain part of the

conditional distribution [50][51]; hence, quantile regression was developed [52]. The goal of quantile regression is to offer "a comprehensive strategy for completing the regression picture" (pg. 20) [53]. In general, this involves minimizing the sum of asymmetrically weighted absolute residuals [52]

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \frac{1}{n}\Bigg[\sum_{i\in\{i:y_i\geq \boldsymbol{x}_i'\boldsymbol{\beta}\}} \tau|y_i-\boldsymbol{x}_i'\boldsymbol{\beta}| + \sum_{i\in\{i:y_i< \boldsymbol{x}_i'\boldsymbol{\beta}\}} (1-\tau)|y_i-\boldsymbol{x}_i'\boldsymbol{\beta}|\Bigg] \tag{4}$$

for some given quantile level τ. In this way, different weights are placed on positive (under-prediction) and negative (over-prediction) errors corresponding to the desired quantile. Note that when $\tau = 0.5$, this simply reduces to median regression. As with the linear case, the coefficients in quantile regression can be penalized the same way. Using lasso has been a popular choice due to its sparse nature [54][55]. However, it has been shown that lasso has some limitations in high-dimensional situations or ones with severe multicollinearity [45]. In addition, it lacks oracle properties [56][57]. That is, lasso does not select the correct subset of predictor variables while also efficiently estimating the non-zero coefficients as if only the truly influential predictor variables are included in the model, asymptotically [58]. Thus other penalties, such as the smoothly clipped absolute deviation (SCAD) [56], have been developed. This has been shown to retain oracle properties for penalized quantile regression models [59]. It can be defined as a quadratic spline function with knots at λ and $a\lambda$ to make the following objective function:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \frac{1}{n}\Bigg[\sum_{i\in\{i:y_i\geq \boldsymbol{x}_i'\boldsymbol{\beta}\}} \tau|y_i-\boldsymbol{x}_i'\boldsymbol{\beta}| + \sum_{i\in\{i:y_i< \boldsymbol{x}_i'\boldsymbol{\beta}\}} (1-\tau)|y_i-\boldsymbol{x}_i'\boldsymbol{\beta}|\Bigg] + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \tag{5}$$

where

$$p_\lambda(|\beta|) = \begin{cases} \lambda|\beta| & 0\leq|\beta|<\lambda \\ \dfrac{a\lambda|\beta|-(\beta^2+\lambda^2)/2}{a-1} & \lambda\leq|\beta|\leq a\lambda \\ \dfrac{(a+1)\lambda^2}{2} & |\beta|>a\lambda \end{cases}$$

for some $a > 2$ and $\lambda > 0$. By assigning different weights depending on $|\beta|$, SCAD avoids over-penalizing large coefficients, as is a common problem in lasso [56][59][49]. Traditionally, solving Eq. (5) is difficult due to its non-convex nature. Fortunately, efficient algorithms have been developed to increase the computational speed for solving these non-differentiable and non-convex optimization problems [49]. Because of the advantages of using the SCAD penalty, this work employs this type of regularization on a quantile regression model at the meta-level. Note that subsequent uses of SCAD refer to the quantile regression model in Eq. (5).

D. A Two-stage Approach

Given the success of multi-step approaches, this work executes two rounds of stacked generalization using two data sources:

1) Initial CME properties taken at the time of ejection
2) Initial CME properties taken at the time of ejection plus the IPI

The execution of stacked generalization on the first data source, noted as stage one, can provide a preliminary estimate as to how strong a CME will be. Then, after adding the important IPI in stage two, the forecast can be updated to more accurately reflect the potential danger from the respective CME. This two-stage meta-learning approach seeks to emulate Kim et al.'s [27] medium-term to short-term forecast for predicting geomagnetic storms. To increase in the interpretation of the framework, the variable importance strategy for stacked generalization described by Larkin [33] is instituted. This involves calculating model-specific variable importance scores for each base-learner and then weighting these scores based on the coefficients from SCAD to produce a final aggregated variable importance score for each predictor variable.

III. Methodology

A. Data

Four sources are considered to construct the experimental dataset: near-Earth CME information provided by Richardson and Cane [60][61], OMNI hourly averaged solar wind data at one AU (astronomical unit) from the Coordinated Data Analysis (Workshop) Web [62], CME measurements given by the Large Angle and Spectrometric Coronagraph (LASCO) located on the Solar and Heliospheric Observatory (SOHO) satellite [63], and some Sun characteristics recorded by NOAA [64]. These data are combined so that each CME has been assigned IPI values (such as B_z) prior to the DST minimum during a predicted area of effect on Earth. Establishing these values before the DST minimum gives a lead time prior to the climax of the geomagnetic storms and allows for a more realistic prediction scenario, especially since B_z typically minimizes prior to the minimization of the DST value [65]. Also included are the initial measurements about the speed and angle of a CME at the time of ejection from the Sun and daily Sun characteristics on the day of ejection. After filtering out missing values and some unnecessary rows, a dataset composed of 2,811 CME events from 1996 to 2014 with 28 predictor variables is ready for analysis [33]. Note only 16 of the 28 predictor variables will be be used in the first stage. Approximately 5% of the observations in the dataset are deemed as strongly geoeffective (i.e. produce a geomagnetic storm with a DST ≤ -100nT). The list of predictor variables is divided into initial CME and solar characteristics in Table I and the subsequent IPI in Table II. Predictor variables types are denoted as continuous (C), discrete (D), or binary (B).

