JGR Space Physics

TECHNICAL REPORTS: METHODS
10.1029/2021JA029198

Key Points:

- We developed an automated technique for step-like pressure front detection
- It uses nonlinear least squares fitting of an analytical step-like function
- We demonstrate it with an application on the response of the transpolar potential to solar wind dynamic pressure fronts

Correspondence to:
A. Boudouridis,
thanasis@spacescience.org

Citation:
Boudouridis, A., & Zesta, E. (2021). Automated technique for the detection of step-like solar wind dynamic pressure changes: Application to the response of the transpolar potential to solar wind dynamic pressure fronts. *Journal of Geophysical Research: Space Physics*, *126*, e2021JA029198. https://doi.org/10.1029/2021JA029198

Received 30 JAN 2021
Accepted 11 JUN 2021

Automated Technique for the Detection of Step-Like Solar Wind Dynamic Pressure Changes: Application to the Response of the Transpolar Potential to Solar Wind Dynamic Pressure Fronts

A. Boudouridis[1,2] **and E. Zesta**[3]

[1]Space Sciences Institute, Boulder, CO, USA, [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [3]NASA Goddard Space Flight Center, Greenbelt, MD, USA

Abstract The input from the solar wind (SW) and the accompanying Interplanetary Magnetic Field (IMF) is the defining feature of the solar wind-magnetosphere interactions. One of the most important SW drivers is the solar wind dynamic pressure, and most significantly its sharp changes known as pressure fronts. Sudden enhancements in solar wind dynamic pressure, P_{SW}, have been shown to lead to intensification of the auroral electrojets (and Region 1 field-aligned currents), increase in auroral emissions, and significant enhancement in ionospheric convection and the transpolar potential. Case studies of the interaction between pressure fronts and the terrestrial space environment shed great light on their importance to the structure and dynamics of the system. However, thorough understanding of the interaction requires fully statistical investigations. These statistical studies require a large number of suitable pressure front events. In the past these have been obtained through laborious manual inspection of solar wind data. We present an automated procedure that identifies sudden step enhancements in P_{SW} that satisfy a desired set of limiting step-like characteristics, such as magnitude of the change, sign of the change (positive/negative), and rate of change. The procedure yields a list of suitable events, along with their step features within minutes. The technique is tested for missed events and false positives by direct comparison to manual inspection of the data. Its utility is illustrated by application to the statistical effect of pressure fronts to the transpolar potential as measured by the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) technique.

1. Introduction

The state of the Earth's magnetosphere-ionosphere-thermosphere system depends strongly on the input from the solar wind (SW) and the accompanying Interplanetary Magnetic Field (IMF). Case studies of the response of the terrestrial system to specific SW/IMF drivers can reveal valuable information on the structure and dynamics of the system and its interaction with the solar wind. It is not, however, until large statistical studies of specific system properties are conducted, that a better understanding of the system workings as a function of SW/IMF can be achieved. Such large statistical studies require a large collection of suitable SW/IMF drivers or "events" that, when analyzed together through superposed epoch analyses, can yield a statistical understanding of the behavior of the system. Traditionally, this has been done by manual inspection of SW/IMF data of long periods in search of the suitable drivers, or by automated procedures that isolate (given certain criteria) the specific events. The first method obviously requires long periods of manual labor, and suffers from the bias introduced by the subjective criteria of the performing researcher. An automated procedure that takes the human error out of the equation is more suitable and objective. Several such procedures exist, each adapted to the needs of the specific project.

One of the most interesting and effective SW/IMF drivers is a step change in one or more SW/IMF variables. Such change has the ability to modify the system in very short timescales, and introduce short-lived transients or long-term disturbances alike. We introduce a procedure that involves comparison of the SW/IMF variables, for which a step-like change is sought with a predefined step-like function. The agreement of the SW/IMF data with the step-like function for a period around each point in time will be assessed through the cross-correlation technique (Bevington & Robinson, 1992). Our approach is similar to the technique used by Palmroth et al. (2007) in their study of the coupling efficiency of solar wind impulses. The result of the

procedure will be a set of events that satisfy the desired step-like characteristics, such as magnitude of the change, sign of the change (positive/negative), and rate of change.

The existing automated procedures in the literature that detect SW/IMF step discontinuities are motivated mainly by the desire to detect solar wind interplanetary shocks (Cash et al., 2014; Kruparova et al., 2013; Neugebauer et al., 2003; Vorotnikov et al., 2008). A step change in at least three out of four SW/IMF variables (density, velocity, temperature, and magnetic field magnitude) with magnitude above certain limiting values is sought in order for the discontinuity to be considered a candidate for an interplanetary shock (with additional criteria finalizing the classification). It should be noted here that our intention is not to identify the best limiting values for step detection, but to devise a mechanism that will accurately and consistently capture all step changes in the desired variables, given a set of limiting values. In that respect, the above studies used local variation of the respective parameters, either between two consecutive points or within a short period (<10 min averages) before and after each point in time to determine if there is a step jump in a SW/IMF variable. Our approach, as described in Section 2, utilizes a fit of the data with a predefined step-like function over a longer period before and after each point in the time series, thus more accurately identifying steps, steering away from short-term SW/IMF variations.

2. Step-Like Solar Wind Dynamic Pressure

In this section, we describe the methodology used for the automated detection of step changes in the SW/IMF. The idea behind this procedure is to seek a step change in a SW/IMF variable by comparing the entire vicinity of the step with a predefined step function, instead of looking at local properties of the variable (within a few min from the step). We demonstrate the procedure using only one SW parameter, the solar wind dynamic pressure, P_{SW}, thus seeking sudden enhancements in P_{SW}.

The SW/IMF data used in the following are from the Wind and/or the Advanced Composition Explorer (ACE) spacecraft. The data are first propagated to (17, 0, 0) R_E GSM using the Weimer et al. (2003) minimum variance technique, and are given with 1-min resolution (Weygand & McPherron, 2006a, 2006b, 2006c, 2006d). The SW/IMF data do not need to be propagated for the application of the step detection technique, but the propagation is necessary for application to space physics studies, like in the demonstrated application of Section 3, and therefore we use it throughout for consistency.

2.1. Step Analytical Form

Palmroth et al. (2007) used a similar procedure for the detection of pressure changes for their study of the coupling efficiency of solar wind impulses. They compared their pressure data with the Heaviside step function of Equation 1 by means of a cross-correlation coefficient, $0 \le R \le 1$ (Bevington & Robinson, 1992). By accepting only coefficients with values higher than 0.95, they chose only pressure variations that resemble a step change.

$$H(x) = \begin{cases} 0 & : \quad \boldsymbol{x} < 0 \\ 1 & : \quad \boldsymbol{x} \ge 0 \end{cases} \tag{1}$$

We follow a similar procedure, however with two modifications. First, the Heaviside step function exhibits a discontinuity in both its value and its derivative at $x = 0$. At this point, its derivative becomes the Dirac delta function.

$$\delta(x) = \begin{cases} +\infty & : \quad x = 0 \\ 0 & : \quad x \ne 0 \end{cases} \tag{2}$$

Even though this does not pose any problems with the cross-correlation of $H(x)$ with the data, it also does not provide information on the rate of change of the SW/IMF variable. A more useful function is the so-called "logistic" function

$$f(x) = \frac{1}{1 + e^{-\boldsymbol{ax}}} \tag{3}$$

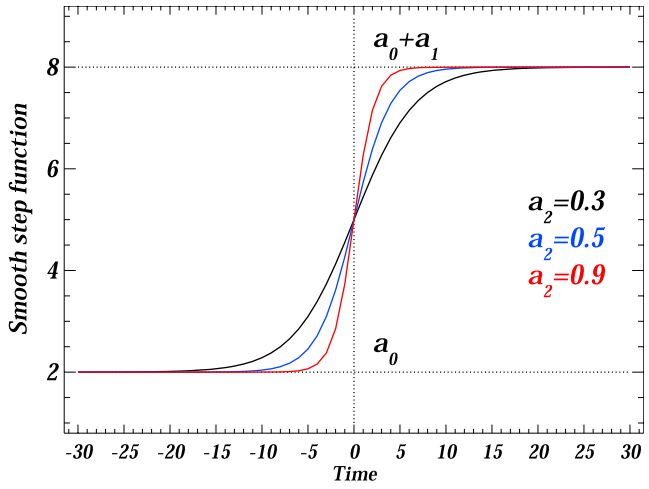

Figure 1. Plot of the function of Equation 4, representing a "smooth" step function.

which is a smooth approximation of $H(x)$, resembling $H(x)$ when $a \to \infty$. For our purposes where the initial and final values can be other than 0 and 1, a more appropriate form of the logistic function is

$$f(t) = a_0 + \frac{a_1}{1 + e^{-a_2 t}} \tag{4}$$

where the coefficients a_i represent the properties of the SW/IMF step, and t is the time with $t = 0$ at the center of the step change. With the above convention, $\lim_{t \to -\infty} f = a_0$, which is the prestep value. On the positive t side, $\lim_{t \to \infty} f = a_0 + a_1$, therefore $a_1 = \Delta f$, the change in f after the step. Finally, a_2 represents the rate of change of f, with higher values leading to steeper changes. Figure 1 shows a plot of Equation 4 for $a_0 = 2$, $a_1 = 6$, and three values of a_2, 0.3, 0.5, and 0.9.

