Figure 1.

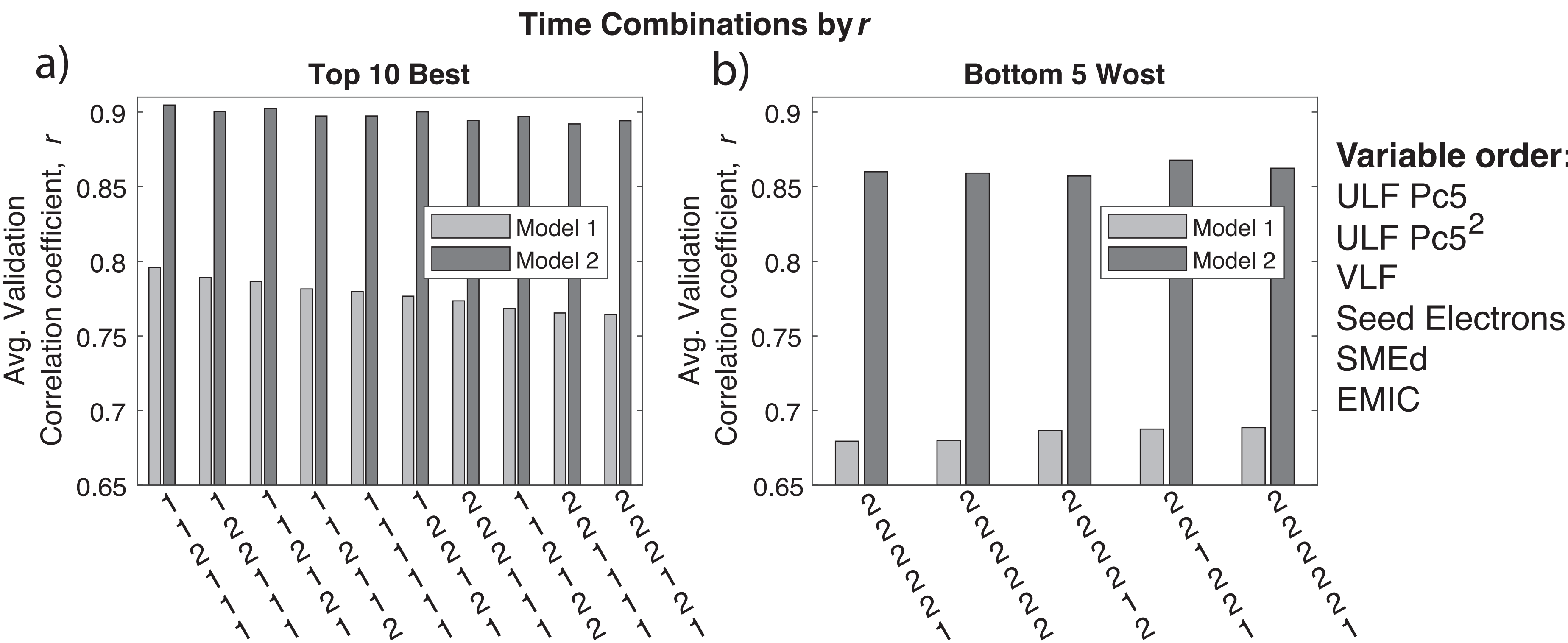
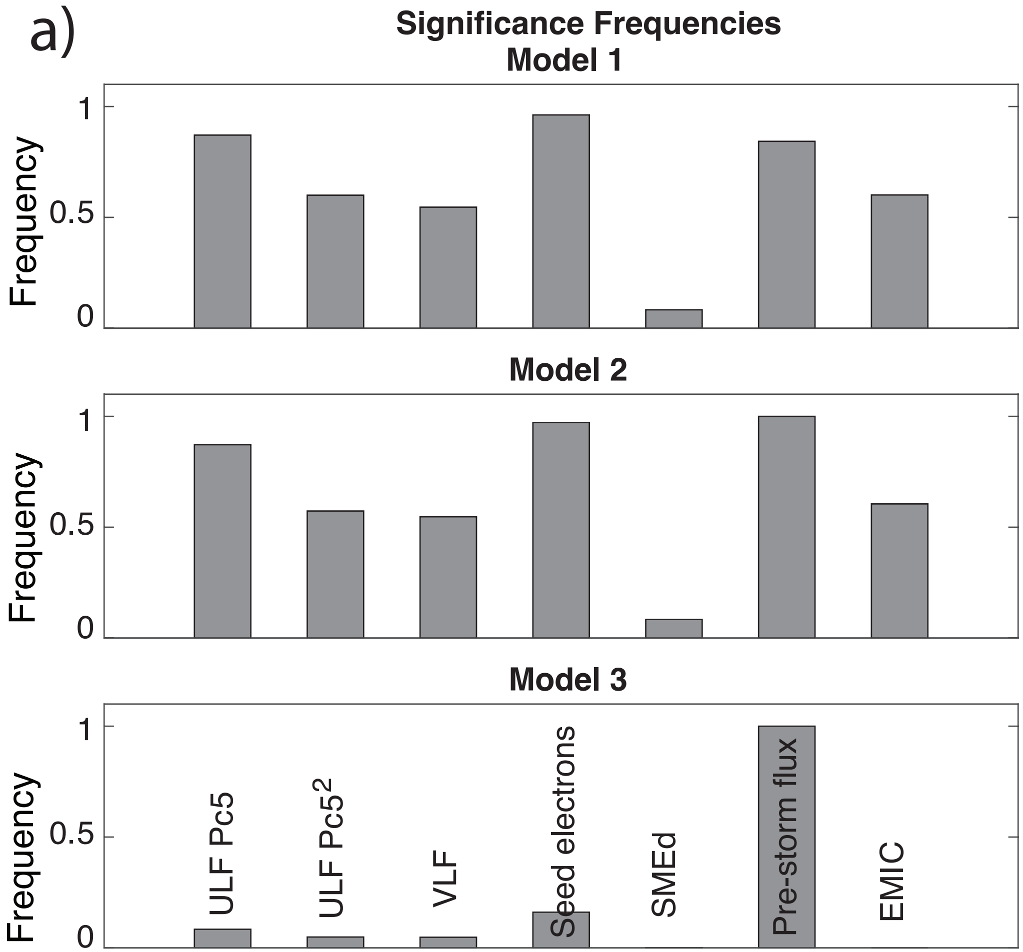

Time Combinations by r
a)
Top 10 Best
Avg. Validation
Correlation coefficient, r
0.9
0.85
0.8
0.75
0.7
0.65
Model 1
Model 2
1 1 2 1 1 1
1 2 2 1 1 1
1 1 2 1 2 1
1 1 2 1 1 2
1 1 1 1 1 1
1 2 2 1 2 1
2 2 2 1 1 1
1 1 2 1 2 2
2 2 1 1 1 1
2 2 2 1 2 1
b)
Bottom 5 Wost
Avg. Validation
Correlation coefficient, r
0.9
0.85
0.8
0.75
0.7
0.65
Model 1
Model 2
2 2 2 2 2 1
2 2 2 2 2 2
2 2 2 2 1 2
2 2 1 2 2 1
2 2 2 2 2 1
Variable order:
ULF Pc5
ULF Pc5^2
VLF
Seed Electrons
SMEd
EMIC

Figure 2.

a)