B. Experimental Set-up

The analysis is performed in the R environment version 3.3.2 [66] using the **caret** (Classification And REgression Training) package [67]. This package allows for a streamlined user interface for applying a diverse set of predictive models from different packages with options to perform various pre-processing, post-processing, resampling, and visualization

TABLE I. LIST OF INITIAL CME PROPERTIES AND SUN CHARACTERISTICS

Variable	Type	Description
MPA	C	Measurement position angle of CME at the height-time measurements (degrees)
AW	C	Sky-plane width of CME (degrees)
LS	C	Linear speed of CME (km/s)
SOI	C	Quadratic speed of CME at initial height measurement (km/s)
SOF	C	Quadratic speed of CME at final height measurement (km/s)
SOR	C	Quadratic speed of CME at height of 20 solar radii (km/s)
Acc	C	Acceleration of CME in (m/s^2)
$Poor$	B	Noted as a poor event in the comments
Very_Poor	B	Noted as a very poor event in the comments
$RFlux$	C	Daily average 10.7cm flux values of solar radio emissions on CME ejection day in 10^{-22}J/s/m^2/Hz
SSN	D	Number of sunspots recorded on CME ejection day
SSA	C	Sum of the corrected area of all observed sunspots on CME ejection day in millionths of the solar hemisphere
NR	D	Number of new sunspot regions on CME ejection day
$XrayC$	D	Number of C-class solar flares on CME ejection day
$XrayM$	D	Number of M-class solar flares on CME ejection day
$XrayX$	D	Number of X-class solar flares on CME ejection day

TABLE II. LIST OF IPI

Variable	Type	Description
E_y	C	Interplanetary electric field in millivolts per meter (mV/m)
B_x	C	X-component magnetic field component (nT)
B_y	C	Y-component magnetic field component (nT)
B_z	C	Southward magnetic field component (nT)
V_{sw}	C	Plasma flow speed (km/s)
Phi	C	Plasma flow direction longitude (degrees)
$Theta$	C	Plasma flow direction latitude (degrees)
D_p	C	Proton density in Newtons per cubic centimeter (N/cm^3)
Na_Np	C	Alpha to proton ratio
T_p	C	Proton temperature in degrees Kelvin (K)
P	C	Flow pressure in nanopascals (nPa)
$Beta$	C	Plasma beta

techniques. In addition, for those models that can perform variable importance estimation, the **caret** package can automatically extract these measures for a practitioner's use. Due to the large number of models available, a rich series of machine learning algorithms and statistical models may be realized to construct the foundation of base-learners. Care is taken to ensure a diverse collection of 50 models and algorithms is used [46]. Unfortunately, in an effort to include a larger number of base-learners, not every model is able to provide model-specific variable importance scores. That is, they either do not have a way to calculate variable importance or **caret** does not implement one. For this study, less than half (42%) of base-learners have model-specific importance scores. For those that do not, the R^2 statistic is calculated using a loess smoother which is fit between the outcome and each predictor variable, as done by default within the package [68]. A summary of the 50 chosen base-learners is listed in Table III. Asterisks "*" indicate those methods that can provide model-specific variable importance scores.

Another advantage to using **caret** is the option to easily tune the parameters for a given learner by simply specifying a number for $tuneLength$ in the $train$ function. Each model has a predefined range of tuning values to search over proportional the $tuneLength$. The higher the $tuneLength$, the more tuning executed. The number of tuning parameters range for each model. In this experimental set-up, $tuneLength$ is left at the default value of three.

For the SCAD implementation, the **rqPen** R package is chosen [69]. This package offers estimation for SCAD as well as other penalized quantile regression models including lasso. In addition, it can utilize the recently proposed and efficient iterative coordinate descent algorithm [49] to compute SCAD solutions using the *QICD* function. Because this function is not offered in **caret**, it is incorporated within the **caret** framework by creating a custom model. It is important to implement this within **caret** to be sure SCAD is trained across the same folds as the base-learners for a fair comparison. To tune SCAD, only two parameters are adjusted: the regularization value λ and the quantile level τ. The parameter a in Eq. (5) is left at the suggested default value of 3.7. The value of λ controls how much to penalize the coefficients and works similarly as λ in the popular **glmnet** package [70]. A diverse range of values are investigated: $\lambda = \{1000, 1, 0.001\}$. For many applications of quantile regression, the selection of the quantile level τ is determined by the user to best suit the research goal. In this work, τ is treated as a tuning parameter to best find a balanced between accurately predicting the much rarer dangerous geomagnetic storms and the more common weaker counterpart. Quantile levels $\tau = \{0.1, 0.2, 0.3\}$ are selected since the 20^{th} percentile of the DST value in this dataset is -49nT, which is approximately the threshold (-50nT) between weak and moderate storms for other works (e.g., [71]). For comparison, τ in the *rqlasso* and *rqnc* methods is set to 0.2. Since the default amount of tuning is instituted, nine different parameter combinations for SCAD are tested. To benchmark the performance of using SCAD as the meta-learner, linear regression is also execute by calling the **caret** method *lm* at the meta-level.

C. Estimating Predictive Performance

For many of the previous studies in predicting geomagnetic storms, the main performance metric utilized has been unweighted error criterion (e.g. root mean square error (RMSE) [23]). While RMSE does penalize larger errors more via the