2.2. Step Detection

The second improvement over the Palmroth et al. (2007) study is that instead of directly performing the cross correlation with the data, using estimates and/or multiple variations of the parameters a_i in order to obtain the best correlation, we perform a nonlinear least squares fitting of the data at each point of the time series with the step function f for a specified interval around that point. We then perform a cross correlation of the fitted function with the data. If there is no step present the cross-correlation coefficient is low. If there is a step and the data point under examination is away from the step, the cross-correlation coefficient is again low. The cross-correlation coefficient maximizes right at the step change. We choose the points with $R > 0.95$. Additional criteria about the step characteristics are then used to isolate the appropriate steps.

Let us demonstrate the steps involved in this process by seeking abrupt P_{SW} enhancements. First, we choose the intervals before and after the step for which semi-stable conditions should exist in order to identify isolated step changes. The period before the pressure step is necessary to avoid any prior activity present in the terrestrial system. We choose this to be $t_{bef} = 30$ min. The interval after signifies the minimum duration of the pressure change before it returns to lower values or is further enhanced. We choose this to be $t_{aft} = 15$ min. For a typical solar wind velocity of 400 km/s this means the high pressure region spans ~56 R_E. This is a high-pressure regime that engulfs the entire magnetosphere, thus causing global effects (Boudouridis et al., 2003). Longer times can be chosen to obtain long-lasting pressure changes. These, however, will miss all the shorter duration pulses. Alternatively, t_{aft} can be varied after the step fitting to determine when R returns to low values, thus determine the duration of the step.

The next step involves the nonlinear least square fitting of the chosen data interval at each time t_{SW}, $\left[t_{SW} - t_{bef}, t_{SW} + t_{aft}\right]$, with the function of Equation 4 in the interval $\left[-t_{bef}, + t_{aft}\right]$. Initial values of the coefficients a_i need to be chosen for this fitting. Initial a_0 and a_1 are chosen by computing the average of the data in the intervals before and after. The initial value for a_2 can be chosen to be 0.5, a middle of the range value. More precisely, this value can be estimated by the linear least square fitting of a few points around t_{SW} which determines the local slope of the data at t_{SW}. Taking this to be equal to the slope of the logistic function at $t = 0$ yields the initial value of a_2 through the equation,

$$\left.\frac{df}{dt}\right|_{t=0} = \frac{a_1 a_2}{4} \tag{5}$$

However, we found the fitting procedure to be robust enough, converging fast even with the constant $a_2 = 0.5$ value. The quality of the fit is then evaluated by computing the cross-correlation coefficient R of the fitted function with the data.

The procedure was applied to one day of Wind data, May 1, 1998. The pressure data were smoothed with a three point running average prior to the application of the procedure, but the original unsmoothed data are displayed. Two pressure steps were detected on this day. The fitted curves and parameters are shown in

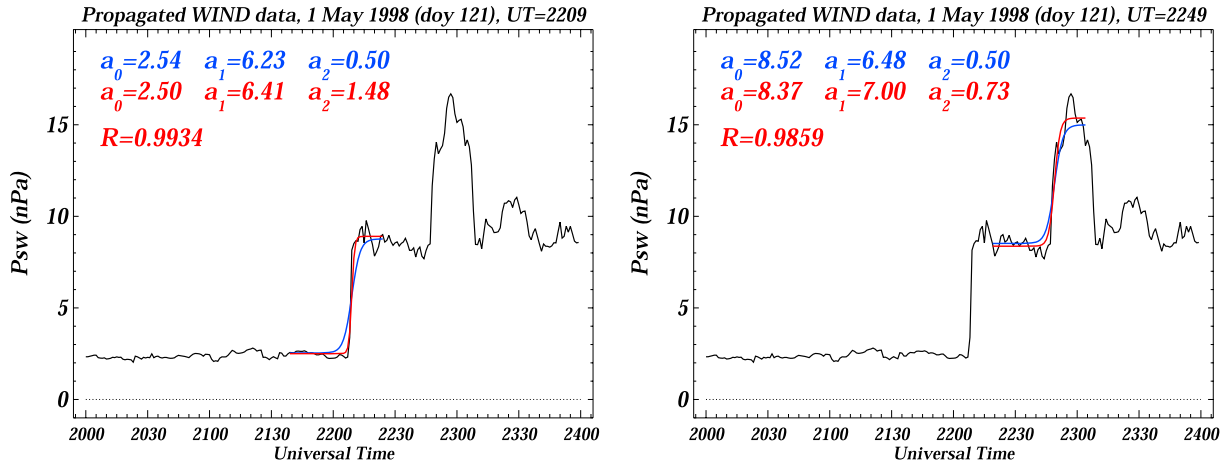

Figure 2. Two examples of the fitting procedure with the fit parameters and curves, initial (blue) and final (red), shown for two P_{SW} steps, 2209 UT (left) and 2249 UT (right), on May 1, 1998, using Wind data.

Figure 2, initial (blue) and final (red). The cross-correlation coefficient R is also shown. Occasionally more than one points around a step have $R > 0.95$ in which case the highest R fit is selected as the final step fit. This is the one shown in Figure 2.

2.3. Step Selection Criteria

As it is the case for this and many other days, more points in the time series have $R > 0.95$. Those points do not qualify as steps in pressure based on two additional criteria we employed in this demonstration of the procedure. The first criterion has to do with the magnitude of the pressure change. We set a minimum positive pressure jump of $\Delta P_{SW} \geq 2$ nPa, as we only wish to consider upward steps of significant magnitude. Therefore, a_1 resulting from the fitting routine has to be ≥2 nPa. The second criterion limits the rate of pressure change. In order for it to be considered an abrupt step, we require that the jump is completed in less than 10 min. However, the analytical logistic function changes from a_0 to $a_0 + a_1$ in infinite time (see plot of Figure 1), as the negative and positive parts of the curve move asymptotically toward the two values, before and after, respectively. So for practical purposes, we consider the step complete when the pressure change reaches 90% of a_1, or $\Delta P = pa_1$, where $p = 0.9$. If we consider this ΔP to be centered around $f(0) = a_0 + a_1 / 2$ then Equation 4 yields that the change occurs in

$$\Delta t = \frac{2}{a_2} \ln\left(\frac{1+p}{1-p}\right) \tag{6}$$

which becomes 0 for $p = 0$ and ∞ for $p = 1$. If we require that $\Delta t \leq \Delta t_{max} = 10$ min, the limiting condition for the time constant a_2 is

$$a_2 \geq \frac{2}{\Delta t_{max}} \ln\left(\frac{1+p}{1-p}\right) \tag{7}$$

For $p = 0.9$, Equation 7 gives $a_2 \geq 0.589$. Any fit resulting in $a_2 < 0.589$ is therefore not considered a pressure step, but more of a pressure ramp.

We illustrate the step detection process based on all three selection criteria (correlation, step magnitude, and step rate) for May 1, 1998 in Figure 3. The bottom panel (d) plots the Wind P_{SW} data. The next three panels, from bottom to top, plot (c) the cross correlation coefficient R, (b) the fit $\Delta P_{SW} = a_1$, and (a) the fit time constant $\boldsymbol{a}_2$. The limiting values for the three criteria are denoted in these panels by the blue horizontal lines. In all panels, the plot line is red when all three criteria are satisfied, and green when $R \geq 0.95$ but only one of the other two criteria is satisfied. In the top two panels, the green points are also marked with green diamonds to highlight their values. The red arrows and times in the bottom panel pinpoint the detected steps. It can be seen from the cross-correlation plot that R is greater than 0.95 in several occasions. However, in all

