Correspondence to:
A. W. Smith,
andy.w.smith@ucl.ac.uk

Probabilistic Forecasts of Storm Sudden Commencements From Interplanetary Shocks Using Machine Learning

A. W. Smith[1] [ID], **I. J. Rae[1,2]** [ID], **C. Forsyth[1]** [ID], **D. M. Oliveira[3,4]** [ID], **M. P. Freeman[5]** [ID], **and D. R. Jackson[6]** [ID]

[1]Mullard Space Science Laboratory, UCL, Dorking, UK, [2]Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle upon Tyne, UK, [3]Goddard Planetary Heliophysics Institute, University of Maryland, Baltimore County, Baltimore, MD, USA, [4]NASA Goddard Space Flight Center, Greenbelt, MD, USA, [5]British Antarctic Survey, Cambridge, UK, [6]Met Office, Exeter, UK

Abstract In this study we investigate the ability of several different machine learning models to provide probabilistic predictions as to whether interplanetary shocks observed upstream of the Earth at L1 will lead to immediate (Sudden Commencements, SCs) or longer lasting magnetospheric activity (Storm Sudden Commencements, SSCs). Four models are tested including linear (Logistic Regression), nonlinear (Naive Bayes and Gaussian Process), and ensemble (Random Forest) models and are shown to provide skillful and reliable forecasts of SCs with Brier Skill Scores (BSSs) of ~0.3 and ROC scores >0.8. The most powerful predictive parameter is found to be the range in the interplanetary magnetic field. The models also produce skillful forecasts of SSCs, though with less reliability than was found for SCs. The BSSs and ROC scores returned are ~0.21 and 0.82, respectively. The most important parameter for these predictions was found to be the minimum observed B_Z. The simple parameterization of the shock was tested by including additional features related to magnetospheric indices and their changes during shock impact, resulting in moderate increases in reliability. Several parameters, such as velocity and density, may be able to be more accurately predicted at a longer lead time, for example, from heliospheric imagery. When the input was limited to the velocity and density the models were found to perform well at forecasting SSCs, though with lower reliability than previously (BSSs ~ 0.16, ROC Scores ~ 0.8), Finally, the models were tested with hypothetical extreme data beyond current observations, showing dramatically different extrapolations.

1. Introduction

Space weather events are ultimately driven by the interaction between the solar wind and the magnetosphere-ionosphere system. These interactions can be characterized as the storage and spontaneous release of energy, leading to intermittent, shorter space weather events (e.g., substorms Freeman et al., 2019), or by the driving of longer extreme space weather events by large-scale structures in the solar wind, such as Coronal Mass Ejections (CMEs) (Illing & Hundhausen, 1983; Gosling, 1993; Richardson & Cane, 2012). CMEs are huge eruptions of plasma from the Sun, exploding into the solar system, often driving fast-forward shocks ahead of them (see review by Webb & Howard, 2012). The impact of CMEs, and their associated interplanetary shocks, on the magnetosphere can drive dynamics and processes such as magnetospheric storms and substorms (e.g., Akasofu & Chao, 1980; Brueckner et al., 1998; Gonzalez et al., 1994; Kamide et al., 1998; Kokubun et al., 1977; Yue et al., 2010; Zhou & Tsurutani, 2001). CMEs associated with high speed interplanetary shocks and significant intervals of southward directed interplanetary magnetic fields (i.e., negative IMF Bz) have been found to be particularly geo-effective (Balan et al., 2014; Echer et al., 2008). Other phenomena can also drive interplanetary shocks, for example, corotating interaction regions (CIRs): structures produced when high speed solar wind interacts with slower moving solar wind in its path (see review by Crooker et al., 1999). However, CIR-driven shocks are not often associated with the same large negative magnetic fields and therefore may drive less severe but longer lasting magnetospheric activity (Alves et al., 2006; Borovsky & Denton, 2006).

On the ground, the impact of a significant interplanetary shock on the magnetosphere can often be inferred through a sharp increase in the northward component of the horizontal magnetic field, known as a Sudden Commencement (or SC) (Araki, 1994; Chree, 1925). If the shock impact is followed by a geomagnetic storm

then it may be further termed a Storm Sudden Commencement (or SSC); if it is not then it can be classed as a Sudden Impulse (SI) (Curto et al., 2007; Joselyn & Tsurutani, 1990). It should be noted that the impact of a shock is not always guaranteed to cause large, measurable changes (i.e., an SC) in the magnetosphere (e.g., Echer & Gonzalez, 2004).

The initial impact of some interplanetary shocks and the induced sharp changes in ground magnetic field have been found to generate large currents in power networks, particularly at low latitude and midlatitude (Beland & Small, 2004; Carter et al., 2015; Kappenman, 2003; Marshall et al., 2012; Zhang et al., 2015). Meanwhile, the magnetospheric storms and substorms that may be caused by the shock, or the phenomena driving the shock, are also likely to be associated with large rates of change of the geomagnetic field (and therefore induced currents) (Dimmock et al., 2019; Freeman et al., 2019; Kappenman & Albertson, 1990; Kappenman, 1996; Ngwira et al., 2013; Pulkkinen et al., 2005, 2012), particularly at higher latitudes where the auroral current systems most often reside (Rogers et al., 2020). Recent work has shown that for the United Kingdom (at midlatitudes) over 90% of extreme geomagnetic field fluctuations occur within 3 days of an SSC (Smith et al., 2019). Interplanetary shocks therefore represent an important source of hazardous space weather, whether directly or in connection with the phenomena that drives the shock. Links have been observed between the properties of interplanetary shocks and the generated geomagnetic (Gonzalez et al., 1999; Oliveira & Raeder, 2015; Tsurutani et al., 1992; Tsurutani & Gonzalez, 1998) and auroral observations (Oliveira et al., 2016). With this in mind it is of great interest to be able to forecast intervals in which infrastructure is at risk, that is, to produce skillful models that are able to accurately forecast the geoeffectiveness of interplanetary shocks.

From an operational perspective, it is possible to automatically identify interplanetary shocks in spacecraft data, for example at L1 (e.g., Cash et al., 2014; Kruparova et al., 2013; Vorotnikov et al., 2008, 2011). At a minimum, the delay from L1 to the subsolar magnetopause provides appropriately 30–120 min of warning of the shock's arrival at Earth, though the precise delay is variable and not simple to predict (Cash et al., 2016). Further classification at the L1 point to determine the phenomena related to the shock, for example, a CME or CIR, is more difficult to automate but would likely aid forecasts (e.g., Echer et al., 2008). Recent work has had success forecasting whether an interplanetary shock will be followed by an extended interval of southward magnetic field (Salman et al., 2018), as these intervals are known to be particularly geo-effective (e.g., Gonzalez & Tsurutani, 1987).

In the last 20 years, studies have begun to leverage machine learning techniques to forecast space weather; the interested reader is directed to Camporeale (2019) for a detailed overview. In particular, machine learning techniques have been used to forecast geomagnetic indices (cf. Morley, 2020), for example the Kp index (e.g., Ji et al., 2013; Tan et al., 2018; Wang et al., 2017; Wing et al., 2005; Wintoft et al., 2017) and Dst/Sym-H indices (e.g., Bhaskar & Vichare, 2019; Chandorkar et al., 2017; Kugblenu et al., 1999; Lethy et al., 2018; Lundstedt et al., 2002; Wu & Lundstedt, 1996). In this work, we investigate the ability of machine learning methods to provide a probabilistic forecast as to whether an observed interplanetary shock will lead to an SC (e.g., a significant ground magnetic field signature), and further whether this will be followed by a geomagnetic storm (i.e., the shock is related to an SSC). We also examine the performance of these models when presented with inputs at and beyond the limits of the training dataset. As discussed above, the geo-effectiveness can likely be more thoroughly determined through inspection of the physical process driving the shock; however, operationally the details of this structure may not be quantified a priori. Therefore, we investigate whether an SSC can be predicted based solely on information about the shock.

The structure of the paper is as follows. In section 2 we discuss the data and catalogs of events employed, while in section 3 we discuss our methods, models, and metrics. Section 4 describes the results of the study, while section 5 discusses the results in terms of the most powerful predictive parameters, the effectiveness of the parameterization of the shock, and the performance of the models when presented with extreme events.

2. Data and Catalogs

For this work we require catalogs of interplanetary shocks (measured at L1) and sudden commencements (SCs), which are further classified into sudden impulses (SIs) and storm sudden commencements (SSCs). We note that any catalog may not be fully comprehensive, as marginal events may not be identified. However, it is reasonable to assume that the largest events will be present, and it is these extreme events that are likely

to be the most significant in terms of space weather. For this purpose, we have selected two independent catalogs, described below.

2.1. Interplanetary Shocks and Parameters

We use a catalog of 547 interplanetary shocks observed by either the Wind or ACE spacecraft at the L1 point between 1995 and 2017, compiled by Oliveira, Arel, et al. (2018). This database is composed of interplanetary shocks selected by Oliveira, Arel, et al. (2018), as well as those identified by Dr. J. C. Kasper for the Wind (`http://www.cfa.harvard.edu/shocks/wi_data/`) and ACE data (`http://www.cfa.harvard.edu/shocks/ac_master_data/`) and also by the ACE team (|`http://www-ssg.sr.unh.edu/mag/ace/ACElists/obs_list.html\#shocks`|), and by Wang et al. (2010).

As a part of this study we wish to evaluate the most powerful predictive properties of solar wind shocks, and we therefore extract a wide variety of parameters from the observations of the shock and the surrounding solar wind. Specifically we record the maximum, minimum, range, and mean of the GSM components of the solar wind velocity (V) and interplanetary magnetic field (B), the magnitude of B, and the solar wind proton density, henceforth referred to as features. We do not consider properties that can be further calculated from such features, for example, dynamic pressure or the various coupling parameters, as they strongly correlate with the existing features. The properties of the interplanetary shocks were extracted from data obtained from the ACE and WIND spacecraft within the hour preceding shock arrival at the Earth (as inferred from an increase in the Sym-H index). The 1 hr window was determined empirically in order to include all shocks and account for the different propagation times between L1 and the subsolar magnetopause. Though this scheme will sample different quantities of pre- and post-shock plasma, depending on the speed of the shock, the extracted parameters are found to be dominated by the shock itself and any associated field rotation. Adjusting the interval of time sampled around the shock was not found to significantly change the results.

From ACE, data were used from the Solar Wind Electron, Proton and Alpha Monitor at 64 s resolution (McComas et al., 1998) and the Magnetic Field Experiment at 16 s resolution (Smith et al., 1998). Meanwhile from WIND, data were used from the Solar Wind Experiment at 92 s resolution (Ogilvie et al., 1995) and the Magnetic Fields Investigation at 1 min resolution (Lepping et al., 1995). We note that the data from ACE and WIND can be incomplete, particularly during more extreme solar wind conditions. As it is our goal to demonstrate a method that could be used for operational forecasting, our feature selection method has been chosen such that they may be extracted from incomplete data. Operationally, that is, from a forecasting perspective, it would be important to minimize the effect of any missing data on the predictions made. However, if the data quality were assured and continuous, then it would likely be of significant value to include more detailed parameters of the shock, such as the shock impact angle (Oliveira & Samsonov, 2018). It should also be noted that this study has been performed using science data (i.e., "Level 2 data"), any future operational form should be trained on the more immediately available data products to provide a more representative comparison.

2.2. Sudden Commencements

To evaluate the geo-effectiveness of the interplanetary shocks we use an independent database of 417 sudden commencements (SCs), for the same time interval as above, provided by the International Service of Rapid Magnetic Variations (part of the International Service of Geomagnetic Indices) based at Ebre Observatory (`http://www.obsebre.es/en/rapid`). These events are identified from the data collected by five low-latitude magnetic observatories (Curto et al., 2007). The catalog has been used in the past to statistically assess the consequences of SCs and related activity (e.g., Carter et al., 2015; Fiori et al., 2014; Smith et al., 2019). We further classify the SCs into Sudden Impulses (SI) or Storm Sudden Commencements (SSC) according to the scheme of Fiori et al. (2014), based on the Sym-H index in the days following the observation (cf. Gonzalez et al., 1994). In this scheme the events are classed as an SI if the Sym-H index is greater than −30 nT for the 48 hr following the SC, an SSC if the Sym-H index falls to less than 30 nT within 4 hr of the SC, and as a delayed SSC (SSCd) if the fall to below 30 nT occurs between 4 and 48 hr after the SC. It should be noted that these definitions do not include any recognition of SSCs that may be identified by the changing magnetic "rhythm" of the stations (Mayaud, 1973); however, it is simple and easily reproducible.

Table 1
Cross-Comparison IP Shocks (Oliveira, Arel, et al. 2018) with SCs Identified by the ISGI, Classified With the Scheme of Fiori et al. (2014)

	IP shock	No IP shock	Totals
SC	307	107	414
No SC	240	-	240
Totals	547	107	
SSC	94	31	125
SSCd	167	49	216
SI	46	27	73
No SC	240	-	
Totals	547	107	

2.3. Cross-Comparison

We can cross-compare the events in both catalogs to evaluate whether the interplanetary shocks identified at L1 caused a ground signature that was recognized as an SC, and whether any further activity was observed. To do so, the SC database was checked for corresponding SCs within ±15 min of each shock impact. This choice of window size was confirmed to correctly match those that would be associated manually. Table 1 shows a comparison between the two catalogs.

First from Table 1 and the perspective of the shock observations, we can see that 240 of the 547 interplanetary shocks in the interval (44%) did not cause a significant ground signature, that is, an identified SC. This relatively large fraction is expected: An interplanetary shock impact is not a sufficient criterion for an SC to be observed (e.g., Echer & Gonzalez, 2004). Meanwhile, 94 out of 125 (75%) SSCs can be directly related to solar wind shocks. This is entirely consistent with the findings of a shorter time interval studied by Wang et al. (2006). From an similar perspective, while 94 interplanetary shocks (out of 547, 17%) directly correspond to SSCs, a total of 261 (out of 547, 48%) can be linked to some form of storm activity (i.e., SSCs or SSCds). This is similar to the observations of Echer and Gonzalez (2004), who found that between 1973 and 2000 57% of their 574 interplanetary shocks were followed by a Dst below −50 nT in the following 3 days.

