Adv. Space Res. Vol. 8, No. 9–10, pp. (9)71–(9)86, 1988
Printed in Great Britain.

0273–1177/88 $0.00 + .50

IMF CONTROL OF GEOMAGNETIC ACTIVITY

R. L. McPherron,*,** D. N. Baker,*** L. F. Bargatze,*,**
C. R. Clauer† and R. E. Holzer*,**

* *Institute of Geophysics and Planetary Physics,*
** *Department of Earth and Space Sciences, University of California, Los Angeles, CA 90024–1567, U.S.A.*
*** *Laboratory for Extraterrestrial Physics, NASA/Goddard Space Flight Center, Greenbelt, MD 20771, U.S.A.*
† *STAR Laboratory, Stanford University, Stanford, CA 94305, U.S.A.*

ABSTRACT

Geomagnetic activity is produced by changes in the solar wind velocity and southward turnings of the interplanetary magnetic field. Two decades of study of this coupling have established a number of empirical relations. The technique of linear prediction filtering has been particularly useful in defining impulse response functions relating solar wind and geomagnetic parameters. For example, the response function relating the rectified solar wind electric field to the AL index is approximately a Rayleigh function with time constant of one hour. The Fourier transform of this function gives the transfer function of the magnetosphere, which is a low pass filter. Filtering analysis shows that less than half the variance of the AL index is predictable by the solar wind implying that the magnetosphere is not a simple, passive element as assumed. When response functions are calculated for different levels of activity we find that they are bi-modal for moderate levels of activity and uni-modal for strong activity. A simple bi-modal model of the magnetospheric response can account for 90% of the variance in AL, provided the parameters are varied from substorm to substorm. This result suggests that both growth phase and expansion phase currents are very closely proportional to the rectified solar wind electric field, but the parameters in this relation depend on internal conditions about which the solar wind has no information. We perform a simple simulation of the magnetosphere incorporating the idea that magnetic activity is the result of superposing two types of currents, those directly driven by the solar wind and those driven by somewhat random unloading of energy stored in the magnetotail by the driven processes. The model response functions have the same behavior as real data, i.e. bi-modal response for moderate activity and uni-modal response for strong activity. The near-earth neutral line model provides a possible explanation for why two types of processes are superimposed in the magnetic variations.

INTRODUCTION

Geomagnetic activity is caused by the interaction of the solar wind with the earth's magnetic field. The form of this interaction is primarily controlled by the interplanetary magnetic field (IMF). When the IMF is northward, disturbances are observed only within the polar cap. If it is southward, magnetic disturbances and aurora occur in the auroral oval. The stronger the southward component of the IMF, the stronger the disturbance. This coupling of the solar wind to the magnetosphere is brought about by the process of magnetic reconnection. As first suggested by Dungey /1/, a solar wind magnetic field oriented antiparallel to the earth's magnetopause near the subsolar point can merge or reconnect with the earth's field. Magnetic field lines initially closed with both feet connected to the earth are opened to the solar wind. The electric field of the solar wind is then projected onto the ionosphere where it drives electrical currents. Gradients in the ionospheric electric field and conductivity create divergences in the currents which must flow along magnetic field lines eventually closing in the solar wind. These currents appear to be directly driven by a solar wind dynamo /2/.

Once field lines become connected to the solar wind they are pulled over the polar caps away from the Sun producing the earth's magnetotail. Eventually they must reconnect or otherwise all the magnetic flux on the day side would be transported to the tail. It seems likely that this reconnection would occur at a point downstream where the open field lines from opposite poles first meet in the center of the tail. Once reconnected they return to the day side completing a closed loop of magnetospheric convection. This simple model is the one envisioned by Dungey, and is the basis of the dynamo model of magnetic activity discussed by Akasofu /2/.