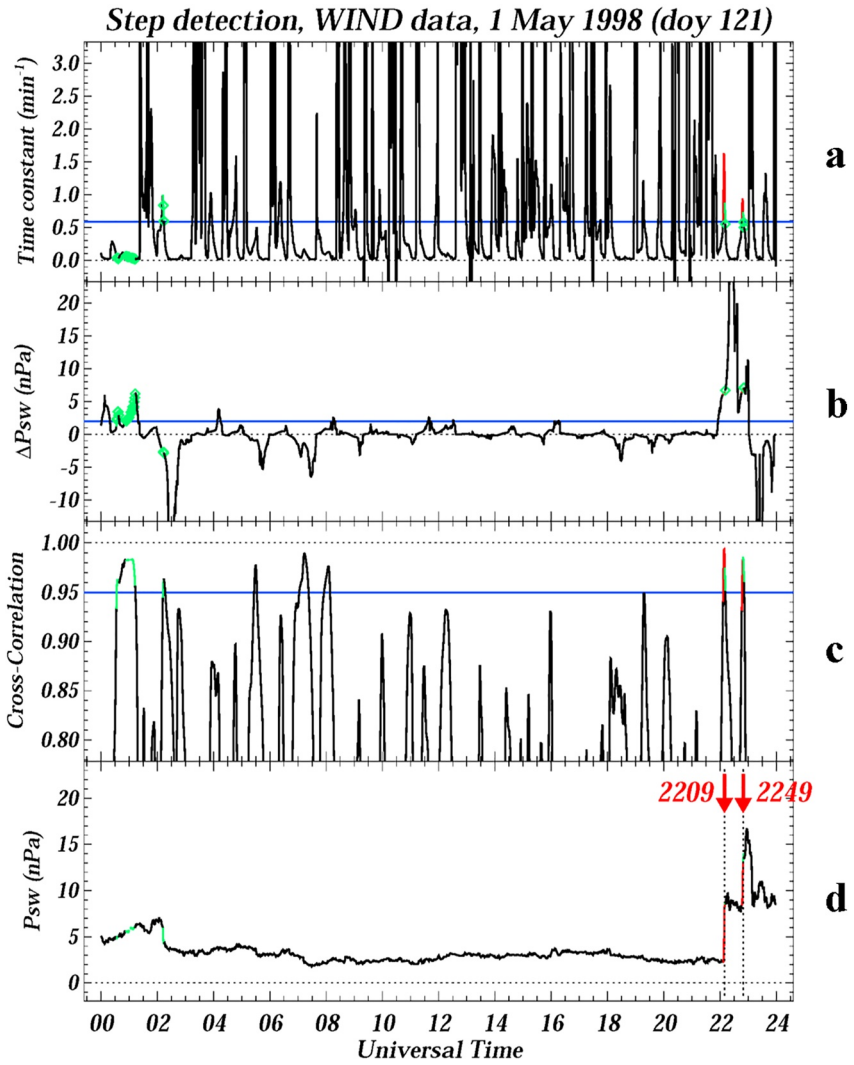

Figure 3. (d) Wind P_{SW} data for May 1, 1998. The next three panels (from bottom to top) plot the three selection criteria: (c) cross-correlation coefficient R, (b) the fit $\Delta P_{SW} = a_1$, and (a) the fit time constant a_2. Blue horizontal lines denote their limiting values. In all panels, the plot line is red when all three criteria are satisfied, and green when $R \geq 0.95$ but only one of the other two criteria is satisfied.

but the two confirmed steps all step candidates fail in one or both of the other two criteria. At around 01 UT, several points have $R \geq 0.95$ and $\Delta P_{SW} \geq 2$ nPa, but the pressure increase rate is very low, $a_2 < 0.1$ min^{-1}, and hence this is a clear pressure ramp. Just after 02 UT there is a sudden drop in pressure with $R \geq 0.95$ and high a_2, but obviously this is a negative a_1 case. In three more instances, 05–09 UT, the correlation is high but both the other two criteria fail to reach the limiting values.

The end result of the procedure is a list of dates and times of the selected step changes in pressure. The pressure characteristics such as ΔP_{SW} and $r = dP_{SW}/dt$ are also recorded, with the latter calculated from Equation 5 since the logistic function slope maximizes at $t = 0$. The final results for the two steps of Figure 2 are shown in Figure 4. The top panels show the specific step in each case with the derived step characteristics in blue. The bottom panels show the entire day fitting results, highlighting the step above in a blue box.

2.4. Procedure Validation

The accuracy, and ultimately the success, of the procedure depend on the numerical values of the input parameters used. For the described technique, these are the two time intervals before and after the examined

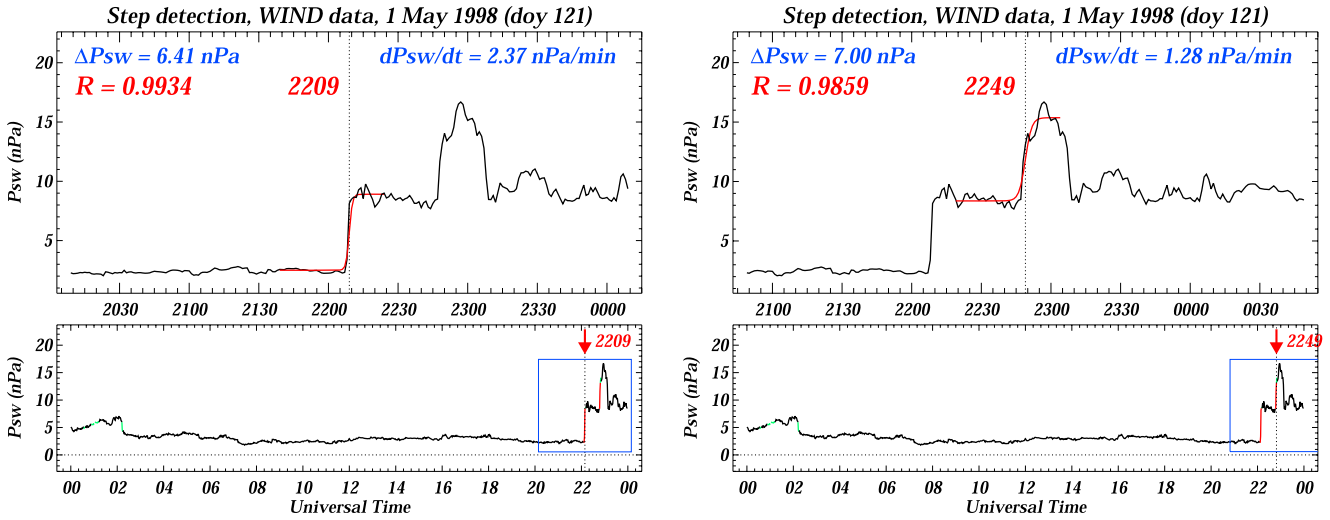

Figure 4. Final-fit results for the two steps of Figure 2, 2209 UT (left) and 2249 UT (right). The top panels show the specific step in each case with the step characteristics in blue. The bottom panels show the entire day fitting results, highlighting the step above in a blue box.

point in the P_{SW} time series, t_{bef} and t_{aft}. A choice other than 30 and 15 min for t_{bef} and t_{aft}, respectively, will affect the results of the procedure in cases when smoothly varying data do not exist before and/or after the potential pressure step. We required that the two periods before and after the pressure step are semistable in order to identify isolated step pressure enhancements, and therefore the procedure works best for those cases. Isolated pressure steps will provide the best results when used in statistical studies of the response of the magnetosphere-ionosphere system, as these events are devoid of any prior activity present in the system that can adversely affect the results of the statistical study. However, it can be argued that a robust procedure should be able to identify the presence of a pressure step, even in those highly variable cases. Below, we investigate the performance and limitations of the technique in that respect.

Another case, when the technique might not yield the desirable results is when the variability in the P_{SW} time series occurs midfront, as opposed to prefront or postfront. This means that the step is not a single, well-defined step, but a series of two or more steps. The intermediate steps could be smaller and less steep than the criteria set for the step selection. Even if they satisfy the given criteria, the time before/after the step is not enough for a good fit of the transition with the function of Equation 4 and $t_{bef} = 30$ min, $t_{aft} = 15$ min, as in these cases the correlation coefficient R will fail to achieve the limiting 0.95 value. Finally, when all the steps are completed and a fit with high correlation is established for the overall step, the rate of change might be too small to qualify under the criterion of Equation 7. Smoothing the data, as we did here, can improve the correlations and detect the overall pressure enhancement as a one-step jump, but there is still a chance of a failed detection. We also examine these multiple step cases further down.

Before we proceed presenting the technique results on a validation interval, and specific examples of the limitations of the procedure during the above mentioned special cases, we need to emphasize one more point. The technique is set to work even with the presence of missing or invalid points in the fit interval. Currently, we have set a limit of half of the combined $t_{bef} + t_{aft}$ interval points necessary for the application of the technique. The fit is simply applied only to the points present. Any fit interval with more than half of its points missing is not considered. In addition, at its present setting, the technique is not applied when the point in examination for a step is itself missing. Therefore, when the missing points are right on the step transition the step is not recognized as such. This could be modified to allow for the technique application using only the points present, before, and after the point in question. However, if all the step transition points are missing, even if a good fit is realized, the step timing and rate of change a_2 will be arbitrary (only a_0 and a_1 could be definitively determined), and thus such avenue was not explored further.