TABLE III. LIST OF BASE-LEARNERS

Model/Algorithm/Learner	**caret** Method	Model/Algorithm/Learner	**caret** Method
Bagged Regression Trees*	*treebag*	Neural Network*	*nnet*
Bayesian Lasso	*blasso*	Neural Network with Feature Extraction	*pcaNNet*
Bayesian Lasso (Model Averaged)	*blassoAveraged*	Non-convex Penalized Quantile Regression	*rqnc*
Bayesian Regularized Neural Network	*brnn*	Non-negative Least Squares*	*nnls*
Bayesian Ridge Regression	*bridge*	Partial Least Squares*	*pls*
Boosted Linear Model*	*glmboost*	Partitioning Using Deletion, Substitution, and Addition Moves*	*partDSA*
Boosted Tree	*bstTree*	Principal Component Regression	*pcr*
Conditional Inference Random Forest*	*cforest*	Projection Pursuit Regression	*ppr*
Conditional Inference Tree	*ctree*	Quantile Random Forest	*qrf*
Cubist*	*cubist*	Quantile Regression with Lasso Penalty	*rqlasso*
Extreme Gradient Boosting with Linear Booster*	*xgbLinear*	Random Forest*	*ranger*
Extreme Gradient Boosting with Tree Booster*	*xgbTree*	Regression Tree with One Standard Error Rule*	*rpart1SE*
Extreme Learning Machine	*elm*	Regularized Random Forest*	*RRFglobal*
Generalized Additive Model using Loess*	*gamLoess*	Relaxed Lasso	*relaxo*
Generalized Additive Model using Splines*	*gamSpline*	Ridge Regression with Variable Selection	*foba*
Independent Component Regression	*icr*	Robust Linear Model	*rlm*
k-Nearest Neighbors Regression	*kknn*	Self-organizing Map	*bdk*
Lasso and Elastic Net Regression*	*glmnet*	Spike and Slab Regression	*spikeslab*
Least Angle Regression	*lars2*	Stacked AutoEncoder Deep Neural Network	*dnn*
Linear Regression*	*lm*	Stochastic Gradient Boosting*	*gbm*
Linear Regression with Stepwise Selection	*leapSeq*	Supervised Principal Component Analysis	*superpc*
Multi-layer Perceptron	*mlp*	Support Vector Machine with Linear Kernel	*svmLinear*
Multi-layer Perceptron Network by Stochastic Gradient Descent*	*mlpSGD*	Support Vector Machine with Polynomial Kernel	*svmPoly*
Multivariate Adaptive Regression Splines*	*earth*	Support Vector Machine with Radial Basis Function Kernel	*svmRadialSigma*
Multivariate Adaptive Regression Splines (Bagged with Generalized Cross-validation Pruning)*	*bagEarthGCV*	Weighted k-Nearest Neighbors	*knn*

squaring operator, it treats each observation the same. In the context of predicting geomagnetic storms, it is more important for a model to accurately forecast the DST value associated for the strongest of storms. At the same time, focusing strictly on these observations can severely bias a model. Hence, the central metric used in this work for comparison as well as optimizing each learner's parameters is weighted mean absolute error (WMAE), which can be defined as

$$\text{WMAE} = \frac{1}{\sum w_i} \sum_{i=1}^{n} w_i |y_i - \hat{y}_i| \tag{6}$$

such that $\hat{y}_i$ is the predicted response value and w_i is the weight associated with the i^{th} observation. This is implemented within **caret** by creating a custom metric. Adopting WMAE allows for the opportunity to penalize models for inaccuracies when forecasting the more important observations. Given the potential impact that dangerous storms can have, strong CMEs are weighted as 10 times more important than the others (DST > -100nT). Using this 10:1 ratio seems to be a conservative balance since strong geomangetic storms can result in economic losses in trillions of U.S. dollars [72]. In addition to WMAE, the overall RMSE and RMSE for the strong CME events will also be reported.

Each of these error metrics are calculated from an average of ten repeats of 10-fold (10×10) nested cross-validation to ensure a good estimation of error in the presence of parameter tuning [73]. Furthermore, significance tests between the SCAD meta-learner and each individual learner are conducted on the population of error metrics (100 estimates for each from the 10×10 nested cross-validation) using the corrected repeated k-fold cross-validation test [74]. It is important to test for significant differences to investigate if the extra computation of stacked generalization is worth the effort compared to simply using the best performing model [75]. All base-learners and meta-learners are trained over the same folds with the only difference being that the meta-learners use the metadata as its inputs instead of the CME predictor variables. The metadata is generated using 10-fold nested cross-validation. Note that this cross-validation is separate from the nested cross-validation used to estimate the error. Finally, after all of the error testing is complete, each learner is trained on all of the data with the parameters optimized via 10-fold cross-validation. The purpose of this is to enable the variable importance scores extracted from SCAD and the base-learners to be based on all of the data.

IV. RESULTS AND DISCUSSION

Table IV reflects the results of the performance in both stages. The first column lists both meta-learners and the ten most accurate base-learners ranked in ascending order by the average WMAE from the second stage. The subsequent columns represent the averaged error metrics for all CME events (WMAE and RMSE) and only those which triggered a strong geomagnetic disturbance (RMSE). Bold and italics indicate the best performing method. The dagger symbol "†" denotes instances where a significant difference between SCAD and the other learners is *not* found at the conventional

0.05 significance level.

Not surprisingly, accuracy greatly increases in the second stage, a direct consequence of including the IPI. In addition, the majority of the best performing base-learners are all bagging or boosting ensemble models. It is natural for these types of techniques to achieve good predictions. Regardless, SCAD yields the lowest WMAE and RMSE on strong CMEs compared to these in either stage. In addition, SCAD performs statistically better than the most accurate base-learners by themselves for these two metrics in stage two and better than the majority in stage one. This provides evidence that the implementation of stacked generalization here has more predictive power than using just one model. Ting and Witten [39] indicated in their analysis that stacked generalization delivers substantial improvements in accuracy for larger datasets. This is likely due to a more accurate estimation from the cross-validation process when generating the metadata. Hence, it is probable that with more data, stacked generalization can continue to enhance geomagnetic storm prediction over using only one technique.

Notably, the dominance of SCAD falters when evaluating RMSE for all events. This is expected since the majority of the other methods are estimating the conditional mean, rather than a particular part of the DST value's distribution. It is important to reinforce that analyzing the overall RMSE alone can be misleading in this context. For instance, the best performing base-learner in stage one, the regularized random forest, achieves a statistically lower error (RMSE = 27.96) compared to SCAD (RMSE = 34.90). However, looking at the strong CME RMSE reveals a much higher value (87.19 compared to 67.61). This occurs for the linear regression meta-learner as well. Hence, if a practitioner only considered this metric, a degradation in accuracy for the strong events will be realized. Therefore, for practical purposes in predicting geomagnetic storms, it is more appropriate to analyze error metrics such as the WMAE or subsets of RMSE, given the most costly and dangerous storms do not occur very often.