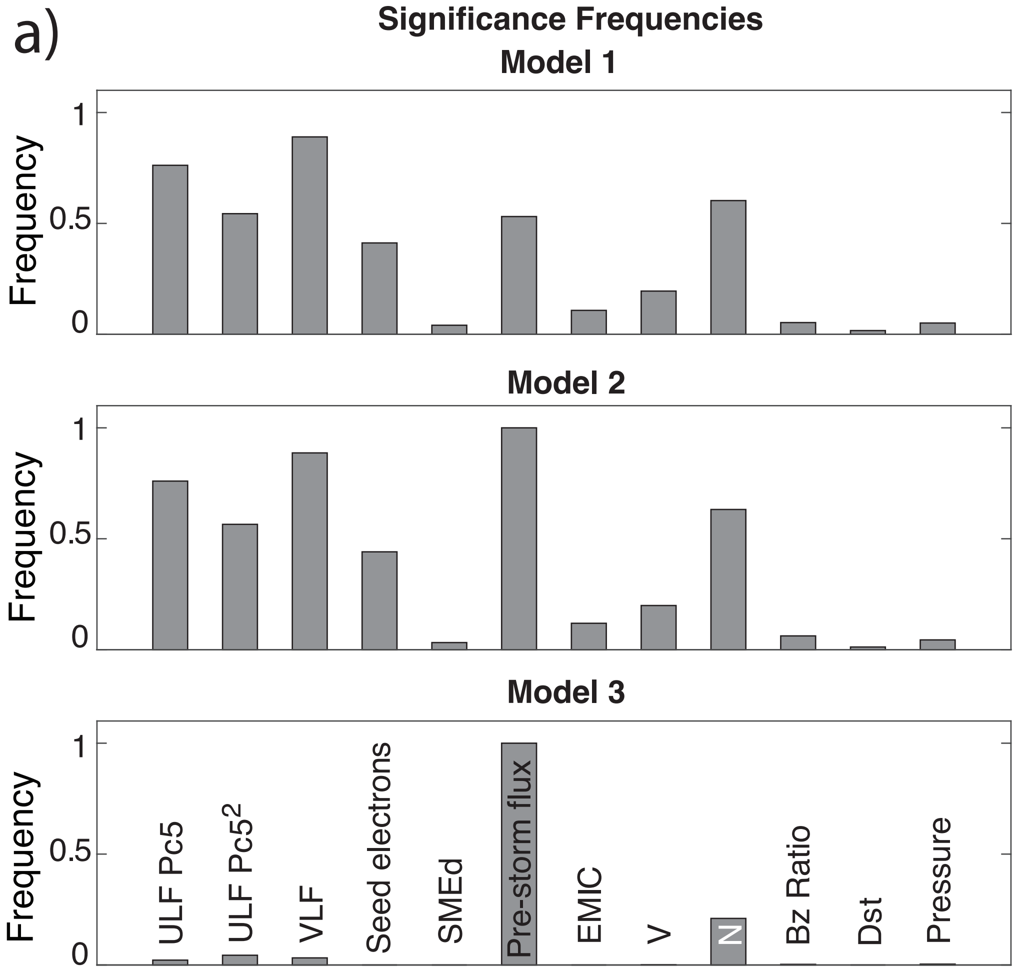
b)

Average Validation Correlation Coefficient r, TPR, TNR, ACC

	Avg. *r*	Avg. TPR	Avg. TNR	Avg. ACC
Model 1	0.80	N/A	N/A	N/A
Model 2	0.91	0.94	0.86	0.92
Model 3	N/A	0.90	0.82	0.87

Figure 3.

a)

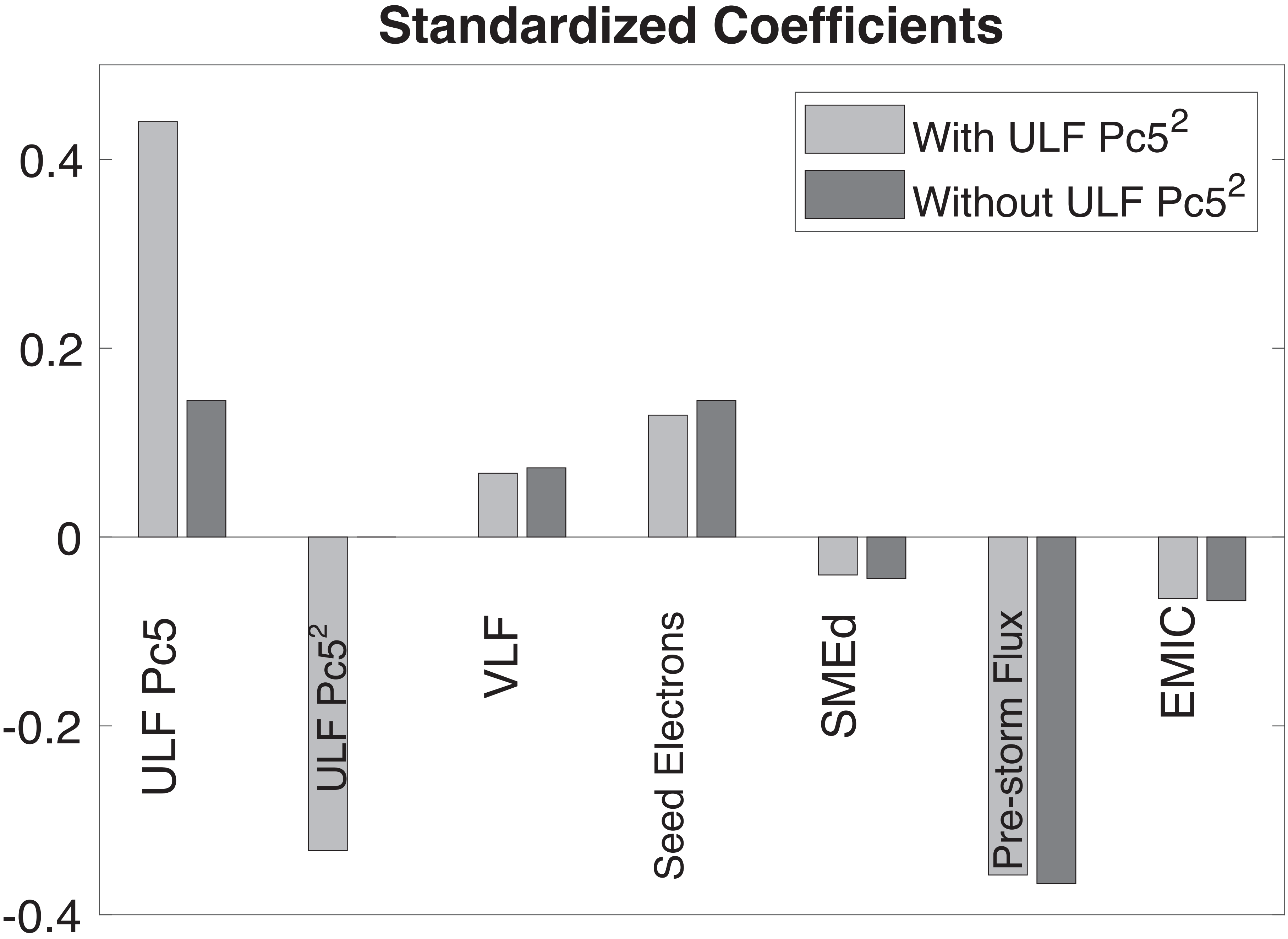


b)

Average Validation Correlation Coefficient r, TPR, TNR, ACC

	Avg. *r*	**Avg. TPR**	**Avg. TNR**	**Avg. ACC**
Model 1	0.80	N/A	N/A	N/A
Model 2	0.90	0.94	0.87	0.92
Model 3	N/A	0.88	0.79	0.85

Figure 4.