The procedure was tested with Wind data for the full year of 1998. In every case of a candidate step (selected or not) the criteria were visually and graphically examined (with plots like those of Figures 2 and 3). Overall, 88 pressure steps were detected. All of them (100%) were accurately selected, there were no false positives,

Table 1
List of Missed Wind Step Detections During the Validation Interval of 1998

Date	Time	Reason for missed detection	Notes
22-01-1998	2116 UT	Prefront variability	Prefront variability lowers R to ~0.88. Removal of the variability results in step detection with $R > 0.99$.
04-05-1998	0536 UT	Prefront spike	Prefront spike inhibits proper identification, $R \sim 0.93$. Removal of the spike results in step identification with $R > 0.99$.
04-05-1998	0722 UT	Prefront and postfront spikes	Prefront and postfront spikes interfere with identification, $R \sim 0.83$. Removing both spikes results in step detection at 0721 UT, with $R > 0.97$.
10-06-1998	1346 UT and 1359 UT	Double step	Double step with two short duration components. Double step inhibits overall detection, with $R < 0.94$. Removal of the first step results in the detection of the second step at 1359 UT with $R \sim 0.97$. Removal of the second step results in detection of the 1st step at 1346 UT with $R > 0.99$.
31-07-1998	1108 UT	Postfront variability	The presence of a postfront dip leads to missing detection with $R < 0.89$. Removal of the dip results in step detection at 1108 UT, with $R > 0.96$.
31-07-1998	1607 UT	Prefront variability	Prefront dip intereferes with the step detection by lowering R below 0.95. Removal of the dip results in a step detection with $R \sim 0.98$.
01-08-1998	1449 UT	Prior variability	The presence of prior variability inhibits step detection as it lowers R to <0.87. The prefront fitting interval of 30 min is too long. Removal of part of the prior variability so that 30 min fit, results in step detection at 1449 UT with $R > 0.99$.
01-09-1998	0527 UT	Prefront variability	Pre-front variability leads to lower R (0.8) and a smaller step, missing the step detection. Removing the variability results in a step with $R \sim 0.99$ and acceptable jump.
01-09-1998	0620 UT	Double step	The presence of a small first step leads to lower rate of change, $a_2 = 0.33$, even though the overall step has $R \sim 0.97$. Removing the first step results in detection of the overall step with $a_2 = 1.05$.
30-09-1998	0236 UT	Postfront spike	Lower rate due to postfront spike, $a_2 = 0.44$, even though $R > 0.98$. Removal of the spike results in similar R and higher rate, $a_2 = 0.95$.

as all conformed to the selection criteria of Section 2.3. Therefore, the technique is 100% accurate in not selecting spurious events. In addition to the detected steps, 48 not selected candidate steps (green lines in plots like that of Figure 3) were closely examined to uncover the reason for the failed selection. It was determined that for 38 of those the technique had an accurate assessment. The remaining 10 events could have been selected, if it was not for some combination of prefront, midfront, or postfront variability. They are listed in Table 1, which includes the reason for the missed selection, and some notes on their further detailed investigation. Of the 10 cases, 8 were missed due to some prefront or postfront variability, and two of them due to the presence of a double step (midfront variability). Also, 8 were missed due to lower resulting correlation coefficient $R < 0.95$, while 2 were not selected due to lower resulting rate of change, $a_2 < 0.589$. Another reason for a failed detection could be a lower step magnitude, $\Delta P_{SW} < 2$ nPa, as prefront and/or postfront variability can lower the fitted a_1 parameter. We now examine in detail two of the events of Table 1.

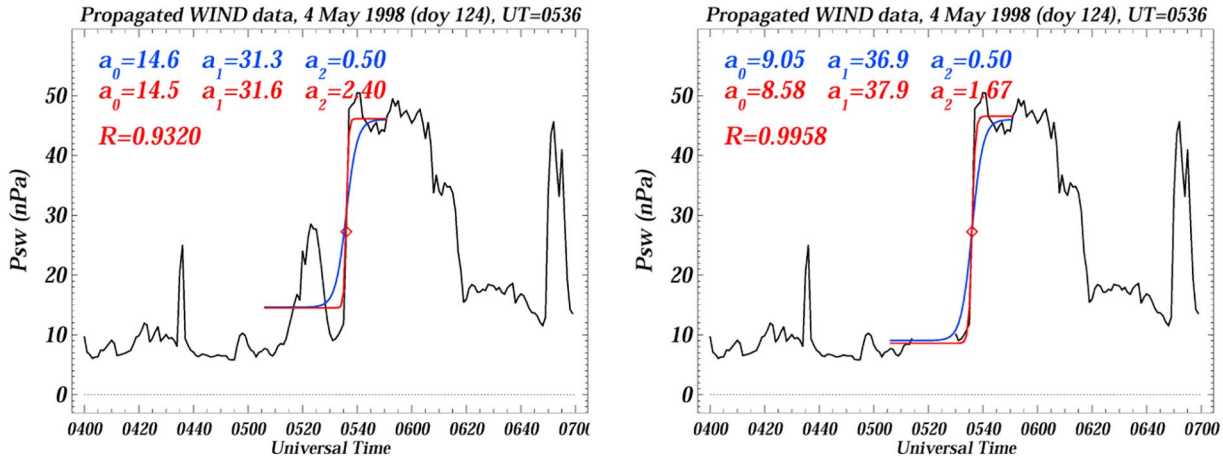

Figure 5. An example of a missed step detection at 0536 UT on May 4, 1998, due to prefront variability. (left) The detection parameters with the variability present. (right) The same with the variability removed.

The first event we examine is an example of a prefront spike/variability that interferes with the step detection at 0536 UT on May 4, 1998 (line 2 of Table 1). Figure 5 shows how this variability affected the fit parameters. The left panel shows the step fit and the derived fit parameters (red) with the presence of a large variation in pressure between 0515 and 0530 UT, approximately, prior to the pressure enhancement at 0536 UT. The red diamond on these plots denotes the point around which the fit was performed. The results show that parameters a_1 and a_2 are in the appropriate range for selection. However, the cross-correlation coefficient R reaches just above 0.93, its highest value in the interval around the pressure step, and therefore the step is not selected. In the right panel of Figure 5 the prefront variability has been artificially removed, setting its values to missing data. Recall that the algorithm can handle missing data, up to half the $t_{\text{bef}} + t_{\text{aft}}$ interval. Without the prefront pressure variability, the technique now yields higher a_1 value, and slightly lower a_2 value (still within range of our criteria) at 0536 UT. However, the coefficient R is now higher than 0.99, easily yielding a step detection.

The second event is an example of a double step on June 10, 1998 (line 4 of Table 1), with the two individual steps being close together at 1346 and 1359 UT. Figure 6 illustrates four fits of the data. The top two panels show the fits of the two individual steps separately, (a) the 1346 UT step on the left and (b) the 1359 UT step on the right. In both cases the step detection algorithm fails to achieve the required $R \geq 0.95$ coefficient. This is due to the fact that the following/preceding step is too close and interferes with the individual fits. The double step fails to qualify in its entirety as well, as the logistic function fit of the overall step never reaches above $R = 0.94$, and in that case the rate of change is only $a_2 = 0.2$. The latter is a common issue with double or multiple steps as mentioned before. The bottom two panels (c and d) show the same fits but now with one step artificially removed as we did in the previous event. In both cases, the correlation coefficient easily achieves higher than 0.95 value, with acceptable fit parameters.

Similar fits and results apply for the remaining cases of Table 1. If we assume the 10 events of Table 1 to be legitimate steps that were missed and considering that the double step of June 10, 1998 contributes two steps, while the double step of September 1, 1998 (line 9 of Table 1) contributes one step, there were supposed to be a total of 99 steps in the Wind data of 1998, of which 88 were accurately identified. This is a success rate of 89%. However, we should put a strong cautionary note here. The procedure presented here is designed to provide a list of events suitable for superposed epoch analyses of the response of the magnetosphere-ionosphere system to solar wind step drivers, in our case the solar wind dynamic pressure. The time intervals $t_{\text{bef}} = 30$ min and $t_{\text{aft}} = 15$ min, before and after the step, respectively, during which semistable pressure conditions were required, were specifically chosen with the intent of selecting events that are "clean," devoid of any prefront or postfront variability. These events have been proven to be the most appropriate drivers for large statistical studies of the solar wind/magnetosphere-ionosphere interaction (e.g., Boudouridis et al., 2011, and references therein), as they steer clear of any possible contamination of the system by adjacent in time activity. In that respect, the events of Table 1 do not qualify as isolated, "clean"