In addition to the arguments above against only considering RMSE on all CME events, additional benefits of using SCAD over linear regression at the meta-level exist, namely the sparsity property. Because linear regression does not perform variable selection, each base-learner prediction is given some weight to make a final estimate. However, with SCAD, certain subsets can be selected, depending on the tuning parameters. This, thereby, reduces the complexity of the problem. During these experiments, SCAD selects 48.31 and 18.71 base-learner predictions on average in stages one and two, respectively. Moreover, any attempt at making any inference at the meta-level using linear regression is frivolous due to the high amount of correlation, which will cause the coefficient estimates to become erratic [76]. Hence, SCAD should be preferred over linear regression for this dataset since it can produce sparser and more interpretable solutions with statistically better error in the important metrics. The quality of being able to dynamically select which base-learners are most useful for prediction at the meta-level may help improve on the fixed form bias issues of stacked generalization mentioned by Vilalta and Drissi [77].

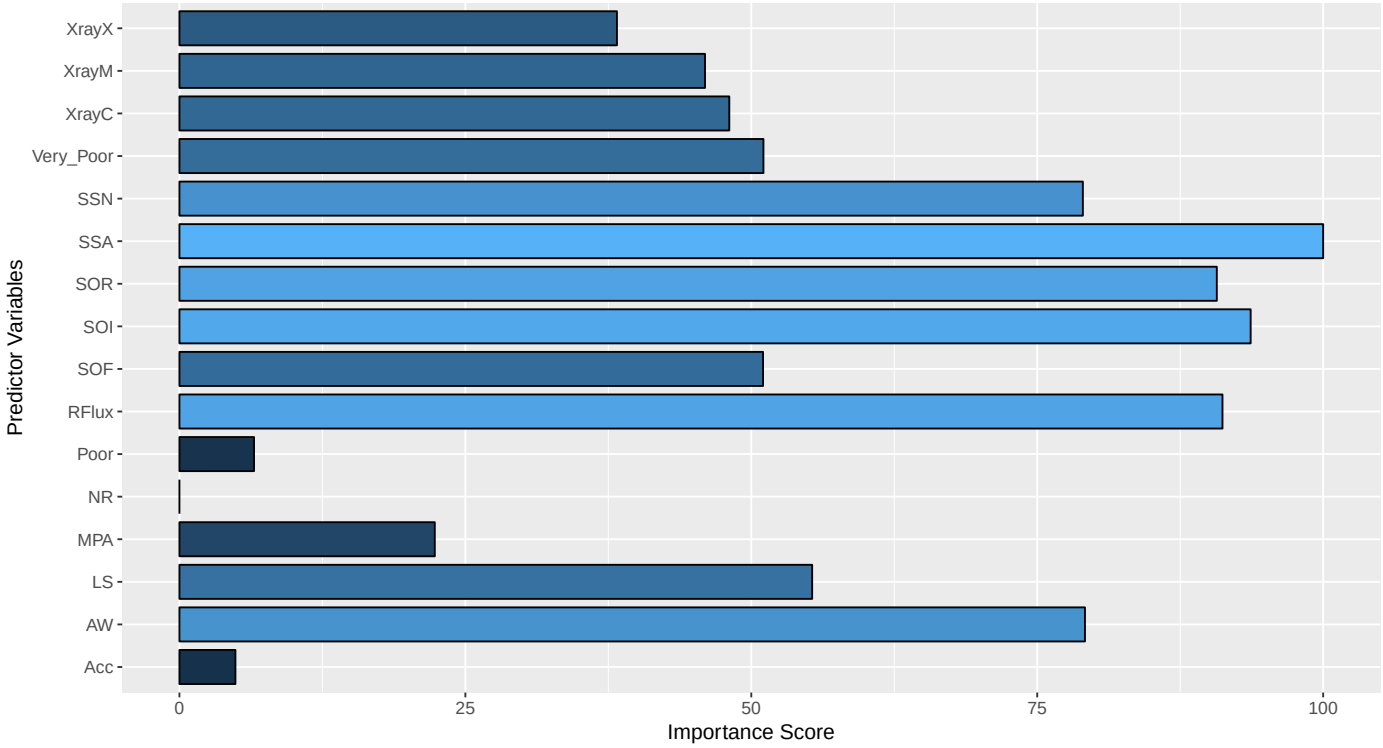
The variable importance scores from stacked generalization can be found in Figure 3 for both stages. Note that these are min-max normalized to represent a score out of 100 where 100 signifies the most useful predictor variable. Note further that since stage one does not use all the predictor variables, not all are listed. The most significant predictor variable in stage one for predicting DST is the sunspot area (SSA). Its high ranking makes sense since sunspot activity can be closely tied with CME occurrences [2]. In stage two, the two most dominating are E_y (which is an interaction between B_z and V_{sw}) and B_z. Given the strong relationship between these predictor variables and the DST value throughout the literature, their contributions towards prediction makes sense. More importantly, the higher values placed on E_y and B_z and lower values on those such as D_p and T_p in determining geomagnetic storm intensity is consistent with other literature (see [26][78][79][80][27] and references therein). Note that when the IPI is introduced, the influence of the stage one predictor variables decreases. This is to be expected given the advantages of using IPI.

Though the study of stacked generalization is not a new concept, this idea has not been explored in the realm of forecasting geomagnetic storm strength from CMEs much if not at all. Given the importance of making forecasts, it becomes all the more important to leverage the best analytical tools for space weather prediction. As shown in other studies, it is necessary to incorporate IPI since these are the most useful for determining the DST value. However, as emphasized by Kim et al. [27], this leaves little time to prepare on Earth once the information is collected at the L1 Lagrangian point. Research in attaining IPI sooner is currently being done. Savani et al. [81] are working towards resolving this type of issue by predicting the magnetic structure of impending CMEs. More accurate forecasts of the IPI will lead to better predictions with more lead time. In addition, since time is such a factor, computationally efficient approaches must be used. Luckily, although stacked generalization requires extra computation, especially for large datasets, it can be easily parallelized across many clusters since creating the metadata is an independent process. This allows for scalability as new models and algorithms are constantly being developed. Incorporating a larger number of faster and smarter base-learners provides the opportunity to increase predictive power.