Meanwhile from the perspective of SCs, 107 of the 414 (25%) SCs recorded during this interval do not correspond to an identified interplanetary shock at L1. This is very similar to the fraction found by Wang et al. (2006) in their study between 1995 and 2004; such events have been attributed to other phenomena in the solar wind (e.g., Park et al., 2015; Tsurutani et al., 2011). Additionally, it is possible that there are data gaps due to extreme plasma conditions at L1 which prohibit the identification of a counterpart shock. This would likely preferentially effect more extreme solar wind conditions. On the other hand, it is also possible that some of the SCs that lack an interplanetary counterpart could be due to the independent nature of the catalogs. It is often the case that when surveying data manually, smaller scale events can be more noticeable when the time period is highlighted by the occurrence of a related phenomenon. With regard to the SSC and SSCds that have an unidentified interplanetary cause, it should be noted that our statistics will include relatively small storms due to our adoption of a low threshold of 30 nT. Previous works have suggested a very strong association between large storms and interplanetary shocks (e.g., Chao & Lepping, 1974; Gosling et al., 1991). Interestingly, if we increase our storm threshold to −50 nT or even −100 nT we still return 22 − 25% of SSCs that do not correspond to an identified interplanetary shock. This suggests that a combination of the factors described above may be responsible.

Finally, we note that if we restrict the study to those events with sufficient in situ L1 data to provide the required features, the number of events reduces down to 93 (from 94) interplanetary shocks related to SSCs (with prompt storm activity within 4 hr) and a total of 303 (from 307) interplanetary shocks that are related to a detected ground signature (SCs). This marginally reduced catalog will form the database for the remainder of the study.

3. Methods, Models, and Metrics

In this work we demonstrate the use of machine learning models to provide probabilistic estimates as to whether observed interplanetary shocks at L1 will be geo-effective, defined in section 2 as causing an identifiable sudden commencement (SC), and also by whether they will be related to a geomagnetic storm (or SSC). In this section we outline the feature selection methods, the machine learning models and metrics used to evaluate the performance of the models.

3.1. Feature Selection

There are a large number of solar wind and IMF properties obtained at L1 that could possibly be used to describe an interval of data around an interplanetary shock. However, not all of these variables will be useful to the models; there will be some that correlate with each other, and some may be confounding variables (e.g., Bentley et al., 2018). A correlation matrix of the solar wind features is presented and discussed in Appendix A. Additionally, the inclusion of a large number of features may result in overfitting, where the model overlearns from the training examples and then is not able to extrapolate to future or unseen events. For this reason, we evaluate which features of the shock are the most useful to the models.

For this study we have chosen to extract the feature importances using a random forest (ensemble) model. The importance of each feature is a measure of how the Gini impurity or information (entropy) is changed by placing requirements upon that feature. It should be noted that this method can have unexpected results when scoring features that have large differences in scales (e.g., Strobl et al., 2007). To minimize the effect of this on our results, we have chosen to standardize the values of each feature using the mean and standard deviation. Additionally, it should be noted that if two features are correlated then the first selected will have the original importance, while the second will have a significantly reduced importance: Its subsequent addition will not significantly improve the forecast. This is despite the fact that both features may individually be good predictors, as would be expected if they correlate strongly. Therefore, an interpretation should be careful to note that the relative feature importance does not necessarily represent the individual skill of each feature in isolation, but instead the combination of features that best contribute to making classifications.

Two other feature evaluation methods were considered, the F Score and Mutual Information (MI) of the features. However, the F Score only measures linear relationships and does not account for correlations between features (Chen & Lin, 2006) and was therefore found to give poorer results. The MI method provided almost identical results to the ensemble feature importance method, but does not easily provide an uncertainty estimate, and therefore, it was not selected over the chosen method.

3.2. Models

In this study, we develop a series of forecasts using relatively simple machine learning techniques such that these forecasts can be run in near-real-time to forecast the consequences of an interplanetary shock that is observed at the L1 point. All models are available from the scikit-learn python package (Pedregosa et al., 2011). In section 5 we will also consider hypothetical or extreme interplanetary shocks, for example, historical events or those that may be predicted, perhaps through ballistic or MHD heliospheric models.

First, we consider a simple linear model. Fitting this model on a single feature would be the equivalent of scaling the probability of an "event" with the value of the parameter, providing a useful benchmark. The model is based on Logistic Regression, a linear technique where the probability of each outcome is modeled with a logistic function (or sigmoid curve) (Hosmer & Lemeshow, 2005).

We also consider two nonlinear models to examine whether the interplanetary shocks can be better evaluated in this way. First, we test a Gaussian Process classifier, a non parametric model that uses a Bayesian approach, assuming a prior distribution on the underlying probability densities (Rasmussen & Williams, 2006). It has the advantage that it natively provides a probabilistic result; however, it is known to be relatively computationally intensive when applied to high dimensional data sets. The second contrasting nonlinear method tested is a Gaussian Naive Bayes model, based on applying Bayes' theorem while assuming conditional interdependence between features, given the assigned class (Pérez et al., 2006; Webb et al., 2011). In this formulation, the likelihood of the features is assumed to be Gaussian. This model has been included as it has been shown to perform well as a classifier with a relatively small training set. However, the probabilities that it assigns are known to be unreliable (Zhang, 2004), and we therefore re-scale the returned probabilities in order to report better calibrated and reliable forecasts using Platt Scaling (Platt, 1999). Platt

scaling involves fitting a parametric logistic regression model to the output of the model. Essentially, this scaling enables the correction of the initial returned probabilities into a more reliable output. The sklearn package provides this feature as a part of the CalibratedClassifierCV class.

Finally, we consider an ensemble model based on decision trees. Each tree divides the parameter space into "leaves," with each split designed to maximize separation of the distinct classes (Breiman et al., 1984). This method is highly nonlinear and is known to be susceptible to overfitting. Therefore, Random Forests average the results of multiple independently derived "trees," thereby reducing the variance of the model and overfitting (Breiman, 2001; Ho, 1995). Probability estimates are obtained from the ensemble of predictions made from the different trees. To calibrate the returned probabilities we once more perform Platt scaling (Platt, 1999).

The Logistic Regression and Random Forest models (in particular) have several hyper-parameters: internal model parameters used during training that can be tuned to provide superior model performance. In this work we apply a simple grid search method to optimize these parameters and improve model performance. To create a completely optimal form of these models a more detailed optimization of these hyper-parameters could be performed.

3.3. Cross-Validation

We apply a k-fold cross-validation procedure to ensure that the model results are not specific to a selected set of training data and can be generalized. This procedure involves splitting the data into k groups; using $k - 1$ groups to train the model and the remaining group to validate the model results (Kohavi, 1995; Schaffer, 1993; Shao, 1993). Specifically, we apply a stratified k-fold (with four folds) to examine the results of the models. The stratification ensures that the classes are evenly represented in all folds.

More generally, it is good practice to test model performance on a portion of data that was initially withheld from the model (i.e., a validation set). However, in this case we report the results of the k-fold cross-validation due to a limited and imbalanced data set, particularly when considering SSCs. Partitioning a validation set in this case returned metrics that varied considerably between distinct model runs, due to the relatively poor quantity of data and random selection processes. Given our chosen method of reporting, the training and test data are not fully independent, and therefore, the uncertainty in the model metrics is likely to be underestimated.

3.4. Baseline and Metrics

In this study, we assess the performance of the models using two standard probabilistic forecast verification metrics: the Brier Skill Score and the ROC score. These metrics were developed for assessing terrestrial weather forecasts but have also been used in a variety of space weather contexts (Azari et al., 2018; Crown, 2012; Forsyth et al., 2020; Murray et al., 2017).

The Brier Score (BS) is a measure of the mean square of the probability error (Brier, 1950), calculated using Equation 1:

$$BS = \frac{1}{N} \sum_{i=1}^{N} (p_i - a_i)^2 \tag{1}$$

where N is the number of observed events, p_i is the forecast probability (between 0 and 1), and a_i is the observation (1 = occurred, 0 = did not occur). The Brier Score is measured between zero (for a perfect forecast) and one (for a completely incorrect forecast). In this study we are interested in comparing the skill of the model as compared to a baseline prediction, and so we calculate the Brier Skill Score (BSS), which shows the improvement of the BS of the model compared to the BS obtained by a reference prediction (climatology, or overall/total probability for example), calculated using

$$BSS = \frac{BS_{Clim} - BS_{Model}}{BS_{Clim}} \tag{2}$$

where BS_{Clim} is the BS obtained by the climatology and BS_{Model} is the BS obtained by the model under investigation. The BSS will be maximized for a perfect forecast and will equal 1, while negative values indicate that the forecast is worse than simply using the climatological forecast.

Reliability diagrams graphically describe the accuracy of a probabilistic forecast, displaying how well the forecast probability of an event corresponds to the actual chance of observing the event. The frequency of the observations is plotted against the frequency of the forecast probability. A perfectly reliable forecast would lie along the diagonal line of gradient unity; for example, a forecast of 30% would correspond to an observation of the event 30% of the time. The comparison between the diagonal and the reliability curve provides a measure of how reliable the model results are. However, it is worth mentioning that a consistently "unreliable" model can be re-calibrated once its reliability curve has been assessed, as described above (cf. Platt, 1999).

Finally, ROC curves describe the ability of a forecast to discriminate between events and non-events (the skill of the model), something that is not tested by the reliability diagram and the Brier Skill Scores. An ROC curve is a plot of the false alarm rate against the hit rate for a series of forecasts in which the parameter which determines a positive or negative forecast is varied (Swets, 1988). In this study, we set a threshold for the forecast probability and vary this to generate the ROC curve. In theory, as the probability threshold is increased a skillful forecast would show an increasing number of "hits," while the number of false positives would grow more slowly. This curve would lie close to the top left of the plot, such that the hit rate approaches 100% while the false positives remain low. To quantify this behavior the area under the curve may be evaluated, known as the ROC score, and is measured between 0 and 1 (Zweig & Campbell, 1993). A perfect forecast will be described by a point at the top left of the plot, maximizing the hits without incurring any false alarms, and return an ROC score of 1. In contrast, if the probability of an event is random then it would be expected that the false positives would grow at the same rate as the hits, and so a score of 0.5 corresponds to a model with low skill. Therefore, a skillful model will show a ROC score greater than 0.5, ideally approaching 1.

4. Results

In this work we apply the models described in section 3.2 to make probabilistic forecasts as to whether an interplanetary shock will be geoeffective. We define "geoeffective" in two ways, first, does the interplanetary shock generate an independently identifiable ground signature: a Sudden Commencement (SC). Second, does the interplanetary shock precede a geomagnetic storm, that is, a Sudden Storm Commencement (SSC).

4.1. Forecasting SCs

We first test the ability of the models to provide a probabilistic forecast as to whether an SC will be observed on the ground, using information derived from data in the interval around the interplanetary shock. As discussed in section 2, out of the 547 interplanetary shocks in the data set, 307 result in an independently identified ground signature. Excluding those events for which insufficient solar wind data are available limits the data set to 540 interplanetary shocks, 303 (56%) of which are related to SCs. Therefore, our climatological forecast is 56%, providing a Brier Score of 0.25.

As discussed in section 3.1, we use a random forest classifier to rank the importance of each of the 32 extracted solar wind features. Figure 1 shows the relative importance of the top 10 features, normalized to the most important, where the uncertainty is the standard deviation from the ensemble of estimators. The most powerful predictive parameter can be seen to be the range of the magnetic field magnitude, while the ranges in density and velocity (X_{GSM} component) are the second and third. The top few features are shown to score highly, while the feature importance quickly drops beyond this. It should be noted that if the range of the solar wind dynamic pressure is included then this becomes the second most important parameter, with a similar importance to the range in density (n_p) shown in Figure 1.

We wish to optimize the number of features that we provide the models, we do this by adding each of the top 10 features, in order of their importance from Figure 1. Figure 2 shows how the addition of features changes the Brier Skill Scores and the ROC scores for each of the different models. All models initially show an increase in both metrics (representing reliability and skill respectively) following the addition of features, however the benefit of additional parameters is seen to plateau at around three features.

Figure 3 shows the results of the models when provided with the top three features. Figure 3a shows a reliability diagram, presenting how the forecast probabilities correspond to the observations. As discussed in section 3.4, a perfect probabilistic forecast will lie along the diagonal dotted line of gradient unity.
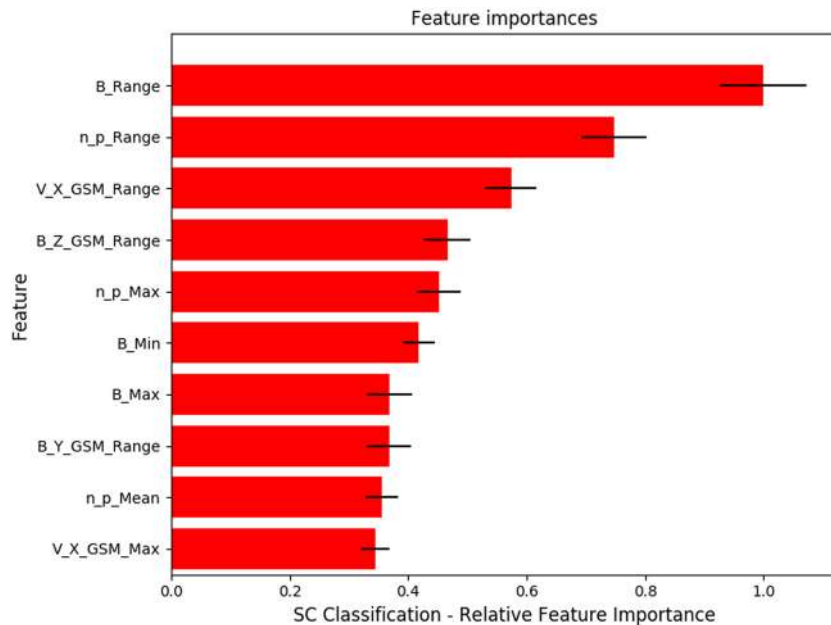
Figure 1. The relative importance of the top 10 features in forecasting an SC, as ranked by a random forest classifier. The uncertainties shown are the standard deviation of the importances returned by the ensemble of estimators.