Observations of auroral and magnetic activity on the ground /3/ and in space /4,5/ reveal a more complicated picture. When the IMF turns southward a growth phase of about 60 minute duration is followed by a sudden explosion of auroral and magnetic activity. This explosion begins on the most equatorward auroral arc in the midnight sector, and then rapidly expands east, west, and poleward. This expansion phase persists for 30-40 minutes and then decays away in a 90 minute recovery phase. The interval of time encompassing the three distinct phase is called a magnetospheric substorm /6/. Such data seem to indicate that the return of magnetic flux to the day side must be initiated close to the earth on the night side, rather than in the deep tail.

Detailed studies of substorms in the magnetotail plasma sheet /7,8,9/ have produced a model of substorms referred to as the near-earth neutral line, or plasmoid model. As the name implies, the neutral line at which nightside reconnection occurs is near to the earth in the range 8-30 R_e. Instead of reconnecting at a distant neutral line as in the dynamo model, reconnection begins first on closed field lines within the near-earth plasma sheet. When reconnection reaches the open field lines of the lobes is it possible to return the flux transported to the tail during the growth phase. Then, circular loops of closed field lines formed by reconnection inside the plasma sheet are pulled from the tail by newly reconnected solar wind magnetic field lines. The structure composed of these loops and plasma energized at the neutral line is called a plasmoid /10/.

This model has been called the unloading model /11/ to distinguish it from the driven model. In the unloading model, energy derived from the solar wind is stored in the magnetic field of the tail lobes. When reconnection of open field lines begins this magnetic energy is "unloaded" and converted to thermal and kinetic energy of particles in the plasma sheet. It has been strenuously argued that the relation between measures of magnetic activity and the IMF demonstrates that the magnetosphere behaves more like a driven system than an unloading system /12/. This seems to imply that the energy storage and unloading process is not a significant component of the substorm.

Numerous investigations have established beyond question the role of the IMF in controlling magnetic activity (e.g. the collection of papers edited by Kamide and Slavin, /13/). Recent studies have considered quantitative aspects of this control. For example, it is now known that the solar wind parameter most linearly related to the AL index is the coupling function $V^2BG(\theta)$, where $G(\theta)$ is a gating function accounting for the strong dependence of activity on the orientation of the IMF. In many studies, $BG(\theta) = U(\pi/2-\theta)B\cos(\theta)$, the southward component of the IMF (where U is the unit step function).

Only a few studies have addressed the question of how closely the waveform of an output such as the AL index resembles the input coupling function. In examining the relation between the epsilon parameter [proportional to $VB^2\sin^4(\theta/2)$] and AE, Akasofu /14/ concluded that magnetic activity is primarily driven because the waveforms of hourly averages of epsilon and AE are so similar. We, however, have used higher time resolution data to examine essentially the same relation and find that only about 40% of the variance in the AL index is related to the rectified solar wind electric field /15,16/. We have argued that this low level of predictability is a consequence of the sporadic unloading of stored energy at times that are not closely correlated with changes in the IMF.

The unloading model is not incompatible with the driven hypothesis. In fact, the near-earth neutral line model assumes that the process of energy storage in the tail is driven by the solar wind /17/. The idea that substorm activity is actually a superposition of two types of processes has been enunciated clearly in several recent papers /18,19/. The essence of this idea is that reconnection on the day side is closely controlled by the IMF and that ionospheric currents related to this process closely follow variations in the IMF. However, energy stored as magnetic flux in the tail by the process must eventually be unloaded. Frequently this occurs at times not closely correlated with changes in the IMF. Ionospheric currents driven by this nightside process thus appear on a short time scale to be uncorrelated with the IMF. Magnetic activity indices contain components related to both types of processes and hence are not completely predictable.

This paper presents a review of our recent work concerned with IMF control of geomagnetic activity. The work was motivated by our desire to express quantitatively the temporal relation between the solar wind input and the ionospheric output from the magnetospheric system. We have done this using linear prediction filtering. This technique treats the magnetosphere as a black box characterized by an impulse response (Fourier transform of the transfer function). Empirical data are used to determine the characteristics of the impulse response. We show that an average impulse response can account for only about 40% of the variance in the AL index. The residual is clearly related to the IMF, but not in a deterministic sense as might be expected for a primarily driven system. Surprisingly, we find that for individual events, both the growth phase and expansion phase currents appear to be quite closely proportional to the rectified electric field of the solar wind. However, the scale factors and delays change from event to event in an unpredictable manner thus reducing the predictability of the system output.