Standardized Coefficients
With ULF Pc5^2
Without ULF Pc5^2
0.4
0.2
0
-0.2
-0.4
ULF Pc5
ULF Pc5^2
VLF
Seed Electrons
SMEd
Pre-storm Flux
EMIC

Comparison of multiple and logistic regression analyses of relativistic electron flux enhancement at geosynchronous orbit following storms

N. S. S. Capman[1,2], L. E. Simms[1], M. J. Engebretson[1], M. A. Clilverd[3], C. J. Rodger[4], G. D. Reeves[5], M. R. Lessard[6], and J. Gjerloev[7]

[1] Department of Physics, Augsburg University, Minneapolis, MN
[2] now at the Department of Mechanical Engineering, University of Minnesota, Twin Cities, Minneapolis, MN
[3] British Antarctic Survey (NERC), Cambridge, UK
[4] Department of Physics, University of Otago, Dunedin, New Zealand
[5] Space Science and Applications Group, Los Alamos National Laboratory, Los Alamos, NM
[6] Space Science Center, University of New Hampshire, Durham, NH
[7] The Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Key Points

1. Following storms, increases in relativistic electron flux at geosynchronous orbit were well predicted by regression models
2. ULF, VLF, and EMIC waves, and seed electrons were all strong predictors
3. Three model types (logistic and linear regressions) had similar validation success

Abstract

Many factors influence relativistic outer radiation belt electron fluxes, such as waves in the ultra low frequency (ULF) Pc5, very low frequency (VLF), and electromagnetic ion cyclotron (EMIC) frequency bands, seed electron flux, Dst disturbance levels, substorm occurrence, and solar wind inputs. In this work we compared relativistic electron flux post storm vs. pre-storm using three methods of analysis: 1) multiple regression to predict flux values following storms, 2) multiple regression to predict the size and direction of the change in electron flux, and 3) multiple logistic regression to predict only the probability of the flux rising or falling. We

determined which is the most predictive model, and which factors are most influential. We found that a linear regression predicting the difference in pre-storm and post storm flux (Model 2) results in the highest validation correlations. The logistic regression used in Model 3 had slightly weaker predictive abilities than the other two models, but had most value in providing a prediction of the probability of the electron flux increasing after a storm. Of the variables used (ULF Pc5 and VLF waves, seed electrons, substorm activity, and EMIC waves), the most influential in the final model were ULF Pc5 waves and the seed electrons. IMF Bz, Dst, and solar wind number density, velocity, and pressure did not improve any of the models, and were deemed unnecessary for effective predictions.

1. Introduction

Relativistic electron flux (>1.8-3.5 MeV) at geosynchronous orbit is influenced by a variety of factors. Ultra low frequency (ULF) Pc5 and very low frequency (VLF) waves have been postulated to accelerate seed electrons (270 keV) to relativistic energies (Jaynes et al., 2015; Rodger et al., 2015). Electromagnetic ion cyclotron (EMIC) waves are thought to precipitate these electrons (Rodger et al., 2008). The ring current index Dst, substorms, and variations in solar inputs such as solar wind velocity, number density, pressure, and interplanetary magnetic field (IMF) Bz have all also been postulated as influences on flux (Simms et al., 2018a, and references therein). Geomagnetic disturbances, during which many of these driving factors increase, can result in flux enhancements during the recovery phase, but only about half of storms result in a dramatic rise in electron flux. Levels can remain unchanged or fall following other storms (Kim et al., 2015; Reeves et al., 2003; Turner et al. 2013; Zhao & Li 2013), and the intensity of Dst during a storm is not sufficient to predict whether electron flux will increase or decrease (Reeves, 1998). Thus, identifying the further storm characteristics that lead to electron enhancement or depletion has been of interest (e.g., O'Brien et al., 2001; Pinto et al., 2018; Simms et al., 2014; Xiong et al., 2015).

Simple correlation or superposed epoch analysis has often been used to study these relationships. However, these approaches can only determine the association between pairs of factors (e.g., between flux and storm Dst). A correlation between a hypothesized predictor and flux may only mean that that possible predictor is itself highly correlated with the actual physical driver of flux enhancement or depletion. It may have no actual physical influence on its own. To avoid false conclusions, analysis methods that control for concomitant changes in all possible predictors are needed. Partial or canonical correlation techniques (e.g., Borovsky & Denton, 2014) can be used to avoid this problem, but multiple regression is better able to compare the relative effects of various interrelated possible predictors and to provide an empirical and tractable prediction equation in the form of a linear combination. However, because treatments (e.g., higher or lower wave activity) cannot be randomly assigned in observational studies such as this, statistically significant effects only prove an association between predictor and response, not a definite causal relationship.

Logistic regression does not predict values but rather the probability of an event. By classifying the response as a binary variable, and with the use of an appropriate transformation (the logistic transformation), regression can be used to produce a model that predicts the probability of occurrence of an event (Neter et al., 1985). We use logistic regression to model the probability of an outer belt electron flux enhancement (over pre-storm levels) following a geomagnetic storm. By adding predictor variables to this model we can determine which processes are the strongest predictors of an increased probability of flux enhancements occurring.

Previously, multiple regression analysis was used to predict relativistic electron flux levels following storms using solar wind and IMF parameters as well as ground-observed ULF and VLF waves (Simms et al., 2014). However, the models in this study did not incorporate either the occurrence of substorms or the presence of EMIC waves. In addition, the available VLF wave data from ground based sensors was only weakly associated with VLF waves occurring at geosynchronous orbit. Recently, a substorm occurrence measure, satellite-observed VLF wave intensity, and EMIC wave activity were all included in a model predicting daily averaged flux (Simms et al., 2018a). Our aim in this present study is to explore which set of the many possible factors best predicts whether the electron flux levels rise or fall following geomagnetic storms.

We compare relativistic electron flux post storm vs. pre-storm using three methods of analysis: 1) multiple regression to predict flux values following storms, 2) multiple regression to predict the size and direction of the change in electron flux, and 3) multiple logistic regression to predict only the probability of the flux rising or falling. Using daily averages from the first and second 24 hours of the recovery period (from minimum Dst until Dst reaches -30 nT), we trained all three model types on a set of predictors thought to have the most direct physical effect on flux: ULF Pc5, VLF lower band chorus (0.1-0.5 of the electron gyrofrequency), and EMIC waves, seed electron flux (270 keV), and Dark Ionosphere SME (SMEd) from the

SuperMAG collaboration as a measure of substorms. It was previously found that these parameters may be more influential on days following a geomagnetic storm than on the same day (Simms et al., 2018a). For this reason, we explore whether predictors are more influential during the first or second 24 hours of recovery following storms. We also use averages of these predictors from storm main phase. Because relativistic electron flux may show a non-linear response to ULF Pc5 waves (Simms et al., 2018b), we introduced a quadratic term (ULF Pc5)^2 to the model. In addition, we trained a model that included the added parameters of solar wind velocity (V), number density (N) and pressure (P), IMF Bz, and minimum storm Dst. This was intended, if possible, to produce a model with more predictive power due to the added explanatory variables.

2. Data

For the years 2005-2009, 126 geomagnetic storms were observed, determined from the Dst values obtained from the Omniweb database. A geomagnetic storm was defined as having a Dst minimum of -30 nT or lower.

As previously described in Simms et al., 2018a, we use the 1.8-3.5 MeV energy channel of relativistic electrons measured by the Energetic Spectrometer for Particles instrument (log10 [electrons/cm2 · s · sr · keV]) and the seed electron flux (270 keV in the same units as above) measured by the Synchronous Orbit Particle Analyzer (SOPA) instrument from the LANL satellites in geosynchronous orbit (Reeves et al., 2011). In one set of regressions, the maximum $\log_{10}$flux in the 7 days following the minimum Dst of storms was predicted. In a second set of regressions, the difference in the $\log_{10}$flux was predicted. This difference was calculated as post storm maximum $\log_{10}$flux (in the 7 days following the minimum Dst) – pre-storm $\log_{10}$flux (maximum flux on the day preceding the storm).