The horizontal and vertical dashed lines indicate where climatological forecasts would lie. All four models can be seen to fairly well correspond to the (diagonal) perfectly reliable forecast, and this is further quantified in Figure 2d, where the Brier Skill Scores (BSSs) are presented, compared to climatology. Also included is the BSS obtained by a simple linear 1-dimensional Logistic Regression model, equivalent to scaling the probability with the most important parameter: the range in magnetic field magnitude in this case. All four models return very good BSSs, outperforming both climatology and the 1-D Logistic Regression model, returning BSSs between 0.26 and 0.32. The Gaussian Process model is shown to provide the most reliable results, being the only model to consistently score above 0.3. For context, as the climatological Brier Score is found to be 0.25, a model BSS of 0.3 corresponds to a Brier Score of 0.175. Figure 3b analyzes the ROC curves, showing the false positive rate plotted as a function of true positive rate (described in section 3.4). The line of no skill is presented as a dashed diagonal line. All four models show ROC scores between 0.8 and 0.83, representing skillful forecasts. The worst performing models in terms of both BSS and ROC scores are found to be the Naive Bayes and Random Forest models, perhaps due to the specifics of their non-linear methods as well as the requirement (and only partial success) of the process to re-calibrate their probabilities (discussed in section 3.2). The Random Forest model in particular may be showing a tendency to over-fit to the training data.

4.2. Forecasting SSCs

Building on the above, we assess the ability of the selected models to provide a probabilistic forecast of whether an observed interplanetary shock will be followed by a geomagnetic storm, that is, a geomagnetic storm is observed within 4 hr of shock impact (cf. Fiori et al., 2014). The dataset for this totals 540 interplanetary shocks, of which 93 (17%) were followed by a geomagnetic storms. The climatological forecast is therefore 17%, which returns a baseline Brier Score of 0.14.

As above, we evaluate the 32 solar wind features extracted from the interval around the shock observation using a random forest classifier. The relative importance of the top 10 features is presented in Figure 4. While the range of the field (B_{Range}) still ranks highly, in contrast to Figure 1 we see that the most important parameter is now the minimum B_Z observed in the interval around the shock. The change in velocity ranks as the third most important in forecasting whether a shock will be related to an SSC. In addition, if the range in dynamic pressure is also included, it ranks as the seventh most important parameter, though this should be noted with the caveat that it does correlate with other features ranked as highly as third most important.
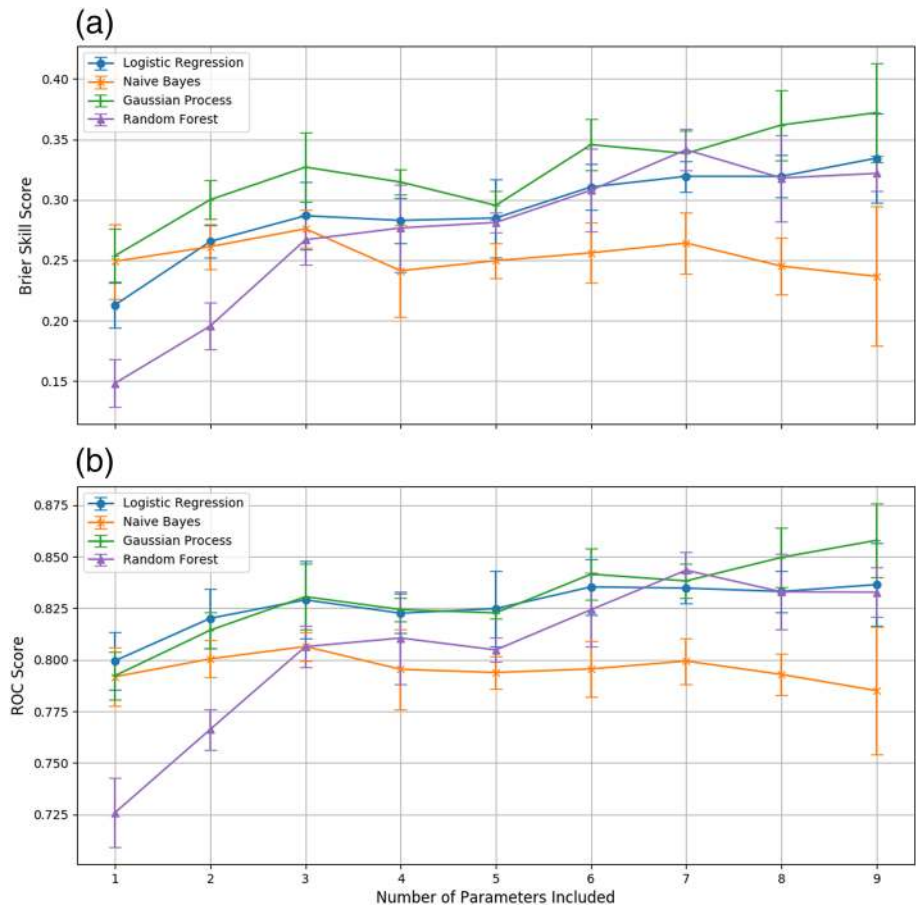
Figure 2. The variation of the Brier Skill Score (a) and ROC scores (b) returned by the selected models when predicting SCs, as a function of the number of input parameters. The error bars indicate the standard error of the metrics across the fourfold validation regime.

Figure 5 shows how the addition of parameters in the order suggested by the random forest feature importances (Figure 4) changes the BSS and ROC score. As with the forecasting of SCs above, the inclusion of a second feature increases the skill of the models; however, the addition of more features is not as beneficial, and the scores can be lower. The variations between the folds can be seen to be quite substantial (from the error bars), which is likely a result of the relatively small number of positive events with which it is possible to train the models.

Figure 6 shows the results of the models when provided with the top 4 features, approximately maximizing the BSSs and ROC scores from Figure 5 for most models. The relatively small number of positive events can be clearly seen in Figure 6c, where the histogram of predicted probabilities is strongly dominated by low values. However, the reliability of the predictions, displayed in Figure 6a and quantitatively assessed in Figure 6d, can still be seen to clearly outperform climatology. The ROC plots and scores presented in Figure 6b also show good scores, above 0.8 for three of the models.

Interestingly in Figure 6d, and in contrast to the SC forecasting in Figure 3d, the addition of more parameters for several of the methods does not seem to provide a strong improvement over the use of a single parameter and a linear method (1-D LR). This could be due to higher dimensional models overfitting the relatively sparse data available. This is particularly the case for the highly nonlinear Random Forest model, which can be seen to give comparatively low BSS and ROC scores. However, it should be noted that in Figure 5 the Random Forest model does appear to benefit more than the other models from the inclusion of additional features. The best models in terms of both ROC scores and BSSs were the Linear Regression and Gaussian
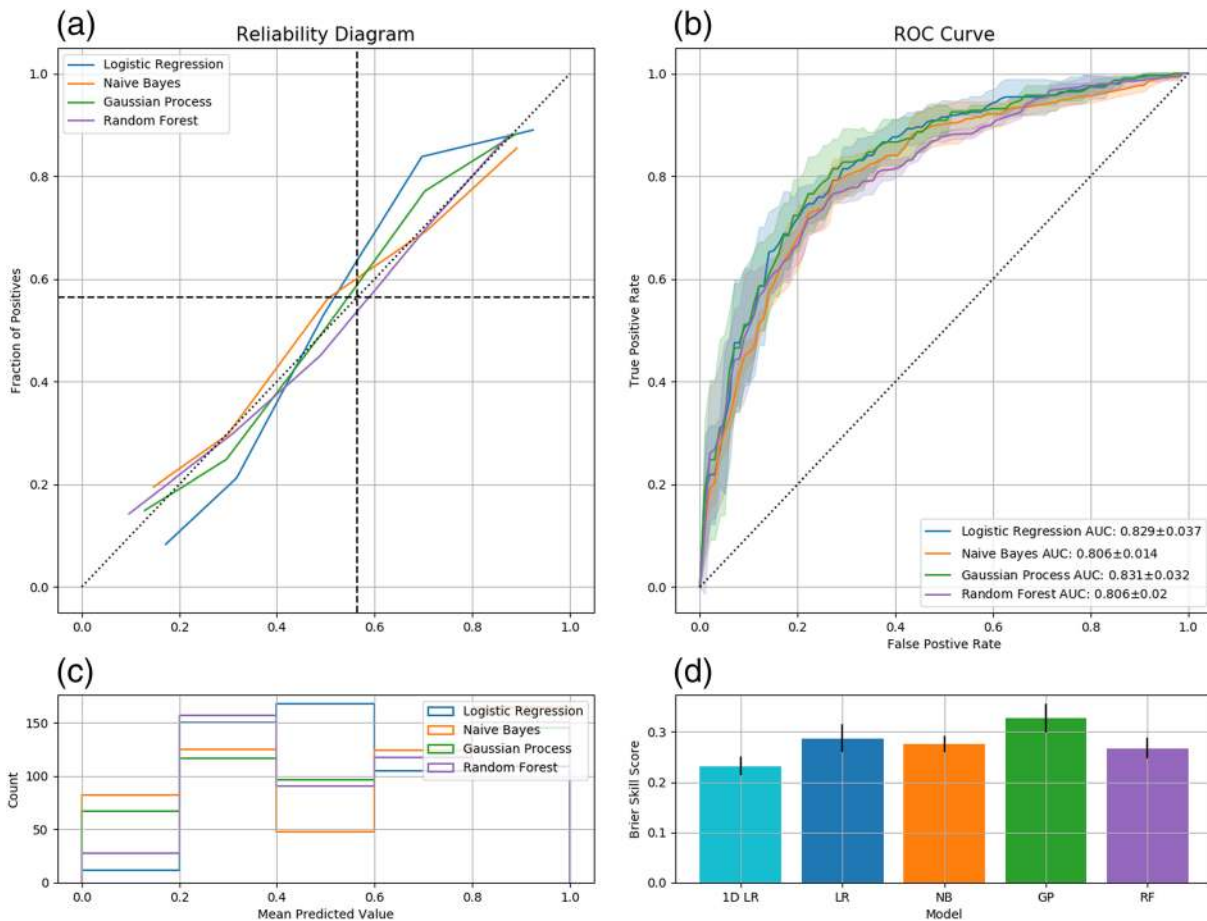
Figure 3. Sudden Commencement forecasting results. The reliability curves (a), ROC curves (b), distribution of predictions (c), and Brier Skill Scores (d) returned by the selected models when provided with the top three features. The uncertainties in panels b and d are extracted from the four fold cross-validation procedure and represent the standard deviation and standard error, respectively. The Brier Skill Scores are shown for the 1-D Logistic Regression (1-D LR), Logistic Regression (LR), Naive Bayes (NB), Gaussian Processes (GP), and Random Forest (RF) models.

Process models with as little as two parameters, with ROC scores and BSSs exceeding 0.82 and 0.21, respectively (e.g., Figure 5). The climatological Brier score was 0.14 for this configuration, and so a BSS of 0.21 represents a Brier Score of 0.11.

5. Discussion

Our results show that simple machine learning techniques can provide models that skillfully and reliably forecast the occurrence of both SCs and SSCs from the properties of interplanetary shocks. However, it is also useful to consider the performance of these models when the available input data is limited or when the features used exceed the limits of the training dataset.

5.1. Most Powerful Predictive Parameters

Figures 1 and 4 evaluated the importance of each of the 32 extracted solar wind features around the shock identifications; we will now discuss and interpret the results in their physical context.

5.1.1. Sudden Commencements

The northward deflections of the horizontal geomagnetic field from which sudden commencements are identified are often modeled as a combination of two main components: a step-like function at low latitudes and a two-pulse structure that dominates at higher latitudes (Araki, 1977, 1994). Observations at low latitudes have found that the magnitude of the low latitude perturbation scales with the square root of the change in solar wind dynamic pressure (e.g., Russell et al., 1992). The SCs used in this study were identified from a series of low latitude magnetic observatories, and it might therefore be expected that the most pow-
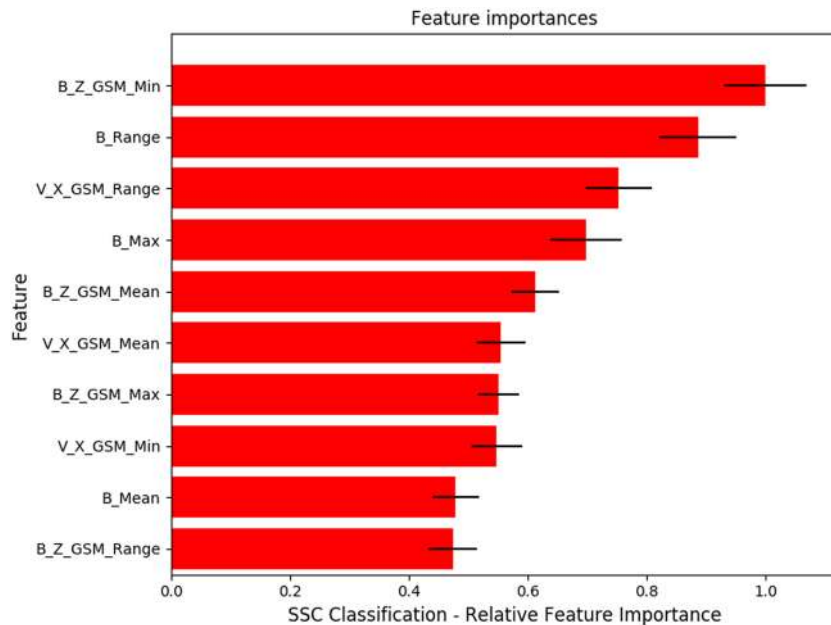
Figure 4. The relative importance of the top 10 features in forecasting an SSC, as ranked by a random forest classifier. The uncertainties shown are the standard deviation of the importances returned by the ensemble of estimators.

erful predictive features would be the features that constitute the change in dynamic pressure: When the range in pressure is larger, it would be expected to generate a more significant ground signature that may be more likely to be identified. However, the ranges of density and velocity are only the second and third most powerful features in Figure 1. Indeed even when explicitly included, the range in dynamic pressure only ranks as the second most important parameter.