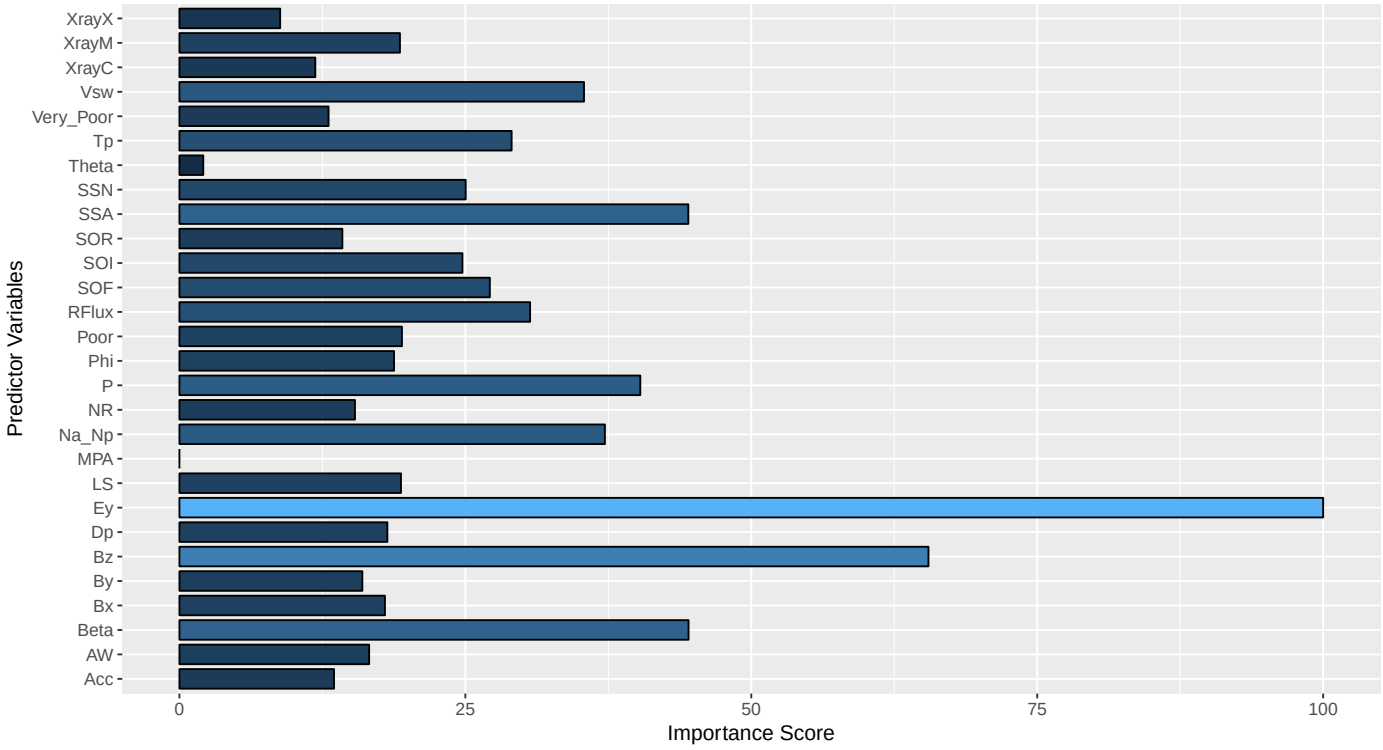
This study brings several future work opportunities. Firstly, as more data is collected on CMEs in more advanced ways, implementation on larger datasets is possible for both classification and regression tasks. With more data, stacked generalization is more probable to find predictive improvements [39]. Secondly, this work only includes 50 base-learners. Increasing this number by incorporating different models and algorithms could yield even better results. With regards to the respective variable importances, exploring ways to extract model-specific measures can be investigated, despite whether a model or algorithm inherently implements them or not. Additionally, analyzing the variable importance scores at different quantiles may reveal some new behaviors regarding the predictor variables, much like in quantile process regression [82]. Furthermore, introducing some type of cost matrix, as done for MetaFraud [35], or re-weighting WMAE can better optimize parameters at both the base and meta-levels.

V. Conclusion

In this work, a meta-learning framework is suggested to predict geomagnetic storms. This approach consists of two stages:



(a) Stage 1



(b) Stage 2

Figure 3. Variable importance scores from stacked generalization in both stages.

TABLE IV. PREDICTIVE PERFORMANCE

	Stage 1			Stage 2		
Learner	All CMEs		Strong CMEs	All CMEs		Strong CMEs
(Meta)	WMAE	RMSE	RMSE	WMAE	RMSE	RMSE
SCAD	***33.59***	34.90	***67.61***	***18.44***	18.76	***39.17***
Linear Regression	35.96	28.86	91.63	19.41†	***17.85***	46.09
(Base)	WMAE	RMSE	RMSE	WMAE	RMSE	RMSE
Cubist	38.35	30.98	96.44	20.04	18.20†	47.04
Extreme Gradient Boosting with Linear Booster	35.11†	29.41	85.28	20.41	19.32†	51.06
Extreme Gradient Boosting with Tree Booster	35.56†	28.80	85.84	20.68	19.25†	49.40
Random Forest	35.70†	28.21	87.98	20.74	18.32†	50.27
Regularized Random Forest	35.48†	***27.96***	87.19	20.88	18.39†	50.95
Boosted Tree	37.59	29.08	92.71	21.27	19.09†	51.86
Multivariate Adaptive Regression Splines (Bagged with Generalized Cross-Validation Pruning)	39.57	29.89	99.45	21.84	19.17†	51.70
Stochastic Gradient Boosting	38.48	29.78	95.03	21.95	19.44†	52.30
Conditional Inference Random Forest	39.70	29.92	101.74	22.07	19.00†	53.87

1) Execute stacked generalization with a SCAD penalized quantile meta-learner to make a preliminary estimate of DST based on initial CME and Sun data.
2) Update the prediction with another round of stacked generalization after collecting the vital IPI.