ULF Pc5 wave power was obtained from a ground-based ULF Pc5 index covering local times 0500–1500 in the Pc5 range (2–7 mHz) obtained from magnetometers stationed at 60°N–70°N corrected geomagnetic latitude (Kozyreva et al., 2007). In training the models described below, it was found that the ULF Pc5 did not have a completely linear relationship with the 1.8-3.5 MeV electron flux. Therefore, we also included $(\text{ULF Pc5})^2$ in our variable set.

We obtained VLF lower band chorus (log10 [$\mu V^2 \cdot m^2 \cdot Hz$]) power spectral density (0.1–0.5 fce; L=4, 4.0-4.99; dayside satellite passes, LT 10:30) from the Instrument Champ Electrique (ICE) on the DEMETER satellite (Berthelier et al., 2006).

EMIC wave activity data were obtained from the induction coil magnetometer located at the Halley, Antarctica, British Antarctic Survey (BAS) ground station at L-shell 4.6. We used the number of hours per day during which there was increased EMIC activity ($>10^{-3}$ nT^2 Hz) in the <1-Hz band.

The SMEd, a measure of only the dark ionosphere (nightside) SuperMAG Auroral Electroject Indices (SME) was obtained from SuperMAG (Gjerloev et al., 2010; Gjerloev, 2012).

From the Omniweb database, we obtained solar wind velocity (V), number density (N), IMF Bz component, Dst, and pressure (P). Each of these was averaged over the main and recovery phases as described above, with the exception of Bz, for which we used the fraction of southward Bz hours out of total hours in each time period.

For the years 2005–2009, we averaged all variables (except when noted) over storm main phase and the first and second 24 hours of recovery. This daily averaging was done to smooth out diurnal fluctuations that occur due to satellite position. Of the 126 geomagnetic storms observed during this time period, only 85 were complete observations (i.e., containing measured values for all parameters per observation) that could be used in the analyses.

3. Methods

Statistical analyses were performed using R and MATLAB.

Three model types were tested:

1. Multiple regression predicting the value of the 1.8-3.5 MeV electron maximum flux at geostationary orbit in the 7 days following the minimum Dst of geomagnetic storms.
2. Multiple regression predicting the log flux difference (pre-storm vs. post storm as defined in the previous section)
3. Logistic regression predicting the probability of a flux increase (post storm higher than pre-storm).

We consider Model Type 1 to be a baseline model (a standard regression model predicting values) and explore the other model types with the hope that they will improve on this model. All three models are expected to show the same general relationship between the explanatory and dependent variables.

The data were randomly split into a 60%:40% ratio of training and test sets (~51:34 storms). For Model Type 1, a linear multiple regression model was created using observations from the training set. Predictions from the test set were calculated using the unstandardized model coefficients for each explanatory variable. A validation correlation coefficient *r* calculated between these predictions and the real value of the electron flux from the test set was used to determine the best model. A larger value of *r* means the corresponding model predicts the test set electron flux better.

The algorithms for Model Type 2 (multiple regression) and 3 (logistic regression) were similar, however rather than predicting the value of the maximum electron flux, these models used the flux difference as a response variable (as described above). Model Type 2 simply predicted the difference between pre and post storm flux log values. However, for Model Type 3, due to the use of logistic regression, we require a binary dependent variable. To produce this, we set all increases in log flux (post storm higher than pre-storm) to 1 and all non-increases to 0. However, this binary response variable will not have a linear relationship with the predictors as the response can only be at the bottom (0) or top (1) of its range. In order to fulfill the linearity requirements of regression, the binary variable must be transformed. This is accomplished in several steps. First, a probability prediction function is assumed, where the probabilities of "success" (p, response=1) and "failure" (1-p, response=0) sum to 1 for any single trial (the discrete Bernoulli distribution). This probability function is not, itself, linear. Although probability responses can now span the range between 0 and 1, the range is still restricted, responses asymptote curvilinearly to 0 and 1, and none of the usual transformations of the data (logs, etc.) will fix this problem. A further transformation is needed to linearize the response and transform the range from negative to positive infinity. This can be accomplished by using the odds instead of the probability. (While probability is the ratio of successes to all trials, the odds are the ratio of successes to failure.) Taking the log of the odds (called the logits) is then a simple transformation to produce a linear function (Neter et al., 1985). Mathematically, this transformation to logits is accomplished via the logistic transformation of the probability π:

$$logit = \log_e\left(\frac{\pi}{1-\pi}\right). \tag{1}$$

The coefficients of the prediction equation are calculated from these observed logits using a nonlinear, iterative process that finds the maximum likelihood estimates for these parameters. The resulting logistic regression equation then predicts logits (log odds) using the fitted model coefficients:

$$\mathrm{logit} = \mathrm{b}_0 + b_1 x_1 + \cdots + b_i x_i \tag{2}$$

which can then be converted to predicted probabilities (probability of log flux increase) with the following back transformation:

$$Pr(event) = \frac{\mathrm{e}^{\mathrm{b}_0 + b_1 x_1 + \cdots + b_i x_i}}{1+e^{\mathrm{b}_0 + b_1 x_1 + \cdots + b_i x_i}}, \tag{3}$$

where $\mathrm{Pr}(event)$ is the predicted probability that an event will occur (in this case a flux increase), x_i refers to the i^{th} predictor and b_i refers to the corresponding coefficient. The predicted probabilities will differ depending on the particular values of the predictors (x_i).

The typical logistic regression algorithm performed on this data set had difficulty converging on model coefficients. It is possible that the data set was too small. Instead, we used a Firth logistic regression, which uses a penalized likelihood method, and often produces a more successful result (Firth, 1993).

Probability predictions from Model Type 3 (logistic regression) ranged from 0 to 1, where 0 was a zero probability of the electron flux going up after a storm, and 1 was a 100% probability of it going up. These were also converted to binary, with probabilities greater than 0.5 being classified as 1 and probabilities less than 0.5 as 0. This allowed a cross tabulation between the predicted and real binary values to be made.

The cross tabulation produces four numbers: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). A true positive refers to an observation that is correctly predicted as an increase in electron flux. A false positive refers to an observation that is incorrectly predicted as an increase in electron flux, when in fact the real data reflected a decrease. The true negative and false negative follow similarly. These values are used to calculate true positive rate (TPR), true negative rate (TNR), and accuracy (ACC), as follows (Fawcett, 2005):

$$TPR = \frac{TP}{TP+FN} \tag{4}$$

$$TNR = \frac{TN}{TN+FP} \tag{5}$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

Model Type 2 used a multiple regression (as in Model Type 1) but now predicting the flux difference. To compare the results with Model Type 3 (logistic), the flux difference from the training set and the predicted flux difference values were converted to binary as in the

Model Type 3 algorithm, and a cross tabulation made. A validation correlation coefficient *r* was also calculated for Model Type 2 as with Model Type 1.

For all 3 model types, 1000 models were trained on 1000 unique, randomly sampled training and test sets, and the measurements obtained were averaged over all 1000 runs. This method, sometimes called bootstrap aggregating or "bagging", can improve the accuracy of a prediction method by providing stability to the training sets (Breiman, 1996). Training single models and comparing the results showed an apparent inhomogeneity in the training sets. Given this, and the small number of data points (85 usable storms), bagging was used to provide averaged metrics over many different subsamples.