The most powerful feature is found to be the range in $|B|$. It is possible that the range in $|B|$ may serve to distinguish between the phenomenon that is driving the shock. For example, it may be that CME-driven shocks more often display large changes in $|B|$, compared to those driven by CIRs, while the effect of the distinct phenomena at the Earth is known to vary (Oliveira & Samsonov 2018; Richter et al., 1985; Smith & Wolfe, 1976; Tsurutani et al., 2006). Another consideration is the combination of parameters that may be used to define a shock. In the frame of the shock, the ratio of upstream and downstream field, velocity, and density should provide the necessary information to evaluate the size of each shock (Hugoniot, 1887, 1889; Rankine, 1870). As the values provided to the algorithms are not in the shock frame, it may be that the change in velocity is not as good a descriptor of the shock as the field and density changes. This may also explain why the addition of parameters beyond the first few is ineffective, once the shock is satisfactorily defined, little information of additional value can be included. From a physical perspective, however, the appearance of the range in B_Z as the fourth most important parameter may indicate that it is effective in discriminating between perpendicular and parallel shocks, which may be important (e.g., Jurac et al., 2002). Nevertheless, the fact that the change in B ranks as the most important parameter highlights the importance of forecasting the nature of the magnetic field upstream of the Earth.

It is also notable that the ranges in the solar wind parameters are selected as important and not the maximum values. This may be explained by the close relationship between the range of a parameter and its ratio, which would define the size of the shock (e.g., Hugoniot, 1887, 1889; Rankine, 1870). Additionally, this could correspond to observations that the nature of the upstream solar wind into which the shock is propagating is also important (e.g., Liu et al., 2014; Riley et al., 1997), as the ranges of these parameters would effectively distinguish the greater geoeffectiveness of a shock propagating into a more tenuous region.

5.1.2. Storm Sudden Commencements

One of the main differences between the feature importances when forecasting SCs and SSCs (Figures 1 and 4) is the dominance of the minimum value of B_Z when considering SSCs. It has long been thought that the southward component of the IMF is strongly associated with ground magnetic disturbances (e.g., Fairfield & Cahill, 1966). Notably, Echer et al. (2008) found that all 90 intense geomagnetic storms ($Dst < -100$
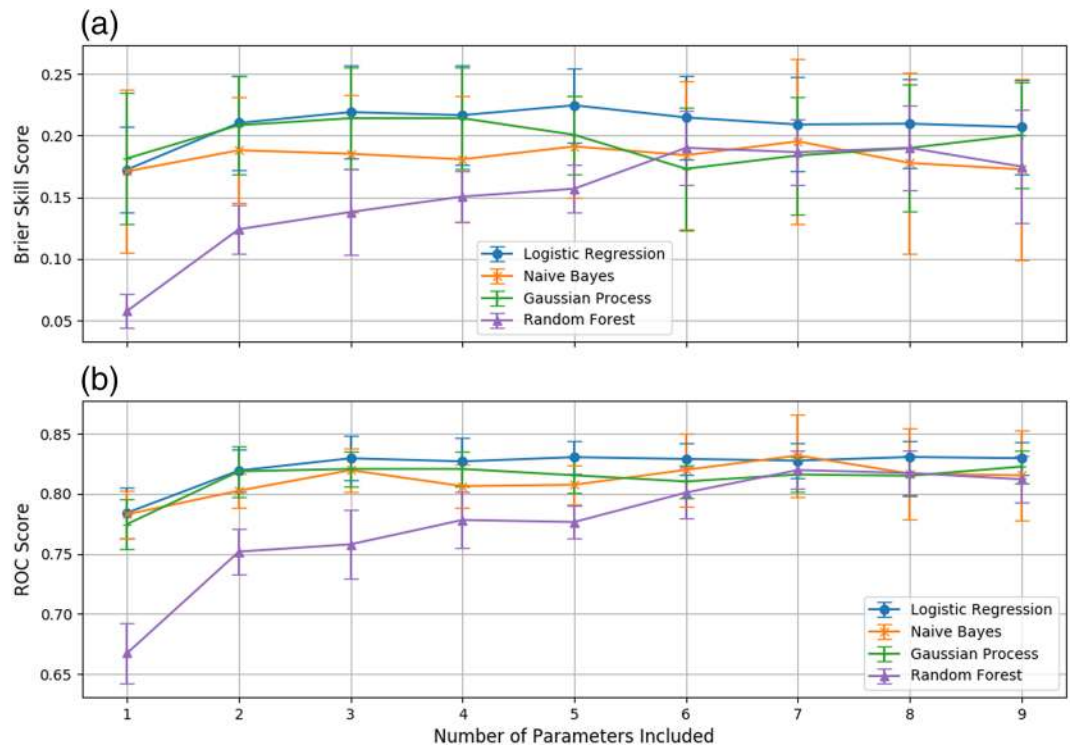
Figure 5. The variation of the Brier Skill Score (a) and ROC scores (b) returned by the selected models when predicting SSCs, as a function of the number of input parameters The error bars indicate the standard error of the metrics across the four fold validation regime.

nT) within their dataset were associated with strong southward B_Z. Therefore, it was expected that it would be a significant parameter. Additionally, strong correlations have been observed between the Dst index and B_Z (e.g., Burton et al., 1975; Temerin & Li, 2002) or the related dawn-dusk electric field (e.g., Ontiveros & Gonzalez-Esparza, 2010; Wang et al., 2003). This work again highlights the importance of forecasting the orientation of the magnetic field.

5.2. Parameterization

The extraction of features from the solar wind around the shock identification is deliberately simple to minimize the effect of missing and incomplete data. However, it is likely that more complex fitting or parameter extraction would increase the effectiveness of the models. For example, recent work has shown the importance of the orientation of the shock front, both through the use of MHD models (Oliveira & Raeder, 2014) and also direct observations (Oliveira & Raeder, 2015).

While more complex characterization of the shock itself may help, obtaining parameters related to the phenomena driving the front may also aid prediction. At the longest lead time, remote observations of CMEs have been shown to be useful for inferring future consequences (Kim et al., 2010), especially when combined with empirical limits on conditions in near-Earth space (Kim et al., 2014). More detailed in situ analysis of CMEs, such as fitting of their structure, has also been shown to be useful (e.g., Kang et al., 2006). In addition, the properties of the sheath between the CME and the shock have been found to be important to determine the magnitude of the interaction with the Earth's magnetosphere (Kilpua et al., 2019). The sheath itself has been inferred to drive 25–50% of intense geomagnetic storms (Richardson et al., 2001; Tsurutani et al., 1988), through several mechanisms (Lugaz et al., 2016). It should be noted that the chosen window of data, 1 hr prior to shock impact, will include some sheath observations. However, with our method the most powerful predictive features are dominated by the ranges in solar wind parameters, and therefore by the shock itself (with some exceptions, for example, the minimum B_Z).

Correlations have been noted between the different regions of the interplanetary phenomena, which may aid the forecasting of SSCs using the methods in this study. For example, correlations have been observed between the properties of CME sheaths and their shocks (Lindsay et al., 1994; Tsurutani et al., 1988), as well
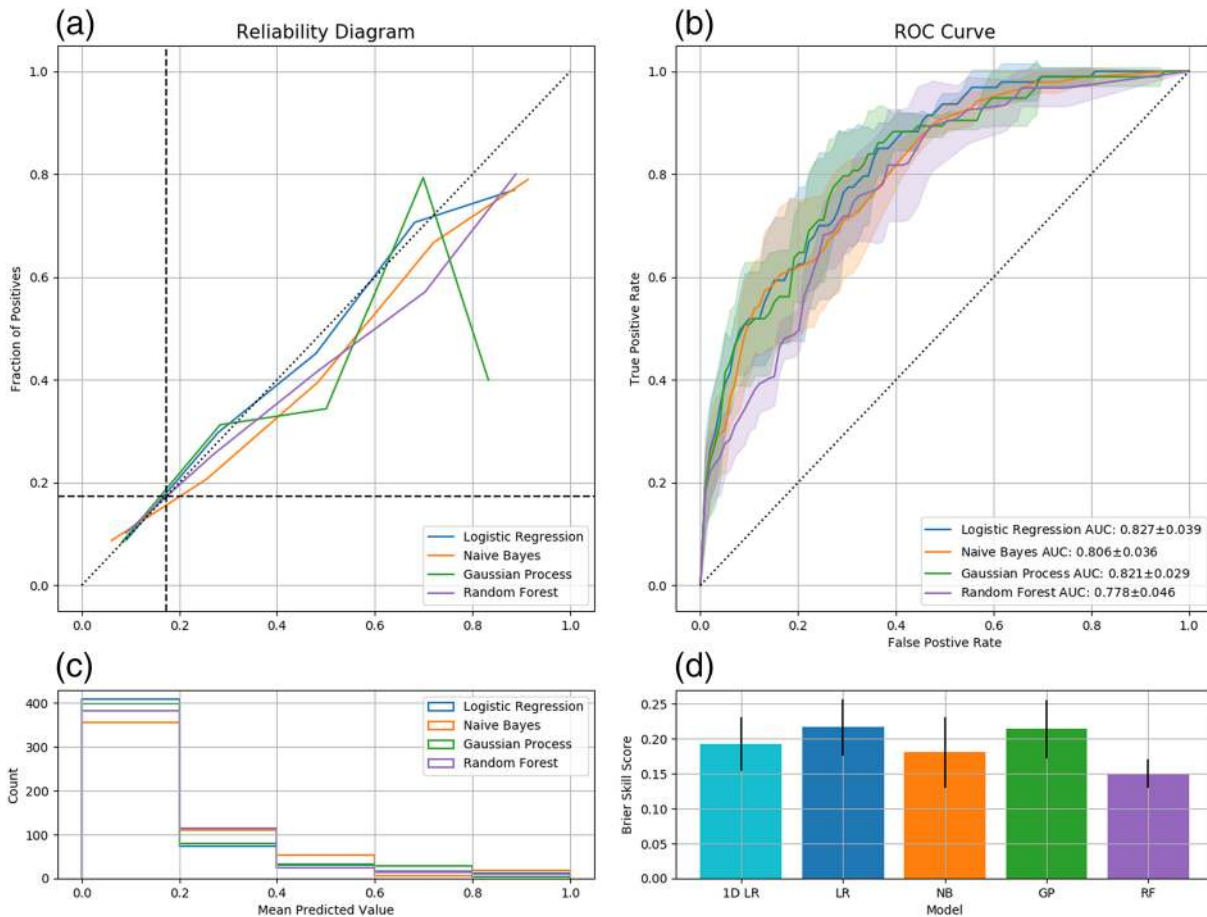
Figure 6. Storm Sudden Commencement forecasting results. The reliability curves (a), ROC curves (b), distribution of predictions (c), and Brier Skill Scores (d) returned by the selected models when provided with the top 4 features. The uncertainties in panels b and d are extracted from the fourfold cross-validation procedure and represent the standard deviation and standard error, respectively. The Brier Skill Scores are shown for the 1-D Logistic Regression (1-D LR), Logistic Regression (LR), Naive Bayes (NB), Gaussian Processes (GP), and Random Forest (RF) models.

as the shocks and the magnetic structure that follows (Lepping et al., 2001; Luhmann, 1997). Additionally, a 30-min interval around the shock has been suggested to be useful in predicting the occurrence of long duration southward B_Z (Salman et al., 2018). Therefore, though information about the larger scale phenomenon driving the shock has not been included directly, it may correlate with the features provided.

5.3. Magnetospheric Information

While the results in Figure 6 show good skill at forecasting SSCs compared to climatology, there is not a large increase with the addition of parameters (i.e., Figure 5), perhaps as all relevant information about the shock has already been included. In fact, the Brier Skill Scores reported for the models mostly do not significantly improve on the benchmark of a 1-D Logistic Regression model (Figure 6d). There are several reasons additional factors that could explain this behavior. First, the extent of the catalogs may not be extensive enough to provide adequate coverage for multidimensional, nonlinear models. Second, it is possible that the models need to include some indication as to the state of the magnetosphere, as this may have a strong influence. Third, it is also possible that the method by which the features have been extracted from the solar wind around the shock impact are insufficient to quantify the nature of the coupling to the geomagnetic field, both initially during the shock and also for any structure that follows (e.g., coronal mass ejection, sheath region, etc.).

To investigate the second and third points, we can explore adding information about geomagnetic indices, providing an estimate of the current magnetospheric conditions and to see if the coupling is adequately captured with the properties extracted from the solar wind data. Similar to the treatment of the solar wind parameters, we extract the minimum, maximum, range, and mean of several magnetospheric indices: AU,
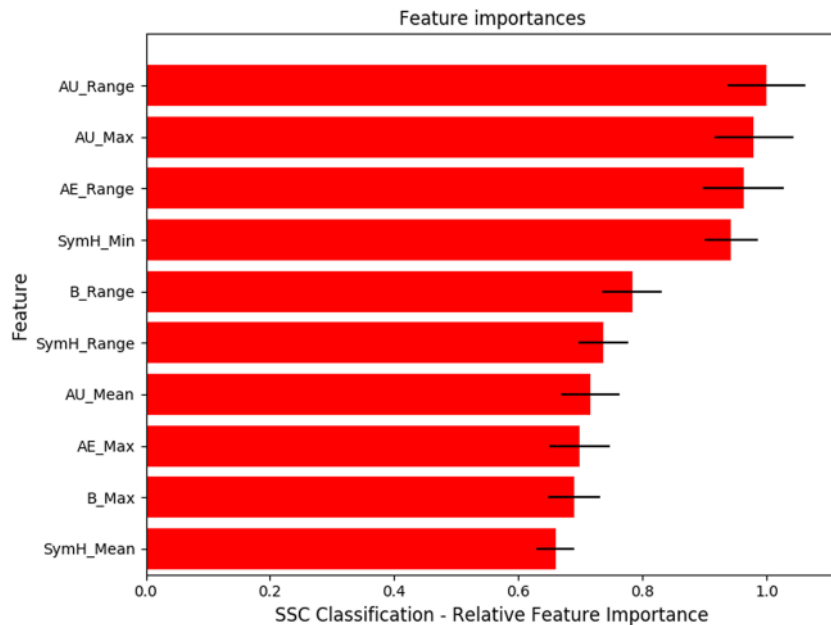
Figure 7. The relative importance of the top 10 features in forecasting an SSC (including magnetospheric indices), as ranked by a random forest classifier. The uncertainties shown are the standard deviation of the importances returned by the ensemble of estimators.

AL, AE, Sym-H, and Sym-D in a 1 hr window prior to shock impact, extending to 15 min after the initial impact. This does mean that we are including any potential SC signature, and therefore, we only test the forecasts of SSCs. Figure 7 shows the top 10 parameters from the total of 52 possible features.