THE LINEAR SYSTEM MODEL

Definition of a Linear, Time Invariant System

A physical system is said to be linear if the output of the system y(t) is linearly related to the input x(t) /20/. Such systems can be described by functional relationships of the form

$$D^{(1)}[y(t)] = D^{(2)}[x(t)]$$

where $D^{(1)}$ and $D^{(2)}$ are linear differential operators. If the coefficients in the differential operators do not depend on time the system is said to be time invariant. This differential equation can then be converted to an algebraic equation by Fourier transformation

$$FT\{y(t)\} = \int_{-\infty}^{+\infty} g(\tau)x(\tau-t)d\tau \tag{1}$$

This equation may be rewritten in the form

$$Y(f)/X(f) = FT\{y(t)\}/FT\{x(t)\} = FT\{D^{(2)}\}/FT\{D^{(1)}\}$$

where X(f) and Y(f) are the Fourier transforms of corresponding input and output to the system. This equation implies that the properties of a physical system may be determined using measured input and output signals.

Definitions of Impulse Response and Transfer Function

The ratio of the Fourier transforms of the output and input is called the transfer function, G(f). It is a complex quantity whose amplitude describes how much the system attenuates a given frequency component of the input, and whose phase describes how much it delays this component.

$$G(f) = Y(f)/X(f) \tag{2}$$

If the transfer function of a system is known, and an arbitrary input is applied to the system, it is possible to predict the output from the relation $Y(f) = G(f)X(f)$. this expression may be inverse transformed to the time domain giving

$$y(t) = \int_{-\infty}^{+\infty} g(\tau)x(\tau-t)d\tau$$

The inverse transform of the transfer function is the impulse response of the system, g(t). This integral equation is called the convolution equation. It provides a method of predicting the temporal output of a physical system in response to an arbitrary input. For real systems there is no output before the input so that $g(t) = 0$ for $t < 0$. Real systems also exhibit dissipation which causes the impulse response to die away after a finite time so that $g(t) = 0$ for $t > t_o$. With these constraints the convolution integral becomes

$$y(t) = \int_{o}^{\tau_o} g(\tau)x(\tau-t)d\tau \tag{3}$$

If the input and output are sampled at equispaced intervals $t_i = i\Delta t$ the integral may be approximated by a finite summation

$$y_i = \sum_{j=0}^{N}(g_j\Delta t)x_{i-j} = \sum_{j=0}^{N} f_j x_{i-j} \tag{4}$$

The coefficients f_j are called the coefficients of the linear prediction filter characterizing the physical system.

PREDICTION FILTERS FOR THE AL INDEX

Calculation of the VBs-AL Filter

The technique of linear prediction filtering has been used by Clauer *et al.* /15/ to study the coupling of the solar wind to the westward electrojet. This study assumed that the AL index measuring the strength of the electrojet is related to a solar wind coupling parameter

through a linear, time-invariant system. Three possible coupling parameters were investigated including the solar wind electric field (VB_s), the epsilon parameter ($VB^2Sin^4\theta/2$), and an electric field with velocity dependent merging efficiency (V^2B_s). Response functions (prediction filters) were calculated for each of the coupling functions using data corresponding to two different levels of magnetic activity. Figure 1 taken from this work demonstrates that all of the empirical filters (thin lines) are qualitatively similar. Each filter has the form of a miniature negative bay peaking at one hour with a total duration of roughly 2.5 hours.