The general outline is similar to the process by Kim et al. [27]. However, instead of focusing on estimating the conditional mean for DST, quantile regression is implemented to find a better balance between predicting dangerous geomagnetic storms effectively without rendering estimation for the weaker ones useless. Using a regularized quantile regression model at the meta-level provides more adaptability since it can specify specific parts of the conditional distribution and choose the best number of base-learners for that particular region. The posited method is evaluated on an inclusive dataset consisting of various characteristics about the solar wind condition, CMEs, and the Sun. In addition, careful experimental methodology is utilized to estimate generalization error and statistical significance. Results show that this framework performs significantly better on the most informative error metrics than the best tuned model or algorithm at the base-level. Moreover, this approach provides an opportunity to study the critical space weather indicators at the beginning of a CME's life and right before its impact on Earth via the variable importance scores from stacked generalization.

Given our dependence on telecommunications and commercial satellites, any disruption in these services could cost millions of dollars for corporations and government agencies world-wide. At the same time, logistically, these entities cannot simply shut down power or telecommunication operations every time a CME approaches Earth. Therefore, it is imperative to make accurate classifications and forecasts as to which of these CMEs that approach Earth can have the potential to trigger devastating geomagnetic storms. Putting into action more sophisticated modeling techniques like stacked generalization have the opportunity to improve predictions. Instituting these in multi-step approaches can greatly benefit in preparation time for geomagnetic storms. Utilizing more complex systems enables the ability to make more accurate predictions, thereby, saving money and reducing the probability for severe geomagnetic storm events wreaking havoc on modern society.

ACKNOWLEDGMENT

We would like to thank NASA for their images and the creation of the CME catalog. This CME catalog is generated and maintained at the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA. In addition, we would like to thank the Goddard Space Flight Center/Space Physics Data Facility (GSFC/SPDF), OMNIWeb, and NOAA for their public use databases. An earlier version of this research was presented at Data Analytics 2016: The Fifth International Conference on Data Analytics.

REFERENCES

[1] T. Larkin and D. McManus, "Impact of analytics and meta-learning on predicting geomagnetic storms: Risk to global telecommunications," in Data Analytics 2016, The Fifth International Conference on Data Analytics, S. Bhulai and I. Semanjski, Eds. IARIA, October 2016, pp. 8–13.

[2] T. Howard, Coronal Mass Ejections: An Introduction. Springer Science & Business Media, 2011, vol. 376.

[3] N. Srivastava and P. Venkatakrishnan, "Solar and interplanetary sources of major geomagnetic storms during 1996–2002," Journal of Geophysical Research: Space Physics (1978–2012), vol. 109, no. A10, 2004, pp. 1–13.

[4] National Oceanic and Atmospheric Administration, "Coronal mass ejections," Available: http://www.swpc.noaa.gov/phenomena/coronal-mass-ejections [accessed: 2017-05-22].

[5] R. MacQueen, "Coronal transients: A summary," Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 297, no. 1433, 1980, pp. 605–620.

[6] National Aeronautics and Space Administration, "Coronal mass ejections," Available: http://helios.gsfc.nasa.gov/cme.html [accessed: 2017-05-22].

[7] National Oceanic and Atmospheric Administration, "Earth's magnetosphere," Available: http://www.swpc.noaa.gov/phenomena/earths-magnetosphere [accessed: 2017-05-22].

[8] National Aeronautics and Space Administration, "Magnetospheres," Available: https://science.nasa.gov/heliophysics/focus-areas/magnetosphere-ionosphere [accessed: 2017-05-22].

[9] J. W. Dungey, "Interplanetary magnetic field and the auroral zones," Physical Review Letters, vol. 6, no. 2, 1961, pp. 47–48.

[10] D. H. Fairfield and L. Cahill, "Transition region magnetic field and polar magnetic disturbances," Journal of Geophysical Research, vol. 71, no. 1, 1966, pp. 155–169.

[11] W. D. Gonzalez and B. T. Tsurutani, "Criteria of interplanetary parameters causing intense magnetic storms (DST <-100 nT)," Planetary and Space Science, vol. 35, no. 9, 1987, pp. 1101–1109.

[12] Y. Wang, P. Ye, S. Wang, G. Zhou, and J. Wang, "A statistical study on the geoeffectiveness of Earth-directed coronal mass ejections from March 1997 to December 2000," Journal of Geophysical Research: Space Physics (1978–2012), vol. 107, no. A11, 2002, pp. SSH 2–1–SSH 2–9.

[13] J. Kappernman and V. D. Albertson, "Bracing for the geomagnetic storms," Spectrum, IEEE, vol. 27, no. 3, 1990, pp. 27–33.

[14] Space Studies Board and others, Severe Space Weather Events: Understanding Societal and Economic Impacts: A Workshop Report. National Academies Press, 2008.

[15] D. Baker, X. Li, A. Pulkkinen, C. Ngwira, M. Mays, A. Galvin, and K. Simunac, "A major solar eruptive event in July 2012: Defining extreme space weather scenarios," Space Weather, vol. 11, no. 10, 2013, pp. 585–591.

[16] J. Gosling, S. Bame, D. McComas, and J. Phillips, "Coronal mass ejections and large geomagnetic storms," Geophysical Research Letters, vol. 17, no. 7, 1990, pp. 901–904.

[17] V. Bothmer and R. Schwenn, "The interplanetary and solar causes of major geomagnetic storms," Journal of Geomagnetism and Geoelectricity, vol. 47, no. 11, 1995, pp. 1127–1132.