It is notable from Figure 7 that the top 4 parameters are all related to the magnetospheric indices, with most corresponding to the initial response of the magnetosphere (i.e., the range of several indices). This would suggest that the simple shock parameterization that has been employed has not captured some of the important information relating to the interaction of the shock and magnetosphere. However, as noted in section 3.1, the fact that the feature representing the minimum B_Z is not present (compared with Figure 4) does not mean it is no longer important at all. It is likely that while the magnetospheric indices are more powerful predictive parameters, they correlate strongly with the minimum B_Z and therefore reduce the returned importance of the B_Z feature.

Figure 8 shows the performance of the models, when provided with the top 5 parameters displayed in Figure 7. The top 5 have been used as this was the point at which no significant additional skill was obtained by adding more features. All four models are shown to once more provide reliable forecasts compared to climatology as shown by the reliability diagram and BSSs (Figures 8a and 8d). The models are also shown to provide good skill, with ROC scores above 0.8. The increase in reliability achieved through the addition of magnetospheric indices amounts to modest increase in BSSs scores of up to 0.05, while the increases in ROC scores are also relatively small at ≤0.02, and not present for all models considered. Therefore, we may conclude that the coupling of the interplanetary structure to the magnetosphere is not completely captured by the simple parameterization employed. The presence and significance of the minimum Sym-H index also suggests that the state of the magnetosphere and ring current is important. However, only a relatively small increase in skill can be achieved through addition of magnetospheric indices.

5.4. Longer Lead Times

When forecasting shocks and associated phenomena at large lead times, for example, from heliospheric imaging, there are certain parameters that may be more accurately predicted. For example, it may be possible to more accurately forecast the velocity (Barnard et al., 2019; Byrne et al., 2010; Davies et al., 2012, 2013; Kahler & Webb, 2007) and density (Barnes, 2020) of a propagating CME, while its internal structure and magnetic field are much more challenging (e.g., Kilpua et al., 2019). Recent work has also confirmed
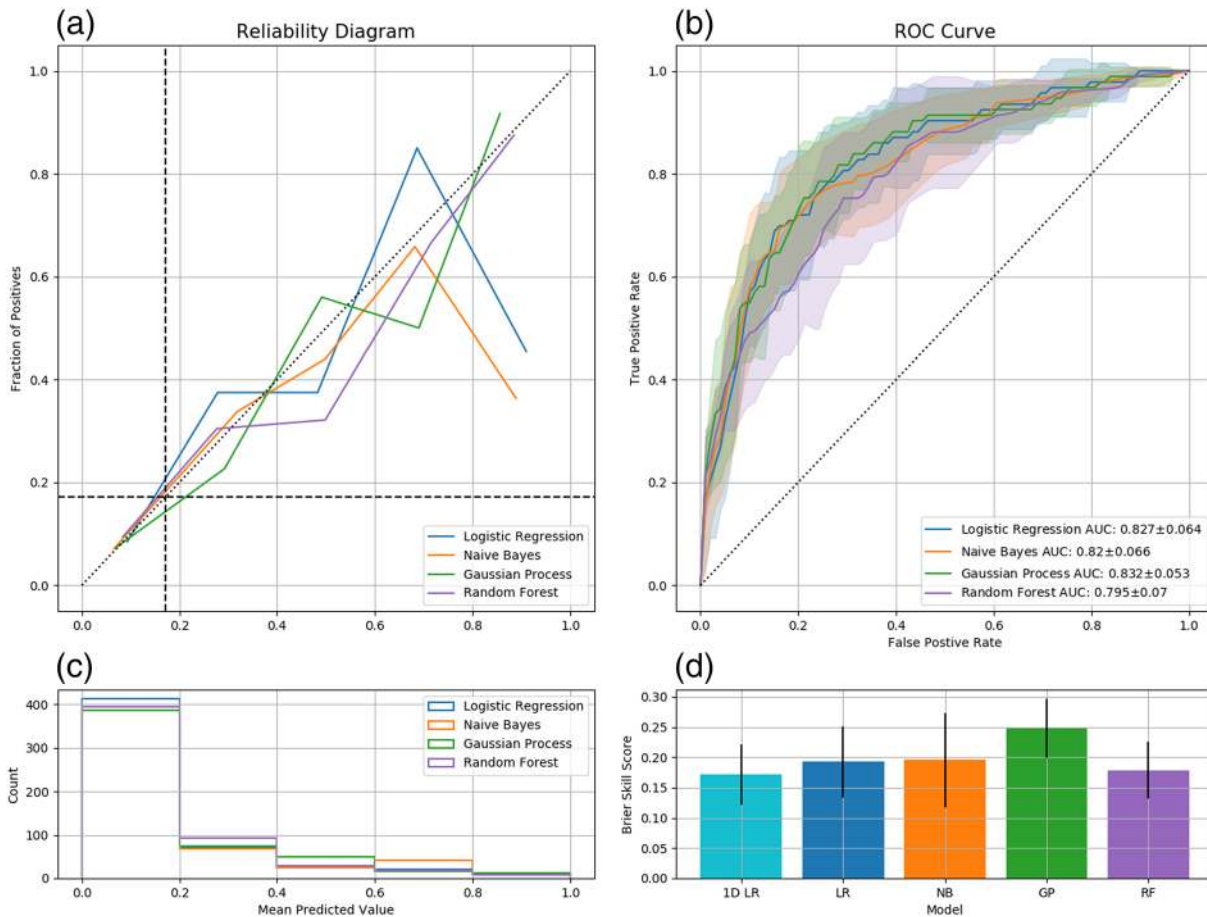
Figure 8. Storm Sudden Commencement forecasting results when magnetospheric indices are included. The reliability curves (a), ROC curves (b), distribution of predictions (c), and Brier Skill Scores (d) returned by the selected models when provided with the top 5 features. The uncertainties in panels b and d are extracted from the fourfold cross-validation procedure and represent the standard deviation and standard error, respectively. The Brier Skill Scores are shown for the 1-D Logistic Regression (1-D LR), Logistic Regression (LR), Naive Bayes (NB), Gaussian Processes (GP), and Random Forest (RF) models.

that accurately forecasting CME velocity provides useful information about their geoeffectiveness, greatly increasing the value of such forecasts (Owens et al., 2020).

It is therefore useful to assess how the models perform when provided with a more limited dataset, corresponding to the parameters than may be predicted with a greater accuracy. For this test, we limit the input features to those related to the velocity and density observed at L1. Figure 9 shows the relative importances of each feature, evaluated once more with the ensemble method. Figure 9 shows that features associated with the velocity appear to provide the greatest predictive power.

However, Figure 10 shows the change in metrics with the addition of multiple features, and we can see that including more than one feature does not aid the majority of the models. For the ensemble (RF, Random Forest) model including up to five features increases the skill, but in general, both ROC and BSS metrics are lower than achieved for the other models. The Gaussian Process and Logistic Regression models give the best BSSs ($\sim$0.16) and ROC scores ($\sim$0.8) with around five features and outperform climatological forecasts. These scores are substantially lower than were achieved with the inclusion of parameters related to the field. This highlights the need to understand and be able to predict some of the magnetic field parameters associated with an incoming interplanetary shock and related structures (e.g., Gosling et al., 1991; Huttunen et al., 2005; Richardson & Cane, 2012; Zhang et al., 2007).

5.5. Extreme Events

The most damaging Space Weather events are those that are extreme; they pose the largest risk to infrastructure and have the potential to have the largest societal consequences. Arguably, the most extreme event
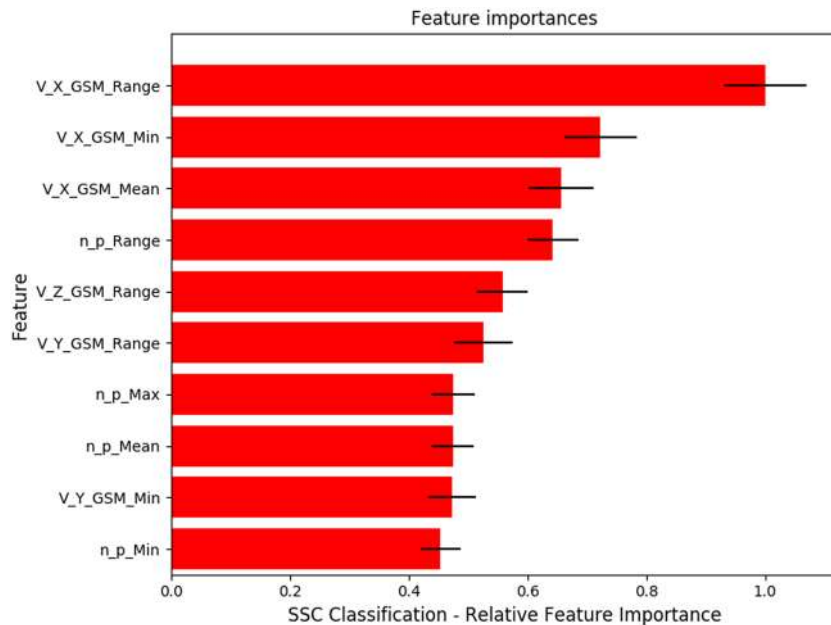
Figure 9. The relative importance of the top 10 features in forecasting SSCs when the features are limited, as ranked by a random forest classifier. The uncertainties shown are the standard deviation of the importances returned by the ensemble of estimators.

recorded in history was the Carrington event in 1859 (Carrington, 1859; Hayakawa et al., 2019; Tsurutani et al., 2003). MHD modeling has suggested that the ground electric fields during this event would be twice as large as during the most extreme geomagnetic storm in the modern era (the March 1989 storm) (Ngwira et al., 2014). In the space age, the fastest CME ever recorded was observed in August 1972 (Hoffman et al., 1975), causing the unexpected detonation of sea mines during the Vietnam war (Knipp et al., 2018). More recently, an interplanetary CME inferred to be similar to the Carrington event, potentially providing a worse-case scenario, was fortunately not Earth-directed but was observed by the STEREO spacecraft in July 2012 (Baker et al., 2013; Ngwira et al., 2013). It is worth noting that the upstream conditions ahead of this CME were highly atypical (Russell et al., 2013). As such CMEs are so rare, and their parameters so unusual, it is interesting to consider how the models presented in this work would react to such an input.

To test this, we can explore how the models react to data provided above the range of the training data. We initially limit the models to the top two most effective features (shown in Figure 4) in order to simplify the visualization. The models are then provided with a grid of hypothetical events, ranging between the minimum of each feature and twice the maximum (calculated for $-B_Z$ minimum). Figure 11 shows the distribution of predictions made by the four different models. The orange contours show the distribution of training data corresponding to the shocks resulting in SSCs, while the red contours shows those that were not related to SSCs. The contours were created using kernel density estimation (Scott, 2015). The red contours representing "non-geoeffective" shocks can be seen to be mostly tightly bunched in the lower right of the panels. Meanwhile, the orange contours representing shocks related to SSCs can be seen to be broader and extend towards larger ranges in *B*.

The Logistic Regression and Naive Bayes models (Figures 11a and 11b) can be seen to asymptote to high probabilities outside of the training region. In contrast, the Gaussian Process and Random Forest models (Figures 11c and 11d) can be seen to have more structured behavior outside of this region, and both tend to be more conservative in their predictions. Without training data at the more extreme values it is not clear which behavior would correspond to better predictions, however, it is very important to take the models response into account when reporting predictions.

It is also notable that for very fast CMEs, for example, the August 1972 event (Hoffman et al., 1975), the transit time from L1 to the Earth's magnetopause will be relatively short. Therefore, the 1 hr window prior to impact will encompass a larger than normal portion of the CME sheath, perhaps effecting the derivation
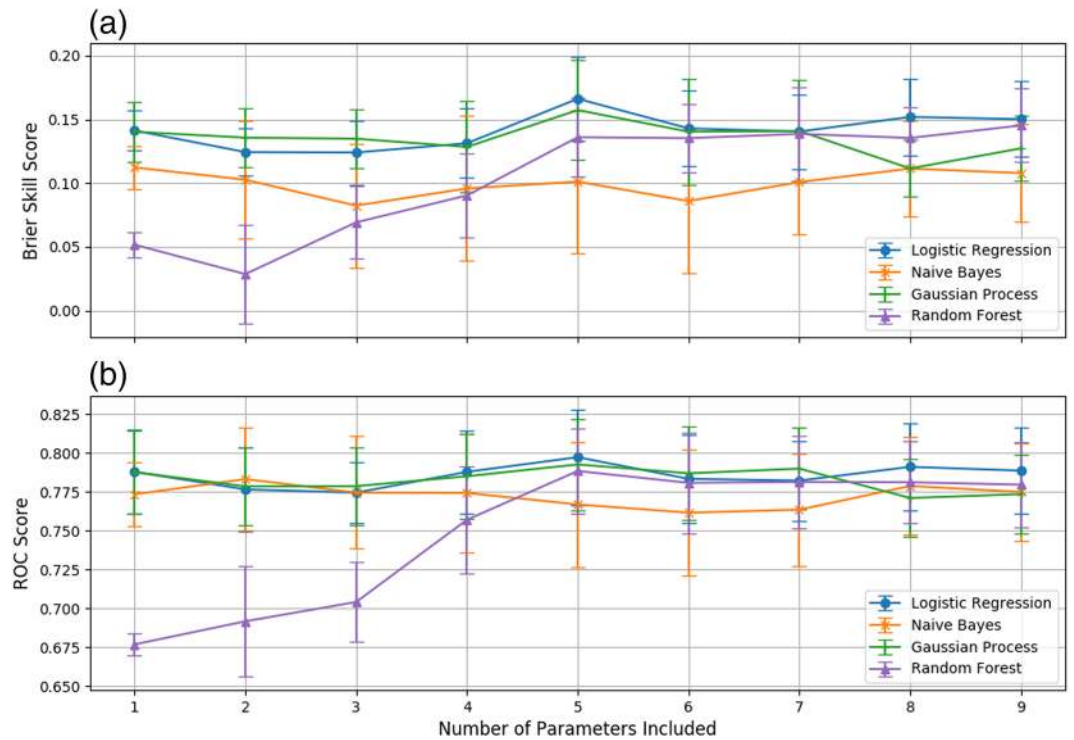
Figure 10. The variation of the Brier Skill Score (a) and ROC scores (b) returned by the selected models when the features are limited, as a function of the number of input parameters The error bars indicate the standard error of the metrics across the four fold validation regime.

of the features used in this method. The consequences of this would have to be be carefully evaluated for any potential operational implementation.