Finally, inspired by Table A1 of O'Brien and McPherron (2003), a table of significance frequencies was also made for each explanatory variable. In each run, a p-value < 0.05 was marked as statistically significant. The p-value is the probability that a particular predictor would show an apparent influence in the regression model when it, in fact, had no association with the response variable at all. In other words, it is the probability of mistakenly believing there is an association when there is not. Using this definition, the frequency of significance is reported for each predictor over the 1000 runs. Obviously, with this number of runs, the actual overall p-value is no longer 5% (i.e., it is not the probability of rejecting the null hypothesis of no association over all the runs), but we use it as a convenient cut off of association vs. no association within each run for the purposes of compiling statistics.

3.2 Variable Sets and Effect of Variable Time Periods

With the electron flux as the dependent variable, a backward elimination stepwise regression procedure was used to select predictor variables, at each step removing the predictor with the least significant p-value and recalculating the regression. This resulted in a more parsimonious but still effective model containing only significant predictors with p-values<0.05 (Neter et al, 1985). The stepwise regression procedure was given only the direct drivers of the electron flux to select from. The resulting variable set included ULF Pc5, ULF Pc5^2, VLF, seed electron flux (270 keV channel), SMEd, pre-storm electron flux (1.8-3.5 MeV channel), and EMIC waves.

Once this variable set was selected, we explored the effect of the two different time periods, the first and the second 24 hours of recovery after a storm, for each of the variables that were measured following storms. This included 6 of the 7 variables (pre-storm electron flux obviously being measured only before storms), resulting in $2^6 = 64$ total combinations of first and second 24 hours of recovery (2 for the number of time period options, 6 for the number of variables with those options). Each of these combinations also included pre-storm electron flux as a predictor. All 64 time period combinations were tested using Model Type 1 (other investigations shown in Figure 1 demonstrated that these time combinations were very similar for Model Type 2). The best of these were determined by finding the top ten highest validation correlation coefficients, which were all above 0.75.

We also determined whether certain variables were more influential on either the first or second 24 hours of recovery. We gathered all models where the first 24 hours of recovery for the first variable was used, all models where the second 24 hour period for the first variable was used, all models where the first 24 hour period of recovery for the second variable was used, and so forth, giving us 12 groups of models. We then counted the number of the top ten best models previously determined that were in each group (Table 1).

Finally, using the top 5 time period combinations determined from Model Type 1, we ran all three model types using a larger predictor set. This set included all the same variables as before, as well as V, N, Bz ratio, Dst, and pressure. The different time period combinations produced very similar results in Model Type 2, therefore the top time period combinations from Model Type 1 were used for the other models. This is demonstrated in Figure 1, where the top 10 and bottom 5 time period combinations found with Model Type 1 follow the same trend in Model Type 2.

4. Results

The goal of these analyses was to determine an effective predictive model to forecast the relativistic electron flux and also to explore some general trends in building such models, such as the time period combinations for each variable, model type, and variable sets. Once the best model was selected, we examined the standardized coefficients of each variable in the regression to see which variables are the most influential.

4.1 Determining the Best Model

4.1.1 Time periods and Model Type

We sorted the validation correlation coefficients from the Model Type 1 runs from largest to smallest and sampled a smaller number of combinations to run with the other two model types. Because we did not find that the overall trends differed greatly, we did not run all

64 for Model Types 2 and 3. Figure 1 shows *r* values for the top ten and bottom five combinations across Model Types 1 and 2.

All 3 model types were run using the time period combination with the highest validation correlation coefficient found in Figure 1a. This combination used most variables measured in the first 24 hours of recovery, with the exception of the VLF. We find that Model Type 2 (linear regression predicting flux difference) produces the best prediction, as seen by the *r*, TPR, TNR, and ACC values across model types in Figure 2b. This is also evident in Figure 1, as Model Type 2 consistently has a higher *r* relative to Model Type 1. In Figure 2a, the pre-storm flux is seen to be significant (p-value < 0.05) in every run of Model Type 2. This is unsurprising, as pre-storm flux is used in calculating the flux difference, but its addition as a covariate improves the model. Otherwise, the significance frequencies between Model Type 1 and 2 are very similar, indicating that similar variables are influencing the two model types equally. We are unable to compare Model Type 1 and 2 by the crosstab metrics, however, the validation correlation coefficient *r* is larger in Model Type 2 (0.9 vs. 0.8).

While the significance frequencies for Model 3 are very low for nearly all variables compared to those for Models 1 and 2 (Figure 2a), this is not necessarily an indication that Model 3 is an ineffective analysis method. The important tests of the models' effectiveness are the crosstab metrics. If the model can predict flux changes well (indicated by good crosstab metrics), poor significance frequencies do not necessarily matter. The crosstab metrics for Model 3 are slightly weaker than for Model 2 (Figure 2b), indicating that Model 2 is a somewhat more effective model.

We conclude that Model Type 2 is a more effective predictive model for these data and this variable set. However, Model Type 1 offers the advantage of predicting the actual values of the electron flux, whereas Model Type 2 predicts the magnitude of the flux increase or decrease. Depending on the application of these models, one or the other type may be more useful despite the difference in validation correlation coefficients.

In the process of determining the best time period combinations for a predictive model, we observed trends in the ideal time periods for each variable. We found that several, though not all, of the variables were significant mostly in either the first or the second 24 hours of recovery. For example, 7 of the top 10 combinations had ULF Pc5 in the first 24 hour time period, suggesting that is the best time period to use for that variable. This information is shown in Table 1.

We found that VLF is significant more in the second day of recovery (8 of the 10 best models of Table 1). ULF Pc5, seed electrons, and EMIC waves are all mostly significant in the first 24 hours of recovery. The seed electrons were especially clear, as all 10 of the best models used data measured in the first 24 hours of recovery. The SMEd and the squared ULF Pc5 were most influential in the first 24 hours about half the time.

Table 1: *Number of times the first or second 24 hours of recovery time periods were used for each variable in the top 10 best time period combinations (using Model Type 1).*

Variable	First 24 hours of recovery	Second 24 hours of recovery
ULF Pc5	**7**	3
ULF $Pc5^2$	5	5
VLF	2	**8**
Seed electrons	**10**	0
SMEd	6	4
EMIC	**8**	2

4.1.2 Variable sets

More variables may make for a better model as they provide more information for the algorithm to work with. We ran all three model types again, this time including five more variables: V, N, Bz ratio, Dst, and pressure, all measured in the first 24 hours of recovery. All other variables are based on observations during the first 24 hour of recovery, except for the VLF waves, which uses the second 24 hour period. The significance frequencies are shown in Figure 3a, and *r* values, and cross tabulation measurements are shown in Figure 3b. Comparing to Figure 2 in Section 4.1.1, we do not see a significant improvement in the models from adding more variables. Both the *r* values and the cross tabulation measurements are virtually the same for both variable sets in each model type. Evidently, these additional variables are unnecessary for an effective predictive model.

4.1.2 Low Success with Logistic Regression

We see in Figures 2a and 3a that the significance frequencies for Model Type 3 (logistic regression) are quite low – some variables have values of zero, and all others are ≤ 0.16 in the smaller variable set (Figure 2), and ≤ 0.21 in the larger variable set (Figure 3) (pre-storm flux is considered a covariate, and as such its significance frequency value of 1 in both figures is not noteworthy). As previously stated, these low significance frequencies do not necessarily indicate that this model is less effective, however we do see slightly lower crosstab measures in Model 3 as compared to Model 2.

Whereas we have defined an increase or decrease in the electron flux to be any change between pre-storm or post storm flux, Reeves et al. (2003) required a relative change of a factor of at least 2. This corresponds to a cutoff of 0.3 with our log flux values. With this definition, storms during which the electron flux increased or decreased only very slightly are classified as having had no change in flux. Using this 0.3 cutoff rather than our original cutoff of 0 for Model Types 2 and 3, we find a slight improvement in the validation correlation coefficient *r* and crosstab measures, as shown in Table 2.