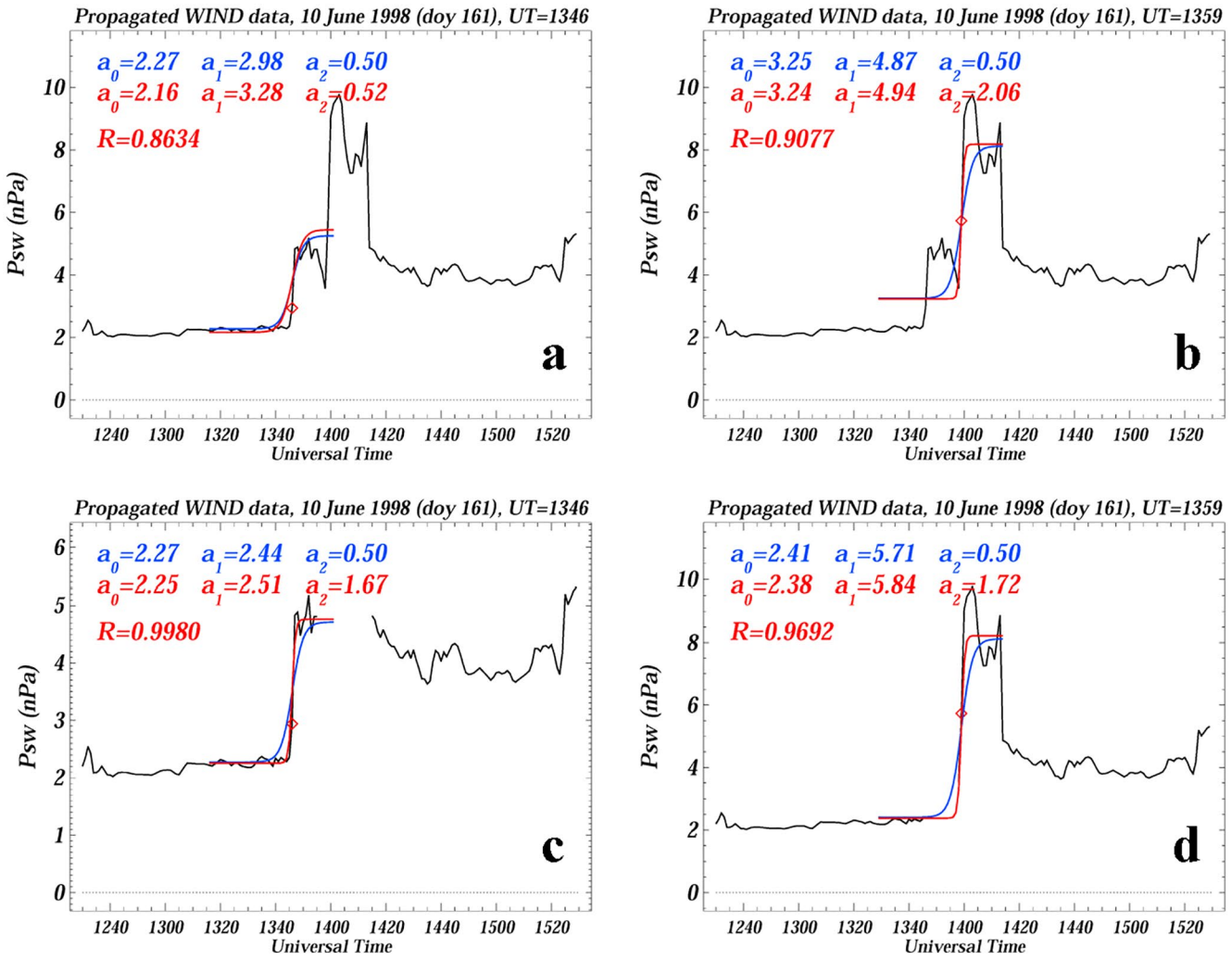

Figure 6. An example of a missed step detection on June 10, 1998, due to the presence of two steps close together at 1346 and 1359 UT. The top two panels (a) and (b) show the step fit for the two steps when both are present. The bottom two panels (c) and (d) show the fits when one of the steps is artificially removed. Left panels (a) and (c) are for the 1346 UT step, right panels (b) and (d) for the 1359 UT step.

steps in pressure, leaving the original list of the 88 events as the most appropriate group of events, given the criteria set forth.

We should note here that, for a given set of selection criteria, namely time intervals before and after, magnitude of the step change and rate of change, the selection can be more or less strict by simply varying the limiting value of the acceptable correlation coefficient, R. This value determines how strictly the event selection adheres to the given criteria. A higher limiting coefficient, will lead to fewer but "cleaner" events. A lower limiting R will yield a larger but less "clean" event sample. In fact, some of the events of Table 1 would have qualified as pressure steps for $R \geq 0.90$. But in that case, the step timing and characteristics would have been less accurate. It is all a balance between the desired number of events and the required quality of events.

2.5. Step Determination Errors

The nonlinear least square fitting procedure also returns the errors of the a_i parameters in the form of their standard deviation, σ_{a_i}. We should emphasize here that, in the rest of this section, these "errors" refer to the natural fluctuation of the fit parameters stemming from the automated nonlinear least squares fitting procedure, denoting the uncertainty associated with the results, and not pointing to any failure of the

technique. The error of the pressure front time can be estimated by considering the number of minutes before and after the selected step UT for which the cross correlation coefficient is still higher than the limiting value, $R \geq 0.95$. For example, if for a selected step there are 2 min before and 1 min after the selected time that still qualify with $R \geq 0.95$, then the step timing error is +1/–2 min. If the selected time is the only point with the acceptable R then the error is ±0 min. This means that for higher limiting R less points qualify, thus the error is smaller, while for lower limiting R more points have a passing correlation around the selected step UT, and therefore the step timing error is larger.

The errors were estimated for the same period as in the validation section, year 1998 of Wind data. Figure 7 shows the error histograms for the parameters involved. From top left we plot the percent fractional error histograms of (a) a_0, (b) $a_1 = \Delta P_{SW}$, (c) a_2, (d) $r = dP_{SW}/dt$, and (e) the histogram of the step UT error, calculated as outlined above. The fractional error is estimated as $\varepsilon = 100 \times \sigma_A / A$, where A is one of a_0, a_1, a_2, and r. The standard deviation of r is calculated through Equation 5 as $\sigma_r = \left(a_1\sigma_{a_2} + a_2\sigma_{a_1} + \sigma_{a_1}\sigma_{a_2}\right)/4$. The error is mostly less than 10% for a_0 (Figure 7a), and less than 15% for a_1 (Figure 7b). The error for a_2 (Figure 7c) is usually below 100%, with a few cases exceeding 100%. This, in combination with the error in a_1, yields a higher error in r (Figure 7d) as calculated above, often higher than 100%. Finally, the step UT error (Figure 7e), calculated as outlined above, is ±2 min, with a preference for +1 min, signifying an impressive accuracy in the determination of the step timing.

The high errors in a_2 and r introduce an additional error in the accuracy of the step determination. At first look, the high errors in a_2 could cast doubt in the step selection process, as any value of a_2 below its limiting value, 0.589 in our analysis, will prevent the event selection as an acceptable step. This can be especially concerning for cases with $\sigma_{a_2} \geq a_2$, or equivalently $\varepsilon_{a_2} \geq 100\%$ (see Figure 7c). This means that any value of a_2 below its selected value, including those less than 0.589 are possible, signifying an uncertainty in the step selection process. However, a closer look at the possible range of a_2 values, and the performance of their step fit based on their cross-correlation coefficient R, puts this worry to rest. Figure 8 shows how the procedure performs for a varying a_2 value within its error range for a pressure front at 2044 UT on November 29, 1998. The left panel shows the step fit curves for $a_{2L} = a_2 - \sigma_{a_2}$ (orange), the lower end of the a_2 error range, and $a_{2H} = a_2 + \sigma_{a_2}$ (green), the higher limit of the a_2 error range, in addition to the selected a_2 step fit curve in red (blue again being the initial guess of the fit parameters). The standard deviation for a_2 in this case is $\sigma_{a_2} = 0.41$, which is about 67% of a_2. The a_0 and a_1 parameters for the orange/green curves are the same as for the red curve. The correlation coefficients R for the selected a_2 value and its extremes are also indicated on the plot. We can see that the technique performs best for the selected a_2 value, both from the three values of R and the fit curves themselves. This becomes obvious by looking at the right panel of Figure 8 (the a_2 colors correspond to those of the left panel). We calculated the correlation coefficient R between the data and step fit curves for successive values of a_2, starting at the lower end of its error range all the way to the high end, 10 values on each side of the selected a_2 value. The plot shows that R maximizes at the a_2 value originally selected by the automated technique. Similar plots were created for all 88 selected steps, and all with no exception exhibit the same behavior of R versus a_2. These results lend strong confidence to the performance of the step detection technique, despite the high errors in a_2 and r.