6. Summary and Conclusions

We have developed four different models to forecast the consequences of interplanetary shocks on the magnetosphere using simple machine learning techniques. These forecasts perform well at predicting whether the shock will lead to an SC or SSC. We tested a linear model (Logistic Regression), two nonlinear models (Naive Bayes and Gaussian Process), and an ensemble model (Random Forest). The interplanetary shocks are simply parameterized by their maximum, minimum, range, and mean of observable solar wind and IMF properties at L1. This scheme was chosen as it is resilient to missing data and does not require manual or complex processing to be performed.

All four models provided skillful and reliable forecasts as to whether a shock would result in the identification of an SC on the ground, with Brier Skill Scores (BSSs) of between 0.26 and 0.32 and ROC scores between 0.8 and 0.83. This outperforms both climatology, which provided the reference to which the BSSs were calculated, and a simple 1-D linear model. Overall, the Gaussian Process model returned the greatest reliability, while both the Gaussian Process and Logistic Regression models showed the best skill. Meanwhile, the ensemble method showed indications that it may tend to over fit the available data. The most powerful predictive power was found to be the range in the magnetic field observed, followed by the ranges of density and velocity. This highlights the importance of forecasting the magnetic field upstream of the Earth. Physically it is possible that the range in the magnetic field serves to distinguish between the phenomena driving the shock, or that it provides a more robust definition of the shock than the change in velocity or density. The fact the ranges of each parameter were shown to provide greater performance (rather than the maximums, for example) confirms that the conditions into which the shock are propagating are important.

When forecasting whether a shock will be related to an SSC all four models provided skillful results (with ROC scores exceeding ~0.78), while significantly outperformed climatology (with BSSs of ~0.21). However, when presented with multiple features, the reliability of all four models were comparable to that obtained
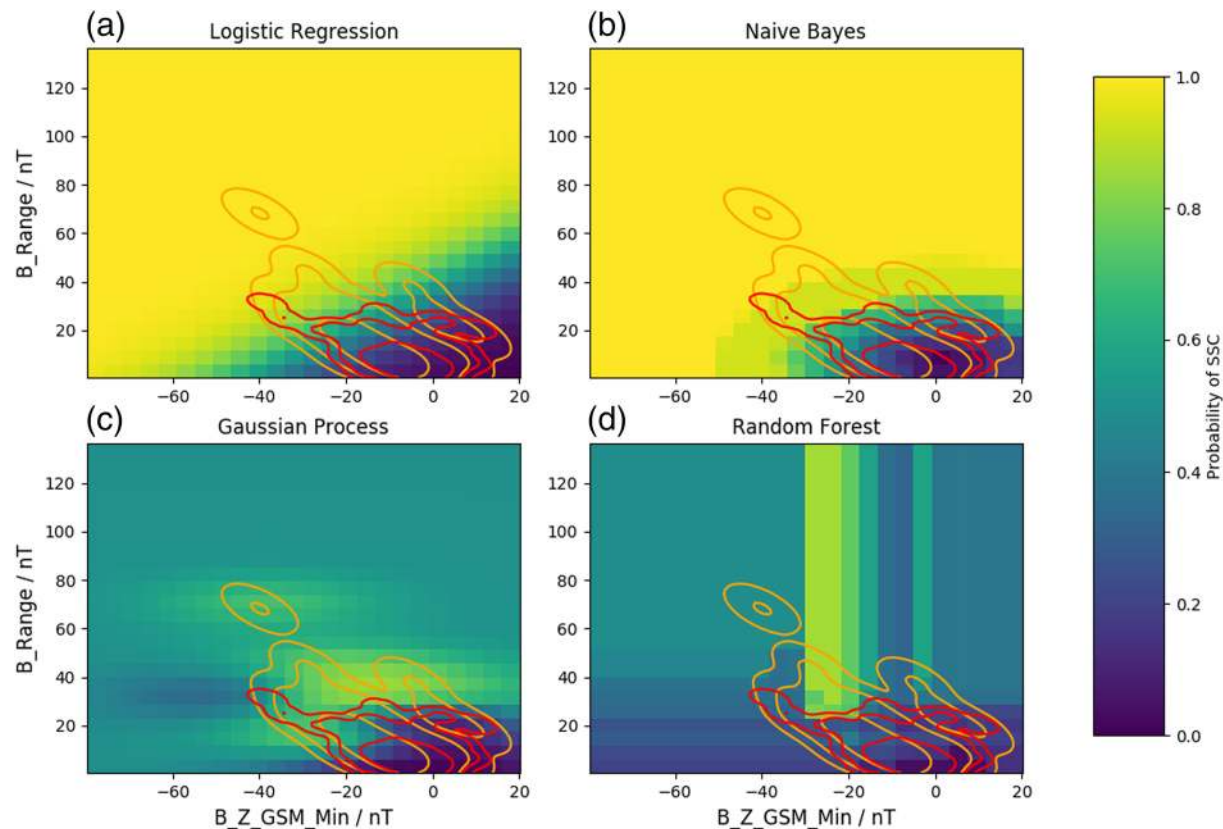
Figure 11. Two dimensional visualizations of the predictions made by the (a) Logistic Regression, (b) Naive Bayes, (c) Gaussian Process, and (d) Random Forest models. The red and orange contours show the distribution of negative (for which no storm was observed) and positive (for which a storm was observed within 4 hr) training data, respectively.

with a more simple one-dimensional Logistic Regression model (BSS ∼ 0.19). Once more, the Logistic Regression and Gaussian Process models were shown to provide slightly greater skill and reliability than the other two models, with the ensemble Random Forest requiring more parameters to achieve comparable performance. The most powerful predictive parameter for SSCs was shown to be the minimum value of B_Z associated with the shock, confirming the importance of the orientation of the IMF when evaluating the effectiveness of a shock.

The chosen parameterization scheme was then tested by the inclusion of magnetospheric indices and their properties, describing the state of the magnetospheric system as well as the initial impact of the shock and sheath. These added features were found to give moderate improvements to the reliability of the models, suggesting that the simple parameterization does not fully capture the coupling of the solar wind and magnetosphere and that an indication of the magnetospheric state is important.

For longer lead times, properties of the magnetic field around interplanetary shocks are challenging to predict, we therefore tested excluding these features from the models results. This more limited data set resulted in decreased predictive performance. However, when solely relying on properties associated with the velocity and density, all models still outperform climatology and provide reliable and skillful predictions (BSSs ∼ 0.16, ROC scores of 0.8). Once more, the Logistic Regression and Gaussian Process models were shown to provide the best relative reliability and skill.

Finally, the models were provided with extreme hypothetical data with conditions beyond that with which they were trained. The responses of the different models were shown and contrasted. Both the Logistic Regression and Naive Bayes models were shown to quickly asymptote to high probabilities outside of the training data, while the Gaussian Process and Random Forest showed much more structured behavior that was heavily dependent on the data at the edge of the training space. The extrapolation of each model therefore needs to be carefully considered when applying such models to extreme data.

Appendix A: Feature Correlations

Many of the properties of the solar wind will be strongly correlated, particularly around shocks, and this should be carefully considered when performing feature selection or evaluating the relative importance of each feature. Figure A1 shows a correlation matrix of the 32 extracted solar wind parameters. The main correlations observed are between the different statistical values (e.g., the range and maximum) of the same property of the solar wind, thereby appearing as squares of similar shades of deep red (or deep blue) along the diagonal. We can also see stronger relative correlations (positive or negative) between the individual magnetic field parameters than for the velocity components. The highest correlation between any density feature and any non-density feature is found between the range in density and the range in the field (B), as may be expected for shocks (cf. Hugoniot, 1887, 1889; Rankine, 1870). Similarly, the correlation between the range in the velocity (V_X) and the range in the field (B) is also notably high. These correlations confirm that the intervals of data likely contain observations of fast forward shocks.

It should be noted that this correlation matrix was produced on the features after they were scaled by the mean and standard deviation of each parameter; this preparation is important and is also applied when the features are supplied as inputs to the models and feature selection methods.

Figure A1. The correlation matrix of the 32 features extracted from the solar wind.

References

Akasofu, S. I., & Chao, J. (1980). Interplanetary shock waves and magnetospheric substorms. *Planetary and Space Science*, *28*(4), 381–385. https://doi.org/10.1016/0032-0633(80)90042-2

Alves, M. V., Echer, E., & Gonzalez, W. D. (2006). Geoeffectiveness of corotating interaction regions as measured by Dst index. *Journal of Geophysical Research*, *111*, A07S05. https://doi.org/10.1029/2005JA011379

Araki, T. (1977). Global structure of geomagnetic sudden commencements. *Planetary and Space Science*, *25*(4), 373–384. https://doi.org/10.1016/0032-0633(77)90053-8

Araki, T. (1994). A physical model of the geomagnetic sudden commencement. In M. Engebretson, K. Takahashi, M. Scholer (Eds.), *Solar Wind Sources of Magnetospheric Ultra-Low-Frequency Waves* (pp. 183). Washington, DC: AGU.

Azari, A. R., Liemohn, M. W., Jia, X., Thomsen, M. F., Mitchell, D. G., Sergis, N., & Vandegriff, J. (2018). Interchange injections at Saturn: Statistical survey of energetic H+ sudden flux intensifications. *Journal of Geophysical Research: Space Physics*, *123*, 4692–4711. https://doi.org/10.1029/2018JA025391

Baker, D. N., Li, X., Pulkkinen, A., Ngwira, C. M., Mays, M. L., Galvin, A. B., & Simunac, K. D. C. (2013). A major solar eruptive event in July 2012: Defining extreme space weather scenarios. *Space Weather*, *11*, 585–591. https://doi.org/10.1002/swe.20097

Balan, N., Skoug, R., Tulasi Ram, S., Rajesh, P. K., Shiokawa, K., Otsuka, Y., & Nakamura, T. (2014). CME front and severe space weather. *Journal of Geophysical Research: Space Physics*, *119*, 10,041–10,058. https://doi.org/10.1002/2014JA020151

Barnard, L. A., Owens, M. J., Scott, C. J., & Jones, S. R. (2019). Extracting inner-heliosphere solar wind speed information from heliospheric imager observations. *Space Weather*, *17*, 925–938. https://doi.org/10.1029/2019SW002226

Barnes, D. (2020). Remote sensing estimates of CME density in the ecliptic using the STEREO heliospheric imagers. *Journal of Geophysical Research: Space Physics*, *125*, e2019JA027175. https://doi.org/10.1029/2019JA027175

Beland, J., & Small, K. (2004). Space weather effects on power transmission systems: The cases of Hydro-Quebec and transpower NewZealand Ltd. In I. Daglis (Ed.), *Effects of space weather on technology infrastructure* (Vol. 176, pp. 287–299): SPRINGER.

Bentley, S. N., Watt, C. E. J., Owens, M. J., & Rae, I. J. (2018). ULF wave activity in the magnetosphere: Resolving solar wind interdependencies to identify driving mechanisms. *Journal of Geophysical Research: Space Physics*, *123*, 2745–2771. https://doi.org/10.1002/2017JA024740

Bhaskar, A., & Vichare, G. (2019). Forecasting of SYMH and ASYH indices for geomagnetic storms of solar cycle 24 including St. Patrick's day, 2015 storm using NARX neural network. *Journal of Space Weather and Space Climate*, *9*, A12. https://doi.org/10.1051/swsc/2019007

Borovsky, J. E., & Denton, M. H. (2006). Differences between CME-driven storms and CIR-driven storms. *Journal of Geophysical Research*, *111*, A07S08. https://doi.org/10.1029/2005JA011447

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st). New York: Routledge. https://doi.org/10.1201/9781315139470

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brueckner, G. E., Delaboudiniere, J. P., Howard, R. A., Paswaters, S. E., St. Cyr, O. C., Schwenn, R., & Wang, D. (1998). Geomagnetic storms caused by coronal mass ejections (CMEs): March 1996 through June 1997. *Geophysical Research Letters*, *25*(15), 3019–3022. https://doi.org/10.1029/98GL00704

Burton, R. K., McPherron, R. L., & Russell, C. T. (1975). An empirical relationship between interplanetary conditions and *Dst*. *Journal of Geophysical Research*, *80*(31), 4204–4214. https://doi.org/10.1029/JA080i031p04204

Byrne, J. P., Maloney, S. A., McAteer, R. J., Refojo, J. M., & Gallagher, P. T. (2010). Propagation of an Earth-directed coronal mass ejection in three dimensions. *Nature Communications*, *1*(1), 74. https://doi.org/10.1038/ncomms1077

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, *17*, 1166–1207. https://doi.org/10.1029/2018SW002061

Carrington, R. C. (1859). Description of a singular appearance seen in the Sun on September 1, 1859. *Monthly Notices of the Royal Astronomical Society*, *20*(1), 13–15. https://doi.org/10.1093/mnras/20.1.13

Carter, B. A., Pradipta, R., Zhang, K., Yizengaw, E., Halford, A. J., & Norman, R. (2015). Interplanetary shocks and the resulting geomagnetically induced currents at the equator. *Geophysical Research Letters*, *42*, 6554–6559. https://doi.org/10.1002/2015gl065060

Cash, M. D., Witters Hicks, S., Biesecker, D. A., Reinard, A. A., de Koning, C. A., & Weimer, D. R. (2016). Validation of an operational product to determine L1 to Earth propagation time delays. *Space Weather*, *14*, 93–112. https://doi.org/10.1002/2015SW001321

Cash, M. D., Wrobel, J. S., Cosentino, K. C., & Reinard, A. A. (2014). Characterizing interplanetary shocks for development and optimization of an automated solar wind shock detection algorithm. *Journal of Geophysical Research: Space Physics*, *119*, 4210–4222. https://doi.org/10.1002/2014JA019800

Chandorkar, M., Camporeale, E., & Wing, S. (2017). Probabilistic forecasting of the disturbance storm time index: An autoregressive Gaussian process approach. *Space Weather*, *15*, 1004–1019. https://doi.org/10.1002/2017SW001627

Chao, J. K., & Lepping, R. P. (1974). A correlative study of SSC's, interplanetary shocks, and solar activity. *Journal of Geophysical Research*, *79*(13), 1799–1807. https://doi.org/10.1029/JA079i013p01799

Chen, Y. W., & Lin, C. J. (2006). Combining SVMs with various feature selection strategies. https://doi.org/10.1007/978-3-540-35488-8_13

Chree, C. (1925). The times of "Sudden Commencements" (S.C.s) of magnetic storms: Observation and theory. *Proceedings of the Physical Society of London*, *38*(1), 35–46. https://doi.org/10.1088/1478-7814/38/1/305