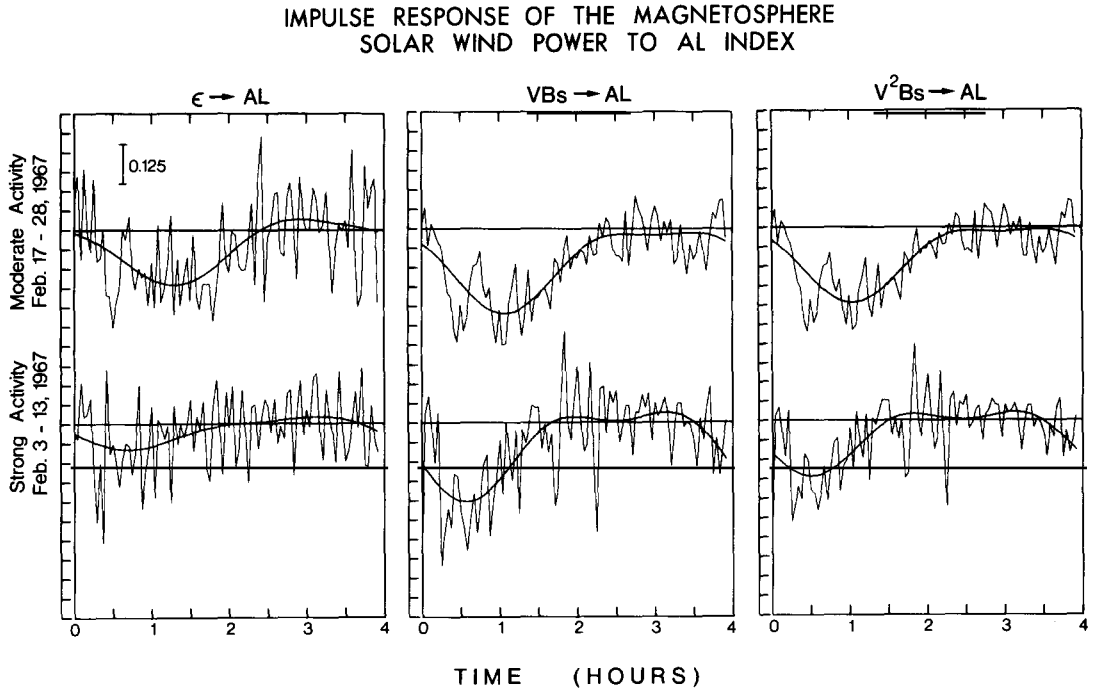

Fig. 1. Prediction filters (thin lines) relating several possible solar wind coupling parameters including epsilon, VB_s and V^2B_s to the AL index. Filters are shown for moderate and disturbed levels of magnetic activity. The general form of the filters is emphasized by smoothing (heavy lines).

Differences in the filters presumably caused by a non-linear relation between the input and output can be seen by comparing the filters for different solar wind coupling functions. The epsilon parameter has very large variance about the low pass filtered version of the response function while the V^2B_s parameter, which empirically is close to optimum /21,22/, has the least variance. Effects due to temporal changes in the system parameters are also evident from a comparison of filters for two different levels of activity. The moderate activity response functions for VB_s and V^2B_s are bi-modal with two peaks at roughly 40 and 60 minutes while the corresponding strong activity filters is uni-modal with only one peak at 30 minutes.

Prediction filters for VB_s-AL coupling are quite similar from interval to interval as illustrated in Figure 2. Thin lines in the diagram represent filters calculated for six different, 7-10 day intervals in the years 1967-1968. The heavy line is a single filter calculated for the entire 52-day data set. Individual filters are similar to each other and to the filter calculated from all of the data. Most of the filters exhibit a hint of the bi-modal response seen in the moderate activity filters of Clauer et al. .

Similar results for epsilon-AL coupling are presented in Figure 3. Although the filter determined from all of the data is very similar to the corresponding VB_s-AL filter, the higher noise level of the epsilon filters is apparent. We attribute the difference to the dependence of epsilon on the product VB^2 rather than VB as in VB coupling. Figure 1 suggests that the product V^2B_s would produce individual filters less variable than the electric field (VB) although we have not checked this.