3. Sample Use: Pressure Front Effects on the Transpolar Potential

In this section, we demonstrate the effectiveness of the procedure on a space physics study. This relates to the effects of solar wind dynamic pressure fronts on the transpolar potential as measured by the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) technique (Richmond & Kamide, 1988). The response of the transpolar potential to solar wind dynamic pressure fronts has been extensively studied in recent decades (Boudouridis et al., 2004, 2005, 2008; Connor et al., 2014). Statistical understanding of this response requires a large superposed epoch analysis of the potentials measured after solar wind pressure fronts. Before we proceed with our analysis, we should mention that other parameters can shed light on the effects of pressure fronts on the magnetosphere-ionosphere system. The combination of high-resolution (1 min) convection-related quantities, such as the polar cap index (e.g., Troshichev et al., 1988, 2007), and the variations in the 1-min ring current SYM-H indices (e.g., Araki & Shinbori, 2016) can be of great use. In addition, possible storm sudden commencement (SSC) events associated with the pressure-related

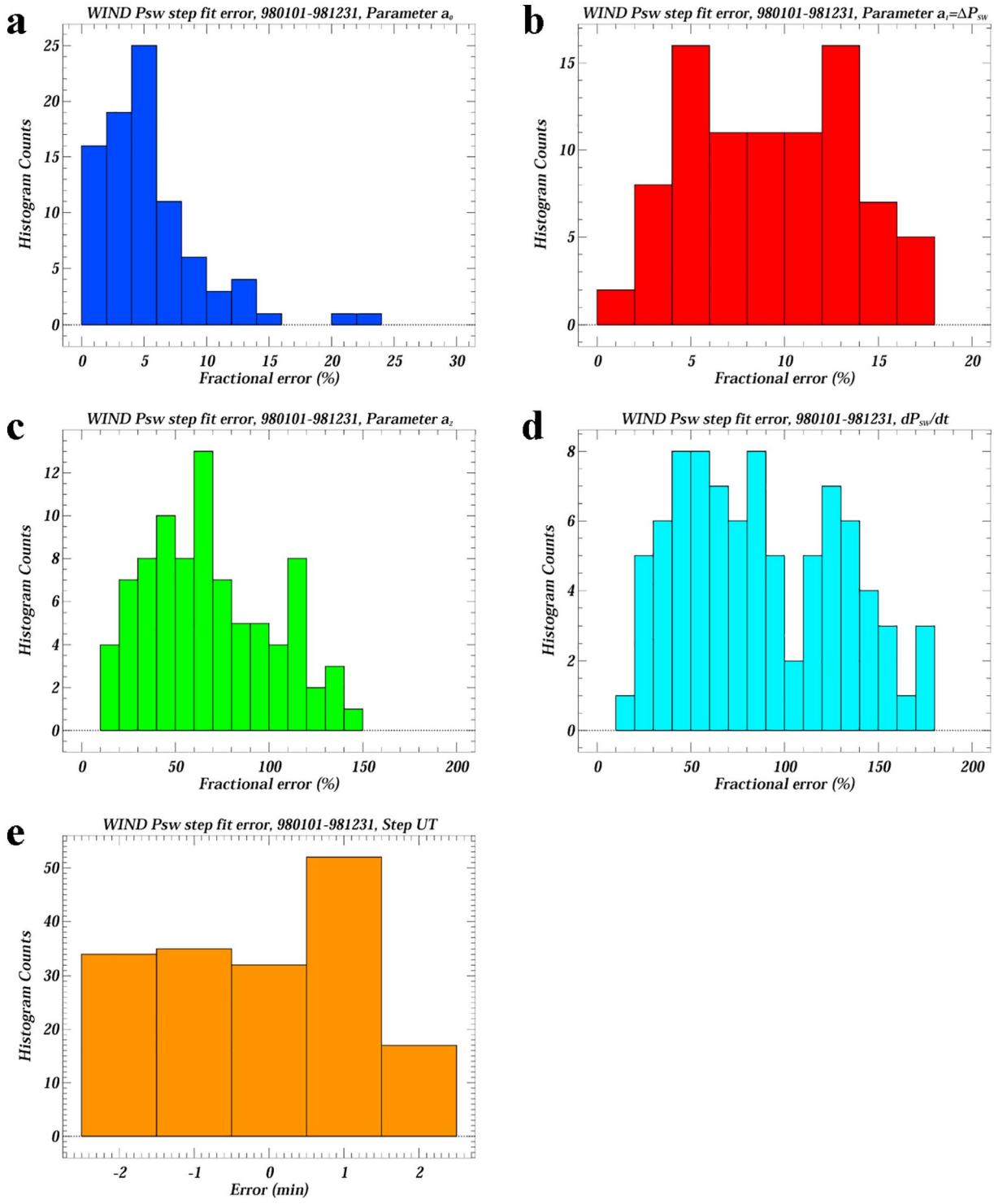

Figure 7. Error histograms of the step fit parameters during the validation period, year 1998 of Wind data. From top left, we plot histograms of the fractional error (%) for parameters (a) a_0, (b) $a_1 = \Delta P_{SW}$, (c) a_2, (d) $r = dP_{SW}/dt$, and (e) the histogram of the step UT error (min).

magnetopause compressional events might provide further or alternative verifications of the magnetospheric reactions to solar wind pressure steps (e.g., Gillies et al., 2012).

The core of the AMIE technique is a weighted least squares fit of observations from various sources, both on the ground (ground magnetometers, Super Dual Auroral Radar Network [SuperDARN] radar

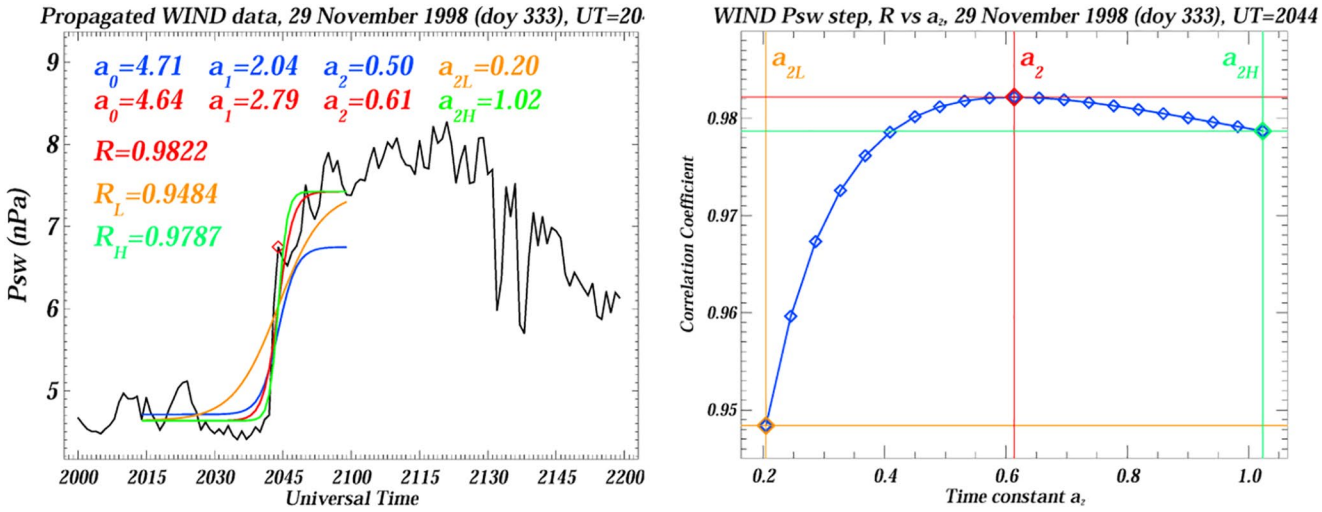

Figure 8. An example of the procedure performance for different time constants a_2 for a P_{SW} step at 2044 UT on November 29, 1998. (left) The step fit curves for the limits of the a_2 range. (right) The correlation coefficient R as a function of a_2 for multiple values within the a_2 range.

measurements), and in space (Defense Meteorological Satellite Program [DMSP] satellites). The result of the assimilative technique is the determination of the high-latitude ionospheric convection pattern, which then yields various ionospheric parameters, including the transpolar potential (Kihn & Ridley, 2005; Kihn et al., 2006; Lu et al., 1996, 1998; Ridley et al., 1998). For the sample statistical study below, we use AMIE results produced with only 1-min resolution magnetometer data from 60 to 130 Northern Hemisphere high latitude magnetometer stations, resulting in a transpolar potential with a continuous, 1-min resolution, global coverage.

For an accurate superposed analysis, the AMIE-derived quantities in the ionosphere need to be associated with the SW/IMF quantities upstream of the magnetopause. Instead of using the propagated SW/IMF values directly, we first apply a time delay of 10 min to account for the travel time from 17 R_E to the magnetopause and the transmission to the ionosphere. Furthermore, we include preconditioning of the magnetosphere by taking an average of these variables for 10 min prior to the delayed time. This averaging smooths any short-term fluctuations that can affect our statistics, and provides a measure of the state of the magnetosphere prior to the pressure front impact. Thus, the SW/IMF value assigned to an AMIE data point at time t_{AMIE} is given by

$$A_{AMIE} = \frac{1}{N} \sum_{t=t_{AMIE}-t_d-t_p}^{t_{AMIE}-t_d} A(t) \quad (8)$$

where $t_d = 10$ min is the applied time delay, $t_p = 10$ min is the preconditioning interval, $A(t)$ is the SW/IMF variable time series at 17 R_E upstream, and $5 \le N \le t_p$ is the number of points used for the averaging (considering that the propagated SW/IMF data have 1-min resolution, and depending on the presence or not of missing points).