Crooker, N. U., Balogh, A., Bothmer, V., Forsyth, R. J., Gloeckler, G., Hewish, A., & Mikić, Z. (1999). Space science reviews. D. Reidel Pub. Co, https://deepblue.lib.umich.edu/handle/2027.42/43796

Crown, M. D. (2012). Validation of the NOAA Space Weather Prediction Center's solar flare forecasting look-up table and forecaster-issued probabilities. *Space Weather*, *10*, S06006. https://doi.org/10.1029/2011SW000760

Curto, J. J., Araki, T., & Alberca, L. F. (2007). Evolution of the concept of Sudden Storm Commencements and their operative identification. *Earth, Planets and Space*, *59*(11), i–xii. https://doi.org/10.1186/BF03352059

Davies, J. A., Harrison, R. A., Perry, C. H., Möstl, C., Lugaz, N., Rollett, T., & Savani, N. P. (2012). A self-similar expansion model for use in solar wind transient propagation studies. *The Astrophysical Journal*, *750*(1), 23. https://doi.org/10.1088/0004-637X/750/1/23

Davies, J. A., Perry, C. H., Trines, R. M. G. M., Harrison, R. A., Lugaz, N., Möstl, C., & Steed, K. (2013). Establishing a stereoscopic technique for determining the kinematic properties of solar wind transients based on a generalized self-similarly expanding circular geometry. *The Astrophysical Journal*, *777*(2), 167. https://doi.org/10.1088/0004-637X/777/2/167

Dimmock, A., Rosenqvist, L., Hall, J., Viljanen, A., Yordanova, E., Honkonen, I., & Sjöberg, E. (2019). The GIC and geomagnetic response over Fennoscandia to the 7–8 September 2017 geomagnetic storm. *Space Weather*, *17*, 989–1010. https://doi.org/10.1029/2018SW002132

Echer, E., & Gonzalez, W. D. (2004). Geoeffectiveness of interplanetary shocks, magnetic clouds, sector boundary crossings and their combined occurrence. *Geophysical Research Letters*, *31*, L09808. https://doi.org/10.1029/2003GL019199

Echer, E., Gonzalez, W. D., Tsurutani, B. T., & Gonzalez, A. L. C. (2008). Interplanetary conditions causing intense geomagnetic storms (Dst ≤ -100 nT) during solar cycle 23 (1996-2006). *Journal of Geophysical Research*, *113*, A05221. https://doi.org/10.1029/2007JA012744

Fairfield, D. H., & Cahill, L. J. (1966). Transition region magnetic field and polar magnetic disturbances. *Journal of Geophysical Research*, *71*(1), 155–169. https://doi.org/10.1029/JZ071i001p00155

Fiori, R. A. D., Boteler, D. H., & Gillies, D. M. (2014). Assessment of GIC risk due to geomagnetic sudden commencements and identification of the current systems responsible. *Space Weather*, *12*, 76–91. https://doi.org/10.1002/2013SW000967

Forsyth, C., Watt, C. E. J., Mooney, M. K., Rae, I. J., Walton, S. D., & Horne, R. B. (2020). Forecasting GOES 15 >2 MeV electron fluxes from solar wind data and geomagnetic indices. *Space Weather*, *18*, e2019SW002416. https://doi.org/10.1029/2019SW002416

Freeman, M. P., Forsyth, C., & Rae, I. J. (2019). The influence of substorms on extreme rates of change of the surface horizontal magnetic field in the U.K. *Space Weather*, *17*, 827–844. https://doi.org/10.1029/2018SW002148

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Ros, G., Tsuru, B. T., & Vasyliunas, V. M. (1994). What is a geomagnetic storm? https://doi.org/10.1029/93JA02867

Gonzalez, W. D., & Tsurutani, B. T. (1987). Criteria of interplanetary parameters causing intense magnetic storms (Dst < -100 nT). *Planetary and Space Science*, *35*(9), 1101–1109. https://doi.org/10.1016/0032-0633(87)90015-8

Gonzalez, W. D., Tsurutani, B. T., & Clúa de Gonzalez, A. L. (1999). Interplanetary origin of geomagnetic storms. *Space Science Reviews*, *88*(3/4), 529–562. https://doi.org/10.1023/A:1005160129098

Gosling, J. T. (1993). The solar flare myth. *Journal of Geophysical Research*, *98*(A11), 18,937–18,949. https://doi.org/10.1029/93JA01896

Gosling, J. T., McComas, D. J., Phillips, J. L., & Bame, S. J. (1991). Geomagnetic activity associated with earth passage of interplanetary shock disturbances and coronal mass ejections. *Journal of Geophysical Research*, *96*(A5), 7831. https://doi.org/10.1029/91JA00316

Hayakawa, H., Ebihara, Y., Willis, D. M., Toriumi, S., Iju, T., Hattori, K., & Knipp, D. J. (2019). Temporal and spatial evolutions of a large sunspot group and great auroral storms around the Carrington Event in 1859. *Space Weather*, *17*, 1553–1569. https://doi.org/10.1029/2019SW002269

Ho, T. K. (1995). Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)*, IEEE Computer Society, USA, pp. 278.

Hoffman, R. A., Cahill, L. J., Anderson, R. R., Maynard, N. C., Smith, P. H., Fritz, T. A., & Gurnett, D. A. (1975). Explorer 45 (S-A) observations of the magnetosphere and magnetopause during the August 4–6, 1972, magnetic storm period. *Journal of Geophysical Research*, *80*(31), 4287–4296. https://doi.org/10.1029/JA080i031p04287

Hosmer, D. W., & Lemeshow, S. (2005). Introduction to the logistic regression model, *Applied Logistic Regression* (pp. 1–30). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471722146.ch1

Hugoniot, P. H. (1887). Mémoire sur la propagation du mouvement dans les corps et plus spécialement dans les gaz parfaits. 1e Partie. *J. Ecole Polytech*, *57*, 3–97.

Hugoniot, P. H. (1889). Mémoire sur la propagation du mouvement dans les corps et plus spécialement dans les gaz parfaits. 2e Partie. *Journal de l'Ecole Polytechnique*, *58*, 1–125.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Huttunen, K. E. J., Schwenn, R., Bothmer, V., & Koskinen, H. E. J. (2005). Properties and geoeffectiveness of magnetic clouds in the rising, maximum and early declining phases of solar cycle 23. *Annales Geophysicae*, *23*(2), 625–641. https://doi.org/10.5194/angeo-23-625-2005

Illing, R. M. E., & Hundhausen, A. J. (1983). Possible observation of a disconnected magnetic structure in a coronal transient. *Journal of Geophysical Research*, *88*(A12), 10,210–10,214. https://doi.org/10.1029/JA088iA12p10210

Ji, E. Y., Moon, Y. J., Park, J., Lee, J. Y., & Lee, D. H. (2013). Comparison of neural network and support vector machine methods for Kp forecasting. *Journal of Geophysical Research: Space Physics*, *118*, 5109–5117. https://doi.org/10.1002/jgra.50500

Joselyn, J. A., & Tsurutani, B. T. (1990). Geomagnetic sudden impulses and storm sudden commencements: A note on terminology. *Eos, Transactions American Geophysical Union*, *71*(47), 1808. https://doi.org/10.1029/90EO00350

Jurac, S., Kasper, J. C., Richardson, J. D., & Lazarus, A. J. (2002). Geomagnetic disturbances and their relationship to interplanetary shock parameters. *Geophysical Research Letters*, *29*(10), 101–1–101–4. https://doi.org/10.1029/2001GL014034

Kahler, S. W., & Webb, D. F. (2007). V arc interplanetary coronal mass ejections observed with the Solar Mass Ejection Imager. *Journal of Geophysical Research*, *112*, A09103. https://doi.org/10.1029/2007JA012358

Kamide, Y., Baumjohann, W., Daglis, I. A., Gonzalez, W. D., Grande, M., Joselyn, J. A., & Vasyliunas, V. M. (1998). Current understanding of magnetic storms: Storm-substorm relationships. *Journal of Geophysical Research*, *103*(A8), 17,705–17,728. https://doi.org/10.1029/98JA01426

Kang, S. M., Moon, Y. J., Cho, K. S., Kim, Y. H., Park, Y. D., Baek, J. H., & Chang, H. Y. (2006). Coronal mass ejection geoeffectiveness depending on field orientation and interplanetary coronal mass ejection classification. *Journal of Geophysical Research*, *111*, A05102. https://doi.org/10.1029/2005JA011445

Kappenman, J. G. (1996). Geomagnetic storms and their impact on power systems. *IEEE Power Engineering Review*, *16*(5), 5. https://doi.org/10.1109/MPER.1996.491910

Kappenman, J. G. (2003). Storm sudden commencement events and the associated geomagnetically induced current risks to ground-based systems at low-latitude and midlatitude locations. *Space Weather*, *1*(3), 1016. https://doi.org/10.1029/2003SW000009

Kappenman, J. G., & Albertson, D. (1990). Bracing for the geomagnetic storms. *IEEE Spectrum*, *27*(3), 27–33. https://doi.org/10.1109/6.48847

Kilpua, E., Lugaz, N., Mays, M. L., & Temmer, M. (2019). Forecasting the structure and orientation of earthbound coronal mass ejections. *Space Weather*, *17*, 498–526. https://doi.org/10.1029/2018SW001944

Kim, R. S., Cho, K. S., Moon, Y. J., Dryer, M., Lee, J., Yi, Y., & Kim, Y. H. (2010). An empirical model for prediction of geomagnetic storms using initially observed CME parameters at the Sun. *Journal of Geophysical Research: Space Physics*, *115*, A12108. https://doi.org/10.1029/2010JA015322

Kim, R. S., Moon, Y. J., Gopalswamy, N., Park, Y. D., & Kim, Y. H. (2014). Two-step forecast of geomagnetic storm using coronal mass ejection and solar wind condition. *Space Weather*, *12*, 246–256. https://doi.org/10.1002/2014SW001033

Knipp, D. J., Fraser, B. J., Shea, M. A., & Smart, D. F. (2018). On the little-known consequences of the 4 August 1972 ultra-fast coronal mass ejecta: Facts, commentary, and call to action. *Space Weather*, *16*, 1635–1643. https://doi.org/10.1029/2018SW002024

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. https://robotics.stanford.edu/\protect\LY1\textbraceleft\protect\LY1\textbracerightronnyk

Kokubun, S., McPherron, R. L., & Russell, C. T. (1977). Triggering of substorms by solar wind discontinuities. *Journal of Geophysical Research*, *82*(1), 74–86. https://doi.org/10.1029/JA082i001p00074

Kruparova, O., Maksimovic, M., Šafránková, J., Němeček, Z., Santolik, O., & Krupar, V. (2013). Automated interplanetary shock detection and its application to wind observations. *Journal of Geophysical Research: Space Physics*, *118*, 4793–4803. https://doi.org/10.1002/jgra.50468

Kugblenu, S., Taguchi, S., & Okuzawa, T. (1999). Prediction of the geomagnetic storm associated Dst index using an artificial neural network algorithm. *Earth, Planets and Space*, *51*(4), 307–313. https://doi.org/10.1186/BF03352234

Lepping, R. P., Acna, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., & Worley, E. M. (1995). The WIND magnetic field investigation. *Space Science Reviews*, *71*(1–4), 207–229. https://doi.org/10.1007/BF00751330

Lepping, R., Berdichevsky, D., Burlaga, L., Lazarus, A., Kasper, J., Desch, M., & Ackerson, K. (2001). The Bastille day magnetic clouds and upstream shocks: Near-Earth interplanetary observations. *Solar Physics*, *204*(1/2), 285–303. https://doi.org/10.1023/A:1014264327855

Lethy, A., El-Eraki, M. A., Samy, A., & Deebes, H. A. (2018). Prediction of the Dst index and analysis of its dependence on solar wind parameters using neural network. *Space Weather*, *16*, 1277–1290. https://doi.org/10.1029/2018SW001863

Lindsay, G. M., Russell, C. T., Luhmann, J. G., & Gazis, P. (1994). On the sources of interplanetary shocks at 0.72 AU. *Journal of Geophysical Research*, *99*(A1), 11. https://doi.org/10.1029/93JA02666

Liu, Y. D., Luhmann, J. G., Kajdič, P., Kilpua, E. K., Lugaz, N., Nitta, N. V., & Galvin, A. B. (2014). Observations of an extreme storm in interplanetary space caused by successive coronal mass ejections. *Nature Communications*, *5*(1), 3481. https://doi.org/10.1038/ncomms4481

Lugaz, N., Farrugia, C. J., Winslow, R. M., Al-Haddad, N., Kilpua, E. K. J., & Riley, P. (2016). Factors affecting the geoeffectiveness of shocks and sheaths at 1 AU. *Journal of Geophysical Research: Space Physics*, *121*, 10,861–10,879. https://doi.org/10.1002/2016JA023100

Luhmann, J. G. (1997). CMEs and space weather, *Geophysical Monograph Series* (Vol. 99, pp. 291–299). Washington, DC: American Geophysical Union (AGU). https://doi.org/10.1029/GM099p0291

Lundstedt, H., Gleisner, H., & Wintoft, P. (2002). Operational forecasts of the geomagnetic Dst index. *Geophysical Research Letters*, *29*(24), 2181. https://doi.org/10.1029/2002GL016151

Marshall, R. A., Dalzell, M., Waters, C. L., Goldthorpe, P., & Smith, E. A. (2012). Geomagnetically induced currents in the New Zealand power network. *Space Weather*, *10*, S08003. https://doi.org/10.1029/2012SW000806

Mayaud, P. N. (1973). A hundred year series of geomagnetic data, 1868-1967: Indices aa, storm sudden commencements (SSC). IUGG Publ. Office, 256.