This new 0.3 cutoff does improve the logistic regression, but the improvement is slight, and additionally, the new cutoff decreased the predictive effectiveness of Model Type 2 in all measurements but the TNR.

Table 2: *Model Types 2 and 3 (linear predicting flux difference, and logistic) results using an electron flux cutoff of 0 and a cutoff of a relative change of a factor of at least 2. The variables*

used were ULF Pc5, ULF Pc5^2, VLF, seed electrons, SMEd, pre-storm flux, and EMIC waves, all measured in the first 24 hours of recovery but the VLF.

Cutoff (log values):	Model Type 2				Model Type 3			
	r	**TPR**	**TNR**	**ACC**	***r***	**TPR**	**TNR**	**ACC**
0	0.91	0.94	0.86	0.92	0.70	0.90	0.82	0.88
0.3	0.90	0.89	0.88	0.89	0.77	0.93	0.84	0.90

4.2 Standardized Coefficients of the Best Model

Having determined the best model to be Model Type 2 (linear regression predicting flux difference, all variables measured in the first 24 hours of recovery except for VLF waves which uses observations in the second 24 hours, and no solar wind parameters), we next calculated the standardized model coefficients (Figure 4). These allow a direct comparison of predictor influences, regardless of differing scales. When including the ULF Pc5^2, the ULF Pc5 shows the strongest influence on electron flux. If the squared term is dropped, ULF Pc5 still has a stronger influence than VLF, SMEd, and EMIC, and similar influence as the seed electrons.

5. Discussion

Geomagnetic disturbances have been associated with relativistic electron flux enhancements during the recovery phase, due in part to the resulting increases in the parameters that are thought to drive electron flux increases. However, not all storms result in appreciable increases in electron flux (Kim et al., 2015; Reeves et al., 2003; Turner et al. 2013; Zhao & Li 2013). Furthermore, we cannot predict the behavior of the electron flux using merely the intensity of the Dst index during a storm (Reeves, 1998). Further parameters are necessary to effectively predict relativistic electron flux. We find that ULF Pc5 waves and seed electrons are the most influential variables in predicting electron flux at geosynchronous orbit, with lower but still observable effects of VLF and EMIC waves. (Flux enhancements have also been observed at altitudes lower than geosynchronous orbit, where they may be driven by different mechanisms than suggested by our present study (Katsavrias et al., 2019)). In our study, we explore prediction from three analysis types: prediction of flux values using regression, prediction of the change in flux using regression, and predicting the likelihood of a flux increase using logistic regression. Logistic regression is a simple classifier model, as it predicts probabilities of observations belonging to a class (in this case, an increase in flux following a storm). Neural networks are more complex examples of classifier models, and several previous studies have utilized neural networks to predict levels of these electrons (O'Brien & McPherron, 2003; Perry et al., 2010). Neural networks can model non-linear, often very complex data, and then predict outcomes from new data using those models. However, given the "black box" nature of these methods, it is difficult to infer physical meaning from the results. If the goal is to learn which physical processes influence electron levels, it is better to use methods such as regression or logistic regression, which provide valuable information on the relative strength of influence of each variable.

As in this paper, Simms et al. (2014) looked at only storm times (removing the quiet periods from the data set). Their analyses showed that ULF Pc5 and seed electrons were influential, similar to what we have found here. However, the VLF data in their work was from ground stations, and they did not find it had good predictive ability. VLF data from ground stations is subject to transionospheric attenuation during periods of solar illumination.

Therefore, ground based VLF measurements are not necessarily representative of what is happening at the altitude of the satellite (Simms et al., 2015; Smith et al., 2010). In this paper, we have a space-based VLF measure from the DEMETER satellite, and this does show good predictive ability (Simms et al., 2019).

Previous work used the AE index as a measure of substorm activity. It was not effective at predicting relativistic electron flux and may not have been a good measure of substorm activity (Simms et al., 2014). In this paper, we used the SMEd index. This is also measured at ground-based magnetometers, but the data comes only from the dark ionosphere (nightside) which would be better able to measure substorm activity from the tail of the magnetosphere. The SMEd also incorporates observations from a wider range of magnetic latitudes and from a much larger number of magnetometer stations than does the AE (Newell & Gjerloev, 2011). Despite this change, waves, particularly ULF Pc5 and VLF, were more effective predictors of relativistic electron flux than substorm activity.

Model Type 2 (linear regression predicting flux difference) was the most effective at predicting the size of electron flux increases. The validation correlation coefficient *r* was larger for this model type than for Model Type 1 (linear regression predicting flux value). The crosstab measures (TPR, TNR, and ACC) were also higher for this model type than for Model Type 3 (logistic regression). The logistic regression used in Model Type 3 had weaker predictive abilities than the other two model types. However, in some circumstances we may want a prediction of the probability of the electron flux increasing after a storm rather than a prediction of the actual value. In this case, Model Type 3 may be the most useful model. These considerations should be taken into account, along with the validation correlation coefficients and crosstab values, when selecting a model type.

The most effective models used waves, seed electrons, and substorm activity. In this predictor set, all but the VLF were measured in the first 24 hours of recovery. The VLF was measured in the second 24 hours of recovery. This time period combination produced the highest validation correlation coefficient and crosstab measures. While ULF Pc5, seed electrons, and EMIC waves were more effective predictor variables when measured during the first 24 hours of recovery, and VLF when measured during the second 24 hours of recovery, the period of measurement was not important for SMEd.

The inclusion of additional parameters (V, N, Bz ratio, Dst, and pressure) did not produce significant improvement, and we did not include them in our final model. The strong correlations of these variables with flux enhancements seen in previous work, together with their apparent redundancy in our models, likely indicates that solar wind and IMF influences are mediated through the driving of waves and seed electrons which then directly influence flux levels.

Significance frequencies and the standardized coefficients for each variable in the final model show which variables are more frequently statistically significant and have higher influence, respectively. ULF wave power and the seed electrons are the most frequently significant of the possible predictor variables. ULF Pc5 waves and seed electrons are also the strongest influences on flux changes as measured by the standardized regression coefficients. Pre-storm flux also shows a high significance frequency and influence, but this is only because it was used to calculate the flux difference (the response variable).

As in Simms et al. (2018a), which looked at daily averages of parameters over the entire year, we again found that the effect of ULF Pc5 as determined by the standardized coefficients was stronger than that of the VLF. However, both the ULF Pc5 and the VLF are important predictors, presumably because they are accelerating seed electrons, which were also associated with increased flux. This supports the argument made in Simms et al. (2018a) that both act independently to increase electron flux levels, rather than only the ULF Pc5 (suggested by Ozeke et al., 2017), or only the VLF (suggested by Jaynes et al., 2015). The ULF Pc5 influence, however, is nonlinear, showing the strongest effect at mid-range values. The positive linear and negative squared ULF Pc5 terms together describe this peak as a quadratic response of flux (Simms et al., 2018b). When the ULF $Pc5^2$ term is not included, the standardized coefficient and influence of the ULF Pc5 is more similar to that of the VLF (Figure 4). The decreased influence of ULF Pc5 at higher values may be related to the hypothesized electron loss during shock events due to outward radial diffusion (Brautigam & Albert, 2000; Degeling et al., 2008; Hudson et al., 2014; Loto'aniu et al., 2010; Shprits et al., 2006; Ukhorskiy et al., 2009; Zong et al., 2012).

A high pre-storm flux has little room to grow substantially, thus a large change in flux will not occur if flux is already high. However, if flux before a storm is low, there could be a substantial increase. This appears as a negative correlation between pre and post storm flux.

There was a negative effect of the EMIC waves due to presumed precipitation (Figure 4). The substorm measure did not show a significant direct effect on the electron flux, although we did not test whether it had an indirect effect through the production of VLF waves.