The transfer functions for the VB_s and epsilon coupling functions have been calculated by Fourier transformation of the 52-day filters shown by heavy lines in Figures 2 and 3. The results are presented in Figure 4. Amplitude responses are at the top and phase responses are at the bottom. The amplitude responses for the 52-day filters show that all solar wind variations with frequencies higher than 0.1 mHz are significantly attenuated in coupling to

the AL index. The linear variation of phase with frequency in the range 0-0.5 mHz indicates that all frequency components of the solar wind input are uniformly delayed in the output by roughly 25 minutes. This delay includes approximately seven minutes of solar wind propagation from the average location of the solar wind monitor.

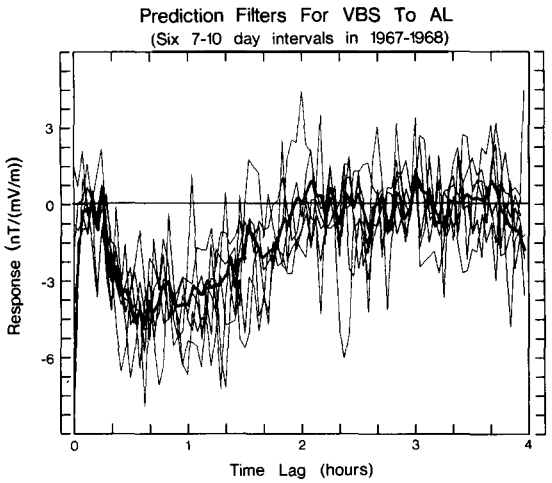

Fig. 2. A comparison of VB_s-AL prediction filters calculated from seven day segments (thin lines) with the average filter for a 52 day interval (heavy line).

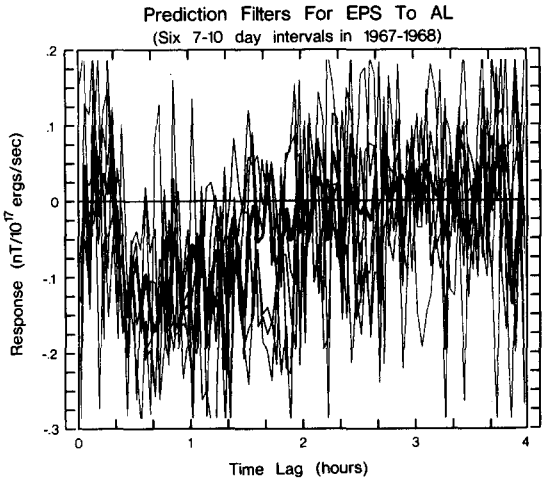

Fig. 3. An illustration of the effect of using an input function (epsilon) which is not linearly related to the output function.

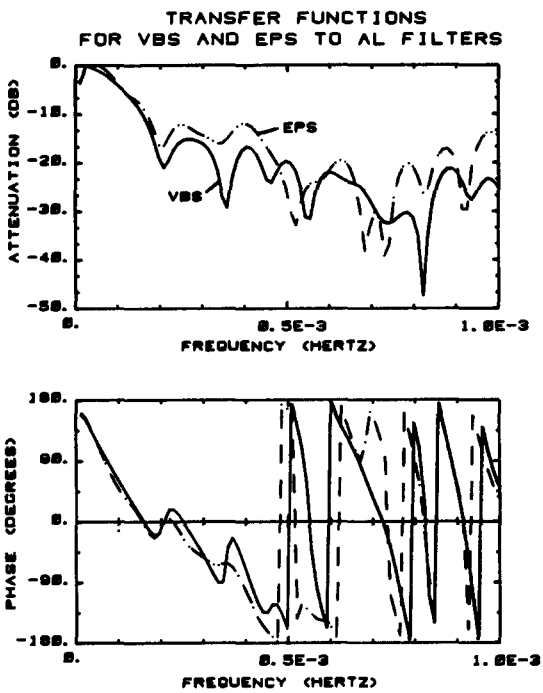

Fig. 4. The transfer function for the VB_s-AL filter. Top panel- shows the logarithmic attenuation as a linear function of frequency. Bottom panel- shows the phase response.