Boudouridis et al. (2008) presented a case study of a P_{SW} step increase on April 30, 1998, and its effects on the transpolar potential as derived from the AMIE technique. They argued that immediately after the increase in pressure the transpolar potential first increases but subsequently declines, thus exhibiting a transient response. Panel (a) of Figure 9 shows 1-min color-coded AMIE transpolar potential for an event on February 28, 1998, as a function of time relative to the pressure front impact time and prevailing IMF B_z. By "prevailing," we mean an IMF B_z value that was calculated according to Equation 8. This process is clear when looking at the Wind SW/IMF conditions shown in the panels on the left below the transpolar potential. From top to bottom we plot Wind spacecraft measurements of (b) P_{SW} (not delayed and preconditioned for precision), (c) IMF B_y component, (d) IMF B_z component, and (e) solar wind electric field, E_{SW}. The Weimer propagated values are in black, the assigned (delayed and propagated) values are in blue and red, red indicating the variable that is reflected on the Y axis of the transpolar potential plot at the top left.

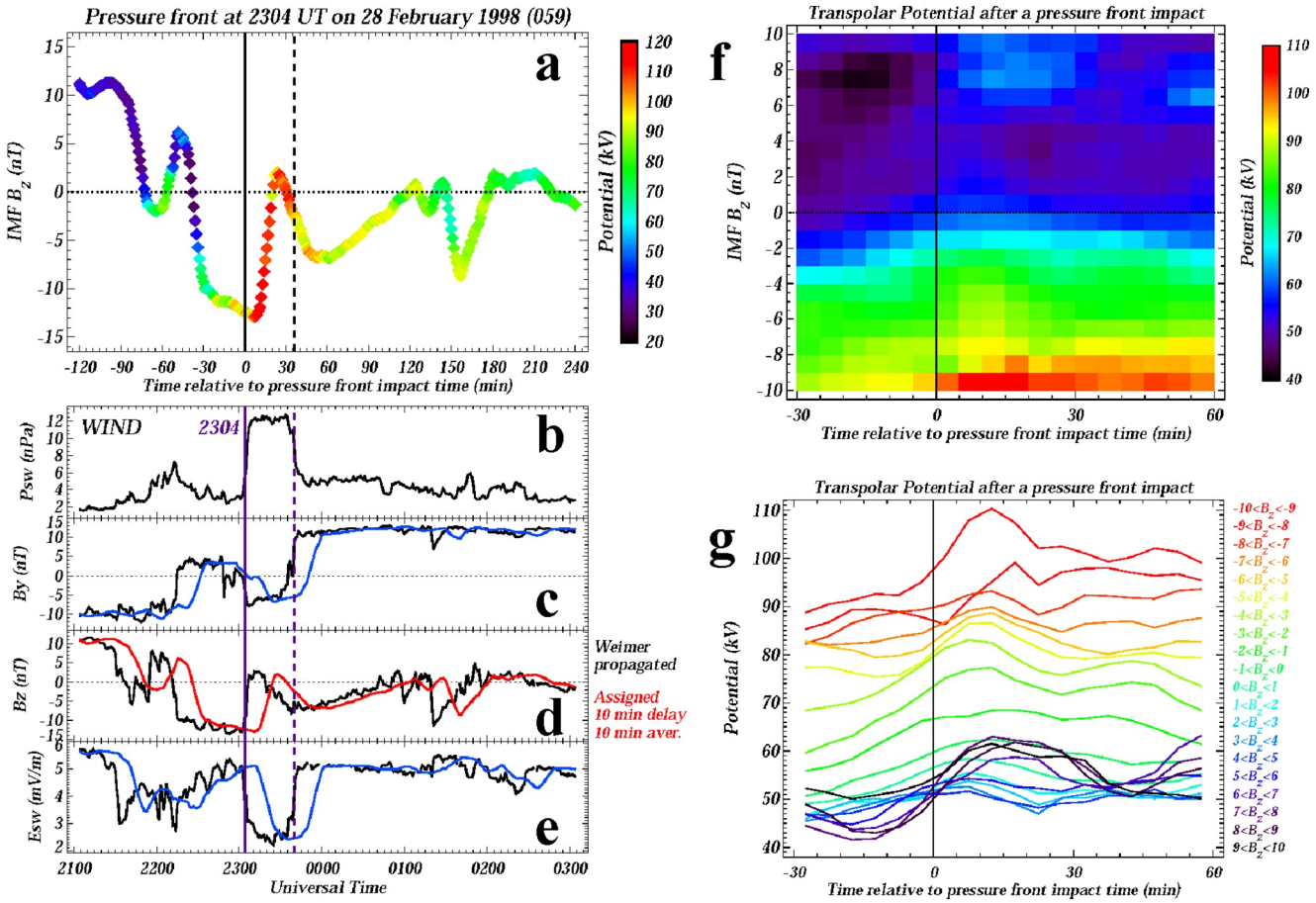

Figure 9. (left top) Assimilative Mapping of Ionospheric Electrodynamics (AMIE) transpolar potential for February 28, 1998 as a function of time relative to the pressure front impact time and prevailing Wind Interplanetary Magnetic Field (IMF) B_z. (left bottom) Solar wind and IMF conditions during the pressure front. (right top) Superposed epoch analysis showing average potential as a function of relative-to-impact time and IMF B_z. The event list contains 263 Advanced Composition Explorer (ACE) pressure fronts detected with the automated step technique for $\Delta P_{SW} \geq 2$ nPa, and a pressure jump completed within 10 min. (right bottom) A line plot every 1 nT of IMF B_z of the plot above.

The reason for the inclusion of the IMF B_z on the Y axis, and the 3-D plot format of the potential is that the IMF B_z, which is the main driver of ionospheric activity, does not always remain constant during a pressure front impact, and thus can influence the result of a superposed epoch analysis by mixing events with different and variable IMF B_z values. The inclusion of the IMF B_z dependence allows us to evaluate the potential behavior separately at different ranges of IMF B_z. The significant transpolar potential increase is clearly seen after the step increase in P_{SW} at 2304 UT, even though the IMF B_z moved northward, thus in theory reducing the ionospheric response.

We can make similar plots for other solar wind pressure fronts, and put them all together in a superposed epoch analysis plot. This requires a consistent pressure front determination with certain characteristics. A list of 263 ACE pressure fronts from February 1998 to December 2000 was generated using the above step detection technique with $R_{min} = 0.95$. We also required that the pressure fronts have $\Delta P_{SW} \geq 2$ nPa, and the pressure jump is completed in less than 10 min. This event list was produced in less than 10 min. Other event lists, with different pressure front characteristics can be produced easily in similar times. Therefore, stronger events can be compared with weaker ones, or faster events compared with slower ones, without the need for laborious manual event searches.

The right panels of Figure 9 show the result of the superposed epoch analysis of the effect of solar wind dynamic pressure fronts on the AMIE-derived transpolar potential, using the list generated with our automated step detection procedure. Panel (f) shows the transpolar potential, color-coded according to the scale

on the right, as a function of time relative to the pressure front impact on the X axis, and IMF B_z on the Y axis. The data from all individual events are averaged within bins of 5 min by 1 nT, and further smoothed with a 9-bin box average. We limited the IMF B_z range between –10 and 10 nT, where enough data points are available. The transpolar potential increases after the pressure enhancement for almost all values of IMF B_z (except for the low northward IMF values). Then it declines again in ~30 min, sometimes remaining elevated compared to the pre-front values. Panel (g) shows the same result in a line format, for cuts at every 1 nT of IMF B_z.

4. Conclusions and Future Directions

Thorough understanding of the interaction of the SW/IMF with the terrestrial environment requires high quality statistical studies that convey information on the most probable response of the magnetosphere-ionosphere-thermosphere system to SW/IMF drivers. These statistical studies rely on the consistent selection of time intervals with specific SW/IMF properties, a list of periods or events for which the system response is sought. We presented an automated technique that uses non-linear least squares fitting of the P_{SW} data to detect sharp step-like changes in solar wind dynamic pressure. The technique uses a given analytical functional form that simulates a smoothly varying step change. In addition to the time of the P_{SW} step change, the procedure yields the magnitude of the change and the rate of change. The effectiveness of the procedure was tested with the application to a magnetospheric problem, the effects of solar wind dynamic pressure fronts on the transpolar potential.