[18] B. T. Tsurutani and W. D. Gonzalez, "The interplanetary causes of magnetic storms: A review," Washington DC American Geophysical Union Geophysical Monograph Series, vol. 98, 1997, pp. 77–89.

[19] J. Zhang, K. Dere, R. Howard, and V. Bothmer, "Identification of solar sources of major geomagnetic storms between 1996 and 2000," The Astrophysical Journal, vol. 582, no. 1, 2003, pp. 520–533.

[20] D. McManus, H. Carr, and B. Adams, "Wireless on the precipice: The 14th century revisited," Communications of the ACM, vol. 54, no. 6, 2011, pp. 138–143.

[21] R. K. Burton, R. McPherron, and C. Russell, "An empirical relationship between interplanetary conditions and DST," Journal of Geophysical Research, vol. 80, no. 31, 1975, pp. 4204–4214.

[22] M. Sugiura, "Hourly values of equatorial DST for the IGY," Ann. Int. Geophys. Yr., vol. 35, 1964, pp. 1–44.

[23] E.-Y. Ji, Y.-J. Moon, N. Gopalswamy, and D.-H. Lee, "Comparison of DST forecast models for intense geomagnetic storms," Journal of Geophysical Research: Space Physics, vol. 117, no. A3, 2012, pp. 1–9.

[24] T. Andriyas and S. Andriyas, "Relevance vector machines as a tool for forecasting geomagnetic storms during years 1996–2007," Journal of Atmospheric and Solar-Terrestrial Physics, vol. 125, 2015, pp. 10–20.

[25] M. Dryer, Z. Smith, C. Fry, W. Sun, C. Deehr, and S.-I. Akasofu, "Real-time shock arrival predictions during the halloween 2003 epoch," Space Weather, vol. 2, no. 9, 2004, pp. 1–10.

[26] N. Srivastava, "A logistic regression model for predicting the occurrence of intense geomagnetic storms," in Annales Geophysicae, vol. 23, no. 9, 2005, pp. 2969–2974.

[27] R.-S. Kim, Y.-J. Moon, N. Gopalswamy, Y.-D. Park, and Y.-H. Kim, "Two-step forecast of geomagnetic storm using coronal mass ejection and solar wind condition," Space Weather, vol. 12, no. 4, 2014, pp. 246–256.

[28] J. Uwamahoro, L. McKinnell, and J. Habarulema, "Estimating the geoeffectiveness of halo CMEs from associated solar and IP parameters using neural networks," Annales Geophysicae-Atmospheres Hydrospheresand Space Sciences, vol. 30, no. 6, 2012, pp. 963–972.

[29] F. Valach, J. Bochníček, P. Hejda, and M. Revallo, "Strong geomagnetic activity forecast by neural networks under dominant southern orientation of the interplanetary magnetic field," Advances in Space Research, vol. 53, no. 4, 2014, pp. 589–598.

[30] R. Schwenn, "Space weather: The solar perspective," Living Reviews in Solar Physics, vol. 3, no. 1, 2006, pp. 1–72.

[31] V. Yurchyshyn, V. Abramenko, and D. Tripathi, "Rotation of white-light coronal mass ejection structures as inferred from lasco coronagraph," The Astrophysical Journal, vol. 705, no. 1, 2009, p. 426.

[32] D. M. Hawkins, "The problem of overfitting," Journal of Chemical Information and Computer Sciences, vol. 44, no. 1, 2004, pp. 1–12.

[33] T. Larkin, "Advanced analytical tools for geomagnetic storm prediction: Ensembles and their insights," Ph.D. dissertation, The University of Alabama, 2017, unpublished thesis.

[34] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, 1992, pp. 241–259.

[35] A. Abbasi, C. Albrecht, A. Vance, and J. Hansen, "Metafraud: A meta-learning framework for detecting financial fraud," MIS Quarterly, vol. 36, no. 4, 2012, pp. 1293–1327.

[36] C.-F. Tsai and Y.-F. Hsu, "A meta-learning framework for bankruptcy prediction," Journal of Forecasting, vol. 32, no. 2, 2013, pp. 167–179.

[37] J. Sill, G. Takács, L. Mackey, and D. Lin, "Feature-weighted linear stacking," arXiv preprint arXiv:0911.0460, 2009, pp. 1–17.

[38] L. Breiman, "Stacked regressions," Machine Learning, vol. 24, no. 1, 1996, pp. 49–64.

[39] K. M. Ting and I. H. Witten, "Issues in stacked generalization," J. Artif. Intell. Res.(JAIR), vol. 10, 1999, pp. 271–289.

[40] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in Proceedings of the Twenty-first International Conference on Machine Learning. ACM, 2004, pp. 18–25.

[41] M. LeBlanc and R. Tibshirani, "Combining estimates in regression and classification," Journal of the American Statistical Association, vol. 91, no. 436, 1996, pp. 1641–1650.

[42] S. Reid and G. Grudic, "Regularized linear models in stacked generalization," in Multiple Classifier Systems. Springer, 2009, pp. 112–121.

[43] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 1970, pp. 55–67.

[44] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), 1996, pp. 267–288.

[45] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, no. 2, 2005, pp. 301–320.

[46] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in Machine Learning: ECML 2001. Springer, 2001, pp. 576–587.

[47] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," Artificial Intelligence, vol. 137, no. 1, 2002, pp. 239–263.

[48] N. Rooney, D. Patterson, and C. Nugent, "Pruning extensions to stacking," Intelligent Data Analysis, vol. 10, no. 1, 2006, pp. 47–66.

[49] B. Peng and L. Wang, "An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression," Journal of Computational and Graphical Statistics, vol. 24, no. 3, 2015, pp. 676–694.

[50] F. Mosteller and J. W. Tukey, "Data analysis and regression: A second course in statistics," Addison-Wesley Series in Behavioral Science: Quantitative Methods, 1977.