A qualitative impression of the degree to which the 52-day filter fits the data from which it is derived is provided by Figure 5. The two panels contrast moderate (top) and strong (bottom) levels of magnetic activity. Thick lines depict the AL index as predicted by the convolution of the average response function with the input while thin lines are the observed AL index. There are significant discrepancies between the observed and predicted index. The most obvious is the filter's inability to predict substorm expansion phases. These can be identified by sudden decreases of the observed AL index below the predicted index. A clear example is seen in the second panel at about 14 hours. This event follows a northward turning of the IMF which causes the rectified electric field to become zero. No system with finite impulse response can predict an output at times later than the duration of the filter. Another form of discrepancy is exemplified by events such as that in the top panel at about 16 hours where an expansion phase occurs during relatively constant input. When the input is constant the output of a linear system must also be constant after a time interval equal to the duration of the system response function.

Changes in VBs-AL coupling with the level of the input suggest that one of our assumptions concerning the system relating the solar wind and the AL index is incorrect. Bargatze et al. /23/ investigated this possibility using the method of piecewise linear approximation. The authors divided their input data into short time intervals characterized by the median AL index for each interval. The segments were then reordered so that

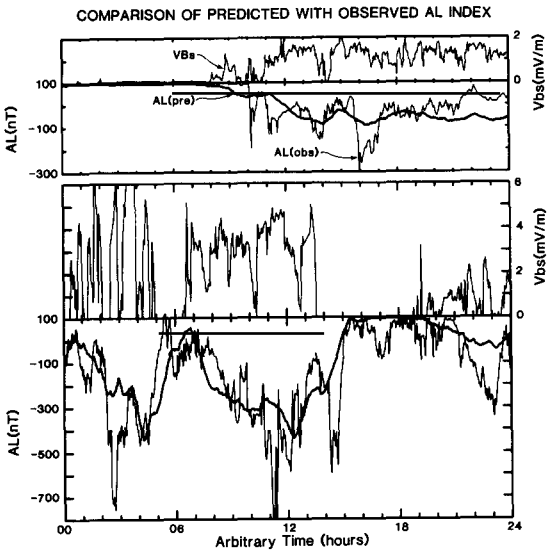

Fig. 5. The time variation of the AL index as predicted by the average VB_s-AL filter compared to some of the data from which it was generated. Top frame shows moderate activity and bottom frame disturbed activity.

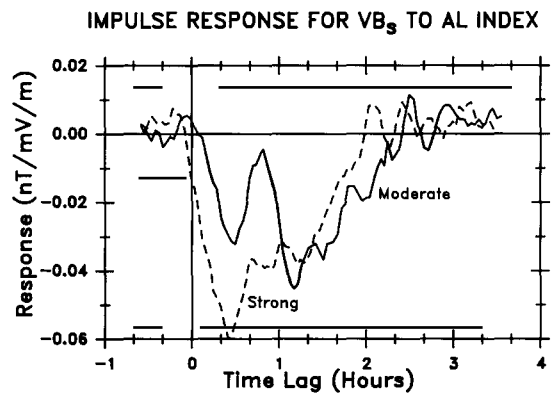

Fig. 6. Filters for VB_s-AL calculated for two different levels of magnetic activity showing the bi-modal character of the magnetosphere's response at moderate levels of activity.

successive intervals had progressively increasing median AL. Filters were generated for 30 overlapping sets of five segments. The results displayed in Figure 6 do not show a gradual transition in the shape of the response function as the level of activity increases as might be expected if the system parameters change gradually with an increase in activity. Instead there appears to be two modes of response for the magnetosphere. At moderate levels of activity the response function is bi-modal with peaks at 25 and 70 minutes. (Slight differences in time delays between the two studies are probably due to corrections for solar wind propagation made in the Bargatze et al. study.) In contrast, during strong activity the amplitude of the first peak increases dramatically relative to the peak for moderate activity. A plateau in the strong activity response function suggests that the second peak is still present but is no longer visible because the tail of the first peak is superimposed. The main difference in the AL index predicted by the two filters is that during strong activity the initial response to the solar wind is much stronger than it is during moderate disturbance. This change causes the strong activity intervals to appear as more directly driven by the solar wind than are moderate activity intervals.