Because this is not a controlled experiment with randomly assigned treatments, we cannot necessarily interpret the significant p-values as implying causation. However, these correlations support the idea that there is a possible causal relationship between variables we have identified as predictors and the rise or fall of relativistic electron flux.

6. Conclusions

1. Following storms, increases in relativistic electron flux at geosynchronous orbit were well predicted by three regression models: 1) multiple regression to predict flux values following storms, 2) multiple regression to predict the size and direction of the change in electron flux, and 3) multiple logistic regression to predict only the probability of the flux rising or falling.
2. The ULF Pc5 waves and seed electrons were the most influential predictors. Additionally, the VLF and EMIC waves were also influential. Including the IMF Bz, Dst, and solar wind number density, velocity, and pressure in the data set did not improve any of the models.
3. The three model types had similar validation success, but Model 2 (linear predicting flux difference) was determined to be the most effective.

Acknowledgments

We thank M. Ohnsted for EMIC wave identification, and R. Gamble for preparing and J.-J. Berthelier for providing DEMETER ICE data. Relativistic and seed electron flux data were obtained from Los Alamos National Laboratory (LANL) geosynchronous energetic particle instruments (https://www.ngdc.noaa.gov/stp/space-weather/satellite-data/satellite-systems/lanl_geo/)). The ULF Pc5 index is available at http://ULFPc5.gcras.ru/. The SMEd index is available from SuperMAG (http://supermag. jhuapl.edu/, Principal Investigator Jesper Gjerloev), derived from magnetometer data from Intermagnet; USGS, Jeffrey J. Love; CARISMA, PI Ian Mann; CANMOS; the S-RAMP Database, PI K. Yumoto and K. Shiokawa; the SPIDR database; AARI, PI Oleg Troshichev; the MACCS program, PI M. Engebretson, Geomagnetism Unit of the Geological Survey of Canada; GIMA; MEASURE, UCLA IGPP, and Florida Institute of Technology; SAMBA, PI Eftyhia Zesta; 210 Chain, PI K. Yumoto; SAMNET, PI Farideh Honary; the institutes who maintain the IMAGE magnetometer array, PI Eija Tanskanen; PENGUIN; AUTUMN, PI Martin Connors; DTU Space, PI Rico Behlke; South Pole and McMurdo Magnetometer, PIs Louis J. Lanzerotti and Alan T. Weatherwax; ICESTAR; RAPIDMAG; PENGUIn; British Artarctic Survey; McMAC, PI Peter Chi; BGS, PI Susan Macmillan; Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN); GFZ, PI Juergen Matzka; MFGI, PI B. Heilig; IGFPAS, PI J. Reda; University of L'Aquila, PI M. Vellante. IMF Bz, Dst, and solar wind V, N, and P data are available from Goddard Space Flight Center Space Physics Data Facility at the OMNIWeb data website (http://omniweb.gsfc.nasa.gov/html/ow_data.html). Work at Augsburg University was supported by NSF grant AGS- 1651263.

6. References

Berthelier, J. J., Godefroy, M., Leblanc, F., Malingre, M., Menvielle, M., Lagoutte, D., et al. (2006). ICE, the electric field experiment on DEMETER. *Planetary and Space Science*, 54(5), 456-471. https://doi.org/10.1016/j.pss.2005.10.016

Brautigam, D. H., & Albert, J. M. (2000). Radial diffusion analysis of outer radiation belt electrons during the October 9, 1990, magnetic storm. *Journal of Geophysical Research*, 105(A1), 291-309. https://doi-org.ezp1.lib.umn.edu/10.1029/1999JA900344

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. https://doi.org/10.1007/BF00058655

Borovsky, J. E. & Denton, M. H. (2014). Exploring the cross correlations and autocorrelations of the ULF Pc5 indices and incorporating the ULF Pc5 indices into the systems science of the solar wind-driven magnetosphere. *Journal of Geophysical Research*, 119, 4307-4334. doi:10.1002/2014JA019876

Degeling, A. W., Ozeke, L. G., Rankin, R., Mann, I. R., & Kabin, K. (2008). Drift resonant generation of peaked relativistic electron distributions by Pc 5 ULF waves. *Journal of Geophysical Research*, 113, A02208. https://doi-org.ezp1.lib.umn.edu/10.1029/2007JA012411

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874. doi:10.1016/j.patrec.2005.10.010

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 80(1), 27-38. doi:10.2307/2336755

Gjerloev, J. W. (2012). The SuperMAG data processing technique. *Journal of Geophysical Research*, 117, A09213. https://doi.org/10.1029/2012JA017683

Gjerloev, J. W., Hoffman, R. A., Ohtani, S., Weygand, J., & Barnes, R. (2010). Response of the Auroral Electrojet Indices to Abrupt Southward IMF Turnings. *Annales Geophysicae*, 28, 1167-1182. doi: 10.5194/angeo-28-1167-2010

Hudson, M. K., Baker, D. N., Goldstein, J., Kress, B. T., Paral, J., Toffoletto, F. R., & Wiltberger, M. (2014). Simulated magnetopause losses and Van Allen Probe flux dropouts. *Geophysical Research Letters*, 41, 1113-1118. https://doi-org.ezp1.lib.umn.edu/10.1002/2014GL059222

Jaynes, A. N., Baker, D. N., Singer, H. J., Rodriguez, J. V., Loto'aniu, T. M., & Ali, A. F., et al. (2015). Source and seed populations for relativistic electrons: Their roles in Radiation Belt changes. *Journal of Geophysical Research: Space Physics*, 120, 7240-7254. https://doi.org/10.1002/2015JA021234

Katsavrias, C., Sandberg, I., Li, W., Podladchikova, O., Daglis, I. A., Papadimitriou, C., et al. (2019). Highly relativistic electron flux enhancement during the weak geomagnetic storm of April-May 2017. *Journal of Geophysical Research: Space Physics*, 124. https://doi.org/10.1029/2019JA026743

Kim, H.-J., Lyons, L., Pinto, V., Wang, C.-P., & Kim, K.-C. (2015). Revisit of relationship between geosynchronous relativistic electron enhancements and magnetic storms. *Geophysical Research Letters*, 42, https://doi.org/10.1002/2015GL065192

Kozyreva, O., Pilipenko, V., Engebretson, M. J., Yumoto, K., Watermann, J., & Romanova, N. (2007). In search of a new ULF Pc5 wave index: Comparison of Pc5 power with dynamics of geostationary relativistic electrons. *Planetary and Space Science*, 55, 755-769. https://doi.org/10.1016/j.pss.2006.03.013

Loto'aniu, T. M., Singer, H. J., Waters, C. L., Angelopoulos, V., Mann, I. R., Elkington, S. R., & Bonnell, J. W. (2010). Relativistic electron loss due to ultralow frequency waves and enhanced outward radial diffusion. *Journal of Geophysical Research*, 115, A12245. doi:10.1029/2010JA015755

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models*. Homewood, Ill: Richard D. Irwin, Inc.