[51] B. S. Cade and B. R. Noon, "A gentle introduction to quantile regression for ecologists," Frontiers in Ecology and the Environment, vol. 1, no. 8, 2003, pp. 412–420.

[52] R. Koenker and G. Bassett Jr, "Regression quantiles," Econometrica: Journal of the Econometric Society, 1978, pp. 33–50.

[53] R. Koenker and K. Hallock, "Quantile regression: An introduction," Journal of Economic Perspectives, vol. 15, no. 4, 2001, pp. 43–56.

[54] R. Koenker, "Quantile regression for longitudinal data," Journal of Multivariate Analysis, vol. 91, no. 1, 2004, pp. 74–89.

[55] Y. Li and J. Zhu, "L1-norm quantile regression," Journal of Computational and Graphical Statistics, 2012.

[56] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," Journal of the American Statistical Association, vol. 96, no. 456, 2001, pp. 1348–1360.

[57] H. Zou, "The adaptive lasso and its oracle properties," Journal of the American Statistical Association, vol. 101, no. 476, 2006, pp. 1418–1429.

[58] F. Audrino and L. Camponovo, "Oracle properties and finite sample

inference of the adaptive lasso for time series regression models," arXiv preprint arXiv:1312.1473, 2013.

[59] Y. Wu and Y. Liu, "Variable selection in quantile regression," Statistica Sinica, 2009, pp. 801–817.

[60] H. Cane and I. Richardson, "Interplanetary coronal mass ejections in the near-earth solar wind during 1996–2002," Journal of Geophysical Research: Space Physics (1978–2012), vol. 108, no. A4, 2003, pp. SSH 6–1–SSH 6–13.

[61] I. Richardson and H. Cane, "Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): Catalog and summary of properties," Solar Physics, vol. 264, no. 1, 2010, pp. 189–237.

[62] J. King and N. Papitashvili, "Solar wind spatial scales in and comparisons of hourly wind and ACE plasma and magnetic field data," Journal of Geophysical Research: Space Physics, vol. 110, no. A2, 2005, pp. 1–8.

[63] N. Gopalswamy, S. Yashiro, G. Michalek, G. Stenborg, A. Vourlidas, S. Freeland, and R. Howard, "The SOHO/LASCO CME catalog," Earth, Moon, and Planets, vol. 104, no. 1-4, 2009, pp. 295–313.

[64] National Oceanic and Atmospheric Administration, "Index of /pub/warehouse," Available: ftp://ftp.swpc.noaa.gov/pub/warehouse [accessed: 2017-05-22].

[65] G.-H. Moon, "Variation of Magnetic Field (By, Bz) Polarity and Statistical Analysis of Solar Wind Parameters during the Magnetic Storm Period," Journal of Astronomy and Space Sciences, vol. 28, no. 2, 2011, pp. 123–132.

[66] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2017, Available: https://www.R-project.org/ [accessed: 2017-05-22].

[67] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt., caret: classification and regression training, 2016, R package version 6.0-70. Available: https://CRAN.R-project.org/package=caret [accessed: 2016-07-27].

[68] M. Kuhn, "Building predictive models in R using the caret package," Journal of Statistical Software, vol. 28, no. 5, 2008, pp. 1–26.

[69] B. Sherwood and A. Maidman, rqPen: Penalized Quantile Regression, 2016, R package version 1.4. Available: https://CRAN.R-project.org/package=rqPen [accessed: 2016-07-27].

[70] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Journal of Statistical Software, vol. 33, no. 1, 2010, p. 1.

[71] C. Loewe and G. Prölss, "Classification and mean behavior of magnetic storms," Journal of Geophysical Research: Space Physics, vol. 102, no. A7, 1997, pp. 14 209–14 213.

[72] Lloyd's and the Atmospheric and Environmental Research Inc., "Solar storm risk to the north american electrical grid," 2013, Available: https://www.lloyds.com/~/media/lloyds/reports/emerging%20risk%20reports/solar%20storm%20risk%20to%20the%20north%20american%20electric%20grid.pdf [accessed: 2017-05-22].

[73] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," BMC Bioinformatics, vol. 7, no. 1, 2006, p. 91.

[74] R. R. Bouckaert and E. Frank, "Evaluating the replicability of significance tests for comparing learning algorithms," in Advances in Knowledge Discovery and Data Mining. Springer, 2004, pp. 3–12.

[75] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?" Machine Learning, vol. 54, no. 3, 2004, pp. 255–273.

[76] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2009.

[77] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," Artificial Intelligence Review, vol. 18, no. 2, 2002, pp. 77–95.

[78] E. Echer, W. Gonzalez, and B. Tsurutani, "Interplanetary conditions leading to superintense geomagnetic storms (DST <= -250 nT) during solar cycle 23," Geophysical Research Letters, vol. 35, no. 6, 2008.

[79] Y. I. Yermolaev, M. Y. Yermolaev, I. Lodkina, and N. Nikolaeva, "Statistical investigation of heliospheric conditions resulting in magnetic storms: 2," Cosmic Research, vol. 45, no. 6, 2007, pp. 461–470.

[80] E.-Y. Ji, Y.-J. Moon, K.-H. Kim, and D.-H. Lee, "Statistical comparison of interplanetary conditions causing intense geomagnetic storms (DST <= -100 nT)," Journal of Geophysical Research: Space Physics (1978–2012), vol. 115, no. A10, 2010.

[81] N. Savani, A. Vourlidas, A. Szabo, M. Mays, I. Richardson, B. Thompson, A. Pulkkinen, R. Evans, and T. Nieves-Chinchilla, "Predicting the magnetic vectors within coronal mass ejections arriving at earth: 1. initial architecture," Space Weather, 2015.

[82] R. Koenker and J. A. Machado, "Goodness of fit and related inference processes for quantile regression," Journal of the American Statistical Association, vol. 94, no. 448, 1999, pp. 1296–1310.