In its current form, the presented technique has a narrow scope, the detection of step-like changes in a single SW quantity, the solar wind dynamic pressure. There are many avenues of improvement, expansion, and generalization of the described procedure. One improvement that has been explored (not presented here) is the variation of the time t_{aft} in order to determine the duration of the high-pressure regime. Two automated algorithms are investigated in this respect and will be presented in future work. However, a lot more can be done to expand and generalize this procedure. Following are a few suggestions for future directions:

1. Application of the technique to additional SW/IMF variables, such as solar wind density, velocity, electric field, IMF B, B_x, B_y, B_z, clock angle, etc.
2. Further assessment of the step fit errors, perhaps with additional criteria introduced to exclude results with higher than normal errors in the a_i parameters.
3. Implementation of simultaneous step changes for more than one variables, so that structures such as shocks and other discontinuities, where more than one variables exhibit a step-like behavior, can be studied.
4. Application of the procedure to OMNI Web data set that is most commonly used by the space sciences community.
5. It is often desired that, when a step occurs in one SW/IMF variable, other variables remain steady so that they will not affect the response of the terrestrial system. This option can be added to the algorithm so that a choice will be given to select certain variables to be stable during the step variation of the primary variables. Various stability requirements can be explored, such as the limiting of the range of a variable, or its standard deviation, during a specific time interval preceding the time under examination. Such automated stability detection procedures have already been developed by the author, and will be presented in a separate publication.
6. The current nonlinear least squares fitting algorithm is applied to the step-like functional form described in Section 2.1. However, there is no limitation as to what functional form the procedure is applied to. Other analytical forms, with the proper parameterization, can be used to study various SW/IMF structures and associated physical phenomena in the solar wind, such as waves, asymptotic structures, etc.

In conclusion, the described technique provides a flexible automated procedure that can be expanded to study a wide range of solar wind structures and phenomena, by furnishing a list of suitable events in a very short time. Its application to statistical studies of the interaction of the solar wind with the magnetosphere-ionosphere-thermosphere system, bypasses the cumbersome, time consuming, manual searches for appropriate events, letting the investigation of this interaction focus on the scientific merits of the physical mechanisms explored, and thus allowing for expedited and more efficient reporting of the results.

Data Availability Statement

The propagated solar wind and IMF data can be found at http://vmo.igpp.ucla.edu/data1/Weygand/. The AMIE results can be found at http://vmr.engin.umich.edu/Model/_amie/plot.php.

Acknowledgments
The work by Athanasios Boudouridis at the Space Science Institute (SSI) was supported by NASA Geospace Sciences award NNX11AJ07G. The authors wish to thank James Weygand of the University of California at Los Angeles for the propagated solar wind and IMF data.

References

Araki, T., & Shinbori, A. (2016). Relationship between solar wind dynamic pressure and amplitude of geomagnetic sudden commencement (SC). *Earth Planets and Space*, *68*, 90. https://doi.org/10.1186/s40623-016-0444-y

Bevington, P. R., & Robinson, D. K. (1992). *Data reduction and error analysis for the physical sciences* (2nd ed.). New York: McGraw-Hill Inc.

Boudouridis, A., Lyons, L. R., Zesta, E., Weygand, J. M., Ribeiro, A. J., & Ruohoniemi, J. M. (2011). Statistical study of the effect of solar wind dynamic pressure fronts on the dayside and nightside ionospheric convection. *Journal of Geophysical Research*, *116*, A10233. https://doi.org/10.1029/2011JA016582

Boudouridis, A., Zesta, E., Lyons, L. R., & Anderson, P. C. (2004). Evaluation of the Hill-Siscoe transpolar potential saturation model during a solar wind dynamic pressure pulse. *Geophysical Research Letters*, *31*, L23802. https://doi.org/10.1029/2004GL021252

Boudouridis, A., Zesta, E., Lyons, L. R., Anderson, P. C., & Lummerzheim, D. (2003). Effect of solar wind pressure pulses on the size and strength of the auroral oval. *Journal of Geophysical Research*, *108*(A4), 8012. https://doi.org/10.1029/2002JA009373

Boudouridis, A., Zesta, E., Lyons, L. R., Anderson, P. C., & Lummerzheim, D. (2005). Enhanced solar wind geoeffectiveness after a sudden increase in dynamic pressure during southward IMF orientation. *Journal of Geophysical Research*, *110*, A05214. https://doi.org/10.1029/2004JA010704

Boudouridis, A., Zesta, E., Lyons, L. R., Anderson, P. C., & Ridley, A. J. (2008). Temporal evolution of the transpolar potential after a sharp enhancement in solar wind dynamic pressure. *Geophysical Research Letters*, *35*, L02101. https://doi.org/10.1029/2007GL031766

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

Connor, H. K., Zesta, E., Ober, D. M., & Raeder, J. (2014). The relation between transpolar potential and reconnection rates during sudden enhancement of solar wind dynamic pressure: Open GGCM-CTIM results. *Journal of Geophysical Research: Space Physics*, *119*, 3411–3429. https://doi.org/10.1002/2013JA019728

Gillies, D. M., St.-Maurice, J.-P., McWilliams, K. A., & Milan, S. (2012). Global-scale observations of ionospheric convection variation in response to sudden increases in the solar wind dynamic pressure. *Journal of Geophysical Research*, *117*, A04209. https://doi.org/10.1029/2011JA017255

Kihn, E. A., Redmon, R., Ridley, A. J., & Hairston, M. R. (2006). A statistical comparison of the AMIE derived and DMSP-SSIES observed high-latitude ionospheric electric field. *Journal of Geophysical Research*, *111*, A08303. https://doi.org/10.1029/2005JA011310

Kihn, E. A., & Ridley, A. J. (2005). A statistical analysis of the assimilative mapping of ionospheric electrodynamics auroral specification. *Journal of Geophysical Research*, *110*, A07305. https://doi.org/10.1029/2003JA010371

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*, *118*, 4793–4803. https://doi.org/10.1002/jgra.50468

Lu, G., Baker, D. N., McPherron, R. L., Farrugia, C. J., Lummerzheim, D., Ruohoniemi, J. M., et al. (1998). Global energy deposition during the January 1997 magnetic cloud event. *Journal of Geophysical Research*, *103*, 11685–11694. https://doi.org/10.1029/98JA00897

Lu, G., Emery, B. A., Rodger, A. S., Lester, M., Taylor, J. R., Evans, D. S., et al. (1996). High-latitude ionospheric electrodynamics as determined by the assimilative mapping of ionospheric electrodynamics procedure for the conjunctive SUNDIAL/ATLAS 1/GEM period of March 28–29, 1992. *Journal of Geophysical Research*, *101*, 26697–26718. https://doi.org/10.1029/96JA00513

Neugebauer, M., Steinberg, J. T., Tokar, R. L., Barraclough, B. F., Dors, E. E., Wiens, R. C., et al. (2003). Genesis on-board determination of the solar wind flow regime. *Space Science Reviews*, *105*, 661–679.

Palmroth, M., Partamies, N., Polvi, J., Pulkkinen, T. I., McComas, D. J., Barnes, R. J., et al. (2007). Solar wind-magnetosphere coupling efficiency for solar wind pressure impulses. *Geophysical Research Letters*, *34*, L11101. https://doi.org/10.1029/2006GL029059

Richmond, A. D., & Kamide, Y. (1988). Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique. *Journal of Geophysical Research*, *93*, 5741–5759. https://doi.org/10.1029/JA093iA06p05741

Ridley, A. J., Lu, G., Clauer, C. R., & Papitashvili, V. O. (1998). A statistical study of the ionospheric convection response to changing interplanetary magnetic field conditions using the assimilative mapping of ionospheric electrodynamics technique. *Journal of Geophysical Research*, *103*, 4023–4039. https://doi.org/10.1029/97JA03328

Troshichev, O. A., Andrezen, V. G., Vennerstrøm, S., & Friis-Christensen, E. (1988). Magnetic activity in the polar cap: A new index. *Planetary and Space Science*, *36*, 1095–1102. https://doi.org/10.1016/0032-0633(88)90063-3

Troshichev, O. A., Janzhura, A. S., & Stauning, P. (2007). Magnetic activity in the polar caps: Relation to sudden changes in the solar wind dynamic pressure. *Journal of Geophysical Research*, *112*, A11202. https://doi.org/10.1029/2007JA012369

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

Weimer, D. R., Ober, D. M., Maynard, N. C., Collier, M. R., McComas, D. J., Ness, N. F., et al. (2003). Predicting interplanetary magnetic field (IMF) propagation delay times using the minimum variance technique. *Journal of Geophysical Research*, *108*(A1), 1026. https://doi.org/10.1029/2002JA009405

Weygand, J. M., & McPherron, R. L. (2006a). *Wind 3DP Weimer propagated 60 s resolution in GSM coordinates*. https://doi.org/10.21978/P82K7X

Weygand, J. M., & McPherron, R. L. (2006b). *Wind Weimer propagated 60 s resolution tri-axial fluxgate magnetometer in GSM coordinates*. University of California, Los Angeles. https://doi.org/10.21978/P8FW5P

Weygand, J. M., & McPherron, R. L. (2006c). *ACE SWEPAM solar wind Weimer propagated 60 s resolution data in GSM coordinates*. https://doi.org/10.21978/P8S62R

Weygand, J. M., & McPherron, R. L. (2006d). *ACE Weimer propagated 60 s resolution tri-axial fluxgate magnetometer in GSM coordinates*. https://doi.org/10.21978/P8933R