Newell, P. T., & Gjerloev, J. W. (2011). Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. *Journal of Geophysical Research*, 116, A12211. https://doi.org/10.1029/2011JA016779

O'Brien, T. P., McPherron, R. L., Sornette, D., Reeves, G. D., Friedel, R., & Singer, H. J. (2001). Which magnetic storms produce relativistic electrons at geosynchronous orbit?. *Journal of Geophysical Research*, 106(A8), 15,533-15,544. https://doi.org/10.1029/2001JA000052

O'Brien, T. P. & McPherron, R. L., (2003). An empirical dynamic equation for energetic electrons at geosynchronous orbit. *Journal of Geophysical Research*, 108(A3), doi:10.1029/2002ja009324

Ozeke, L. G., Mann, I. R., Murphy, K. R., Sibeck, D. G., & Baker, D. N. (2017). Ultra-relativistic radiation belt extinction and ULF Pc5 wave radial diffusion: Modeling the September 2014 extended dropout event. *Geophysical Research Letters*, 44, 2624-2633. https://doi.org/10.1002/2017GL072811

Perry, K. L., Ginet, G. P., Ling, A. G., & Hilmer, R. V. (2010). Comparing geosynchronous relativistic electron prediction models. *Space Weather*, 8(12), doi:10.1029/2010sw000581

Pinto, V. A., Kim, H.-J., Lyons, L. R., & Bortnik, J. (2018). Interplanetary parameters leading to relativistic electron enhancement and persistent depletion events at geosynchronous orbit and potential for prediction. *Journal of Geophysical Research: Space Physics*, 123. https://doi.org/10.1002/2017JA024902

Reeves, G. D., Morley, S. K., Friedel, R. H. W., Henderson, M. G., Cayton, T. E., & Cunningham, G., et al. (2011). On the relationship between relativistic electron flux and solar wind velocity: Paulikas and Blake revisited. *Journal of Geophysical Research*, 116, A02213. https://doi.org/10.1029/2010JA015735

Reeves, G. D., McAdams, K. L., Friedel, R. H. W., & O'Brien, T. P. (2003). Acceleration and loss of relativistic electrons during geomagnetic storms. *Geophysical Research Letters*, 30(10), 1529, doi:10.1029/2002GL016513

Reeves, G. D. (1998). Relativistic electrons and magnetic storms: 1992-1995. *Geophysical Research Letters*, 25(11), 1817-1820, doi:10.1029/98gl01398

Rodger, C. J., Raita, T., Clilverd, M. A., Seppälä, A., Dietrich, S., Thomson, N. R., & Ulich, T. (2008). Observations of relativistic electron precipitation from the radiation belts driven by EMIC waves. *Geophysical Research Letters*, 35, doi:10.1029/2008GL034804

Rodger, C. J., Cresswell-Moorcock, K., & Clilverd, M. A. (2016). Nature's Grand Experiment: Linkage between magnetospheric convection and the radiation belts. *Journal of Geophysical Research: Space Physics*, 121, 171-189. doi:10.1002/2015JA021537

Shprits, Y. Y., Thorne, R. M., Friedel, R., Reeves, G. D., Fennell, J., Baker, D. N., & Kanekal, S. G. (2006). Outward radial diffusion driven by losses at magnetopause. *Journal of Geophysical Research*, 111, A11214. doi:10.1029/2006JA011657

Simms, L. E., Engebretson, M. J., Clilverd, M. A., Rodger, C. J., Lessard, M., Gjerloev, J., & Reeves, G. D. (2018a). A Distributed Lag Autoregressive Model of Geostationary Relativistic Electron Fluxes: Comparing the Influences of Waves, Seed and Source Electrons, and Solar Wind Inputs. *Journal of Geophysical Research: Space Physics*, 123(5), 3646-3671, doi:10.1029/2017ja025002

Simms, L. E., Engebretson, M. J., Clilverd, M. A., Rodger, C. J., & Reeves, G. D. (2018b). Nonlinear and synergistic effects of ULF Pc5, VLF chorus, and EMIC waves on relativistic electron flux at geosynchronous orbit. *Journal of Geophysical Research: Space Physics*, 123. https://doi.org/10.1029/2017JA025003

Simms, L. E., Pilipenko, V. A., Engebretson, M. J., Reeves, G. D., Smith, A. J., & Clilverd, M. A. (2015). Analysis of the effectiveness of ground-based VLF wave observations for predicting or nowcasting relativistic electron flux at geostationary orbit. *Journal of Geophysical Research: Space Physics*, 120, 2052-2060. https://doi.org/10.1002/2014JA020337

Simms, L. E., Pilipenko, V., Engebretson, M. J., Reeves, G. D., Smith, A. J., & Clilverd, M. A. (2014). Prediction of relativistic electron flux at geostationary orbit following storms: Multiple regression analysis. *Journal of Geophysical Research: Space Physics*, 119(9), 7297-7318. doi:10.1002/2014ja019955

Simms, L. E., Engebretson, M. J., Clilverd, M. A., & Rodger, C. J. (2019). Ground-based observations of VLF waves as a proxy for satellite observations: Development of models including the influence of solar illumination and geomagnetic disturbance levels. *Journal of Geophysical Research: Space Physics*, 124, 2682-2696. https://doi.org/10.1029/2018JA026407

Smith, A. J., Horne, R. B., & Meredith, N. P. (2010). The statistics of natural ELF/VLF waves derived from a long continuous set of ground-based observations at high latitude. *Journal of Atmospheric and Terrestrial Physics*, 72, 463-475. https://doi.org/10.1016/j.jastp.2009.12.018

Turner, D. L., Angelopoulos, V., Li, W., Hartinger, M. D., Usanova, M., Mann, I. R., Bortnik, J., & Shprits, Y. (2013). On the storm-time evolution of relativistic electron phase space density in Earth’s outer radiation belt. *Journal of Geophysical Research: Space Physics*, 118, 2196-2212. https://doi.org/10.1002/jgra.50151

Ukhorskiy, A. Y., Sitnov, M. I., Takahashi, K., & Anderson, B. J. (2009). Radial transport of radiation belt electrons due to stormtime Pc5 waves. *Annales de Geophysicae*, 27(5), 2173-2181. https://doi.org/10.5194/angeo-27-2173-2009

Xiong, Y., Xie, L., Pu, Z., Fu, S., Chen, L., Ni, B., Li, W., Li, J., Guo, R., & Parks, G.K. (2015). Responses of relativistic electron fluxes in the outer radiation belt to geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 120, 9513-9523. doi:10.1002/2015JA021440

Zhao, H., & Li, X. (2013). Inward shift of outer radiation belt electrons as a function of Dst index and the influence of the solar wind on electron injections into the slot region. *Journal of Geophysical Research: Space Physics*, 118, 756-764. https://doi.org/10.1029/2012JA018179

Zong, Q. G., Wang, Y. F., Zhang, H., Fu, S. Y., Zhang, H., & Wang, C. R., et al. (2012). Fast acceleration of inner magnetospheric hydrogen and oxygen ions by shock induced ULF waves. *Journal of Geophysical Research*, 117, A11206. doi:10.1029/2012JA018024

Figure 1: *Validation correlation coefficient* r *across Model Types 1 and 2 (linear regressions) for the top 10 and bottom 5 Model Type 1 time period combinations. Each bar's label shows a sequence of 1's and 2's for the first or second 24 hours of recovery for variables in the following order: ULF Pc5, ULF Pc5*^{2}*, VLF, seed electrons, SMEd, EMIC. Pre-storm electron flux is also included in these models.*

Figure 2: *Significance frequencies, r, and crosstab values for Model types 1, 2, and 3 for the best time-period model by* r *(all variables are measured in the first 24 hours of recovery except the VLF).*

Figure 3: *Significance frequencies,* r*, and crosstab values for Model types 1, 2, and 3 for the best time-period model by* r *(all variables are measured in the first 24 hours of recovery except VLF).*

Figure 4: *Standardized model coefficients for Model Type 2 (linear predicting flux difference) over the entire data set and best time period (all variables measured in the first 24 hours of recovery, except VLF). Dark gray shows standardized coefficients for the model with ULF Pc5^2, and light gray shows standardized coefficients for the model without ULF Pc5^2. All variables were statistically significant (*$p < 0.05$*), with the exception of SMEd.*