McComas, D., Bame, S., Barker, P., Feldman, W., Phillips, J., Riley, P., & Griffee, J. (1998). Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. *Space Science Reviews*, *86*(1/4), 563–612. https://doi.org/10.1023/A:1005040232597

McKinney, W. (2010). Data structures for statistical computing in Python. https://conference.scipy.org/proceedings/scipy2010/mckinney.html

Morley, S. K. (2020). Challenges and opportunities in magnetospheric space weather prediction. *Space Weather*, *18*, e2018SW002108. https://doi.org/10.1029/2018SW002108

Murray, S. A., Bingham, S., Sharpe, M., & Jackson, D. R. (2017). Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather*, *15*, 577–588. https://doi.org/10.1002/2016SW001579

Ngwira, C. M., Pulkkinen, A., Kuznetsova, M. M., & Glocer, A. (2014). Modeling extreme Carrington-type space weather events using three-dimensional global MHD simulations. *Journal of Geophysical Research: Space Physics*, *119*, 4456–4474. https://doi.org/10.1002/2013JA019661

Ngwira, C. M., Pulkkinen, A., Leila Mays, M., Kuznetsova, M. M., Galvin, A. B., Simunac, K., & Glocer, A. (2013). Simulation of the 23 July 2012 extreme space weather event: What if this extremely rare CME was Earth directed? *Space Weather*, *11*, 671–679. https://doi.org/10.1002/2013SW000990

Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lobell, J., & Gergin, E. (1995). SWE, a comprehensive plasma instrument for the WIND spacecraft. *Space Science Reviews*, *71*(1-4), 55–77. https://doi.org/10.1007/BF00751326

Oliveira, D. M., Arel, D., Raeder, J., Zesta, E., Ngwira, C. M., Carter, B. A., & Gjerloev, J. W. (2018). Geomagnetically induced currents caused by interplanetary shocks with different impact angles and speeds. *Space Weather*, *16*, 636–647. https://doi.org/10.1029/2018SW001880

Oliveira, D. M., & Raeder, J. (2014). Impact angle control of interplanetary shock geoeffectiveness. *Journal of Geophysical Research: Space Physics*, *119*, 8188–8201. https://doi.org/10.1002/2014JA020275

Oliveira, D. M., & Raeder, J. (2015). Impact angle control of interplanetary shock geoeffectiveness: A statistical study. *Journal of Geophysical Research: Space Physics*, *120*, 4313–4323. https://doi.org/10.1002/2015JA021147

Oliveira, D. M., Raeder, J., Tsurutani, B. T., & Gjerloev, J. W. (2016). Effects of interplanetary shock inclinations on nightside auroral power intensity. *Brazilian Journal of Physics*, *46*(1), 97–104. https://doi.org/10.1007/s13538-015-0389-9

Oliveira, D. M., & Samsonov, A. (2018). Geoeffectiveness of interplanetary shocks controlled by impact angles: A review. *Advances in Space Research*, *61*(1), 1–44. https://doi.org/10.1016/J.ASR.2017.10.006

Ontiveros, V., & Gonzalez-Esparza, J. A. (2010). Geomagnetic storms caused by shocks and ICMEs. *Journal of Geophysical Research*, *115*, A10244. https://doi.org/10.1029/2010JA015471

Owens, M. J., Lockwood, M., & Barnard, L. A. (2020). The value of CME arrival-time forecasts for space weather mitigation. *Space Weather*, *18*, e2020SW002507. https://doi.org/10.1029/2020SW002507

Pérez, A., Larrañaga, P., & Inza, I. (2006). Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *International Journal of Approximate Reasoning*, *43*(1), 1–25. https://doi.org/10.1016/J.IJAR.2006.01.002

Park, W., Lee, J., Yi, Y., Ssessanga, N., & Oh, S. (2015). Storm sudden commencements without interplanetary shocks. *Journal of Astronomy and Space Sciences*, *32*(3), 181–187. https://doi.org/10.5140/JASS.2015.32.3.181

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, *10*, 61–74.

Pulkkinen, A., Bernabeu, E., Eichner, J., Beggan, C., & Thomson, A. W. P. (2012). Generation of 100-year geomagnetically induced current scenarios. *Space Weather*, *10*, S04003. https://doi.org/10.1029/2011sw000750

Pulkkinen, A., Lindahl, S., Viljanen, A., & Pirjola, R. (2005). Geomagnetic storm of 29–31 October 2003: Geomagnetically induced currents and their relation to problems in the Swedish high-voltage power transmission system. *Space Weather*, *3*, S08C03. https://doi.org/10.1029/2004SW000123

Rankine, W. J. M. (1870). XV. On the thermodynamic theory of waves of finite longitudinal disturbance. *Philosophical Transactions of the Royal Society of London*, *160*, 277–288. https://doi.org/10.1098/rstl.1870.0015

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Berlin Heidelberg New York: MIT Press.

Richardson, I. G., & Cane, H. V. (2012). Near-earth solar wind flows and related geomagnetic activity during more than four solar cycles (1963-2011). *Journal of Space Weather and Space Climate*, *2*, A02. https://doi.org/10.1051/swsc/2012003

Richardson, I. G., Cliver, E. W., & Cane, H. V. (2001). Sources of geomagnetic storms for solar minimum and maximum conditions during 1972-2000. *Geophysical Research Letters*, *28*(13), 2569–2572. https://doi.org/10.1029/2001GL013052

Richter, A. K., Hsieh, K. C., Luttrell, A. H., Marsch, E., & Schwenn, R. (1985). *Review of interplanetary shock phenomena near and within 1 AU*, pp. 33–50. Washington, DC: American Geophysical Union (AGU). https://doi.org/10.1029/GM035p0033

Riley, P., Gosling, J. T., & Pizzo, V. J. (1997). A two-dimensional simulation of the radial and latitudinal evolution of a solar wind disturbance driven by a fast, high-pressure coronal mass ejection. *Journal of Geophysical Research*, *102*(A7), 14,677–14,685. https://doi.org/10.1029/97JA01131

Rogers, N. C., Wild, J. A., Eastoe, E. F., Gjerloev, J. W., & Thomson, A. W. P. (2020). A global climatological model of extreme geomagnetic field fluctuations. *Journal of Space Weather and Space Climate*, *10*, 5. https://doi.org/10.1051/swsc/2020008

Russell, C. T., Ginskey, M., Petrinec, S., & Le, G. (1992). The effect of solar wind dynamic pressure changes on low and mid-latitude magnetic records. *Geophysical Research Letters*, *19*(12), 1227–1230. https://doi.org/10.1029/92GL01161

Russell, C. T., Mewaldt, R. A., Luhmann, J. G., Mason, G. M., von Rosenvinge, T. T., Cohen, C. M. S., & Simunac, K. D. C. (2013). The very unusual interplanetary coronal mass ejection of 2012 July 23: a blast wave mediated by solar energetic particles. *The Astrophysical Journal*, *770*(1), 38. https://doi.org/10.1088/0004-637X/770/1/38

Salman, T. M., Lugaz, N., Farrugia, C. J., Winslow, R. M., Galvin, A. B., & Schwadron, N. A. (2018). Forecasting periods of strong southward magnetic field following interplanetary shocks. *Space Weather*, *16*, 2004–2021. https://doi.org/10.1029/2018SW002056

Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine Learning*, *13*(1), 135–143. https://doi.org/10.1007/BF00993106

Scott, D. W. (2015). Multivariate density estimation: Theory, practice, and visualization.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486. https://doi.org/10.2307/2290328

Smith, A. W., Freeman, M. P., Rae, I. J., & Forsyth, C. (2019). The influence of sudden commencements on the rate of change of the surface horizontal magnetic field in the United Kingdom. *Space Weather*, *17*, 1605–1617. https://doi.org/10.1029/2019SW002281

Smith, C., L'Heureux, J., Ness, N., Acuña, M., Burlaga, L., & Scheifele, J. (1998). The ACE magnetic fields experiment. *Space Science Reviews*, *86*(1/4), 613–632. https://doi.org/10.1023/A:1005092216668

Smith, E. J., & Wolfe, J. H. (1976). Observations of interaction regions and corotating shocks between one and five AU: Pioneers 10 and 11. *Geophysical Research Letters*, *3*(3), 137–140. https://doi.org/10.1029/GL003i003p00137

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science Science*, *240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

Tan, Y., Hu, Q., Wang, Z., & Zhong, Q. (2018). Geomagnetic index Kp forecasting with LSTM. *Space Weather*, *16*, 406–416. https://doi.org/10.1002/2017SW001764

Temerin, M., & Li, X. (2002). A new model for the prediction of Dst on the basis of the solar wind. *Journal of Geophysical Research*, *107*(A12), SMP 31–1–SMP 31–8. https://doi.org/10.1029/2001JA007532

Tsurutani, B. T., & Gonzalez, W. D. (1998). Magnetic storms, *From the Sun: Auroras, magnetic storms, solar flares, cosmic rays* (pp. 57–66). Washington, D. C.: American Geophysical Union. https://doi.org/10.1029/SP050p0057

Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L. C., Guarnieri, F. L., Gopalswamy, N., Grande, M., & Vasyliunas, V. (2006). Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of Geophysical Research*, *111*, A07S01. https://doi.org/10.1029/2005JA011273

Tsurutani, B. T., Gonzalez, W. D., Lakhina, G. S., & Alex, S. (2003). The extreme magnetic storm of 1–2 September 1859. *Journal of Geophysical Research*, *108*(A7), 1268. https://doi.org/10.1029/2002JA009504

Tsurutani, B. T., Gonzalez, W. D., Tang, F., Akasofu, S. I., & Smith, E. J. (1988). Origin of interplanetary southward magnetic fields responsible for major magnetic storms near solar maximum (1978–1979). *Journal of Geophysical Research*, *93*(A8), 8519. https://doi.org/10.1029/JA093iA08p08519

Tsurutani, B. T., Gonzalez, W. D., Tang, F., & Lee, Y. T. (1992). Great magnetic storms. *Geophysical Research Letters*, *19*(1), 73–76. https://doi.org/10.1029/91GL02783

Tsurutani, B., Lakhina, G., Verkhoglyadova, O., Gonzalez, W., Echer, E., & Guarnieri, F. (2011). A review of interplanetary discontinuities and their geomagnetic effects. *Journal of Atmospheric and Solar-Terrestrial Physics*, *73*(1), 5–19. https://doi.org/10.1016/J.JASTP.2010.04.001

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., & Contributors, S (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vorotnikov, V. S., Smith, C. W., Farrugia, C. J., Hu, Q., Szabo, A., & Yumoto, K. (2011). Use of single-component wind speed in Rankine-Hugoniot analysis of interplanetary shocks. *Space Weather*, *9*, S04001. https://doi.org/10.1029/2010SW000631

Vorotnikov, V. S., Smith, C. W., Hu, Q., Szabo, A., Skoug, R. M., & Cohen, C. M. S. (2008). Automated shock detection and analysis algorithm for space weather application. *Space Weather*, *6*, S03002. https://doi.org/10.1029/2007SW000358

Wang, C., Li, C. X., Huang, Z. H., & Richardson, J. D. (2006). Effect of interplanetary shock strengths and orientations on storm sudden commencement rise times. *Geophysical Research Letters*, *33*, L14104. https://doi.org/10.1029/2006GL025966

Wang, C., Li, H., Richardson, J. D., & Kan, J. R. (2010). Interplanetary shock characteristics and associated geosynchronous magnetic field variations estimated from sudden impulses observed on the ground. *Journal of Geophysical Research*, *115*, A09215. https://doi.org/10.1029/2009JA014833

Wang, R., Lu, Q., Nakamura, R., Baumjohann, W., Russell, C. T., Burch, J. L., & Gershman, D. (2017). Interaction of magnetic flux ropes via magnetic reconnection observed at the magnetopause. *Journal of Geophysical Research: Space Physics*, *122*, 10,436–10,447. https://doi.org/10.1002/2017JA024482

Wang, Y., Shen, C. L., Wang, S., & Ye, P. Z. (2003). An empirical formula relating the geomagnetic storm's intensity to the interplanetary parameters: ′ and Δt. *Geophysical Research Letters*, *30*(20), 2039. https://doi.org/10.1029/2003GL017901

Webb, D. F., & Howard, T. A. (2012). Coronal mass ejections: Observations. *Living Reviews in Solar Physics*, *9*(1), 3. https://doi.org/10.12942/lrsp-2012-3

Webb, G. I., Keogh, E., Miikkulainen, R., Miikkulainen, R., & Sebag, M. (2011). Naïve Bayes, *Encyclopedia of Machine Learning* (pp. 713–714). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-30164-8_576

Wing, S., Johnson, J. R., Jen, J., Meng, C. I., Sibeck, D. G., Bechtold, K., & Takahashi, K. (2005). Kp forecast models. *Journal of Geophysical Research*, *110*, A04203. https://doi.org/10.1029/2004JA010500

Wintoft, P., Wik, M., Matzka, J., & Shprits, Y. (2017). Forecasting Kp from solar wind data: input parameter study using 3-hour averages and 3-hour range values. *Journal of Space Weather and Space Climate*, *7*, A29. https://doi.org/10.1051/swsc/2017027

Wu, J. G., & Lundstedt, H. (1996). Prediction of geomagnetic storms from solar wind data using Elman Recurrent Neural Networks. *Geophysical Research Letters*, *23*(4), 319–322. https://doi.org/10.1029/96GL00259

Yue, C., Zong, Q. G., Zhang, H., Wang, Y. F., Yuan, C. J., Pu, Z. Y., & Wang, C. R. (2010). Geomagnetic activity triggered by interplanetary shocks. *Journal of Geophysical Research*, *115*, A00I05. https://doi.org/10.1029/2010JA015356

Zhang, H. (2004). The optimality of Naïve Bayes. In Flairs 2004 Conference, https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.483.2183

Zhang, J., Richardson, I. G., Webb, D. F., Gopalswamy, N., Huttunen, E., Kasper, J. C., & Zhukov, A. N. (2007). Solar and interplanetary sources of major geomagnetic storms (Dst ≤ −100 nT) during 1996–2005. *Journal of Geophysical Research*, *112*, A10102. https://doi.org/10.1029/2007JA012321

Zhang, J. J., Wang, C., Sun, T. R., Liu, C. M., & Wang, K. R. (2015). GIC due to storm sudden commencement in low-latitude high-voltage power network in China: Observation and simulation. *Space Weather*, *13*, 643–655. https://doi.org/10.1002/2015SW001263

Zhou, X., & Tsurutani, B. T. (2001). Interplanetary shock triggering of nightside geomagnetic activity: Substorms, pseudobreakups, and quiescent events. *Journal of Geophysical Research*, *106*(A9), 18,957–18,967. https://doi.org/10.1029/2000JA003028

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. (Vol. 39; Tech. Rep. No. 4), https://www.floppybunny.org/robin/web/virtualclassroom/dss/articles/roccurves/roc_as_evaluation_tool_zweig_campbell_1993.pdf