PREDICTION FILTERS FOR MODEL SYSTEMS

A Simple Model of Solar Wind-Magnetosphere Coupling

The double peaked response function relating VB_s to the AL index obtained by Bargatze et al. for moderate activity suggests that two distinct physical processes control the westward electrojet at such times. Because the first peak represents the initial response of the electrojet to a southward turning of the IMF we associate it with the substorm growth phase and magnetospheric convection driven by dayside reconnection. In the terminology of Rostoker et al. /24/ the growth phase is produced by processes driven by the solar wind. Because the second peak is centered at about one hour we associate it with the expansion phase and convection driven by nightside reconnection. According to the discussion of Rostoker et al. the expansion phase is driven by unloading processes. We note, however, that the bi-modal VB_s-AL response implies that this internal unloading process is controlled by the solar wind electric field to some extent. Such a relationship might be expected if the rate of nightside reconnection is proportional to the electric field imposed on the magnetosphere by the solar wind once it begins.

These results suggest that it may be possible to model the response of the westward electrojet to the rectified solar wind electric field by a response function consisting of only two delta functions of fixed delay and amplitude. Thus for f_j in equation (4) we assume

$$f_j = a_j d_j + a_k d_k$$

The convolution theorem then implies that

$$AL(i) = a_j VB_s(i-j) + a_k VB_s(i-k)$$

If the time lags j and k are known then the scale factors a_j and a_k (filter coefficients) can easily be determined by least square techniques. If we minimize the variance in the residual between the model and the observations, the scale factors are obtained from equations identical to those normally used in linear prediction filtering (e.g. /25/). Since there are only two unknowns the calculation reduces to a set of two linear equations in which the multipliers of the unknown coefficients are the autocovariance of the input at the assumed lags, and the right hand side of the set is a column vector with elements which are the cross covariance between the input and output at the two lags. The normalized minimum variance can then be determined from the expression

$$P_{min}/A_{oo}(0) = 1 - \sum_{1} C_{oi}(1)a_1/A_{oo}(0)$$

where $A_{oo}(0)$ is the autocovariance of the output at zero lag and $C_{oi}(1)$ is the covariance between the input and output at lag 1. The last term on the right hand side is called the prediction efficiency since it can be shown to have a value of 1.0 if the model perfectly fits the observations, and a value of 0.0 if there is absolutely no covariance between input and output.

The variation of the prediction efficiency with lag provides a method for determining the optimum delay in the bi-modal model. We start with a guess for the growth and expansion time lags j and k obtained by visual examination of the data for a given substorm. We then perform a grid search varying j and k independently over a small range about the initial guess. For each pair of lags we calculate the prediction efficiency. A contour map showing the efficiency versus lag j and k is plotted and the optimum lags chosen from the map. These lags are then used in the subsequent construction of the predicted AL index.

Filter for a Strictly Driven (Uni-modal) System

For some years Akasofu has argued that the magnetosphere is "primarily a driven system" (see for example /26/). In terms of the model discussed above a directly driven system would be uni-modal, having a response function consisting of a single delta function. The parameters in a uni-modal model could be determined by a procedure similar to that described above. A more familiar approach gives the same results, however. The optimum single lag is determined by choosing the lag of maximum cross correlation between the input and output. The scale factor is then found by a linear regression of the output against the input delayed by the optimum lag and subject to the constraint that zero input produces zero output. Results obtained with this procedure are summarized in Figure 7.

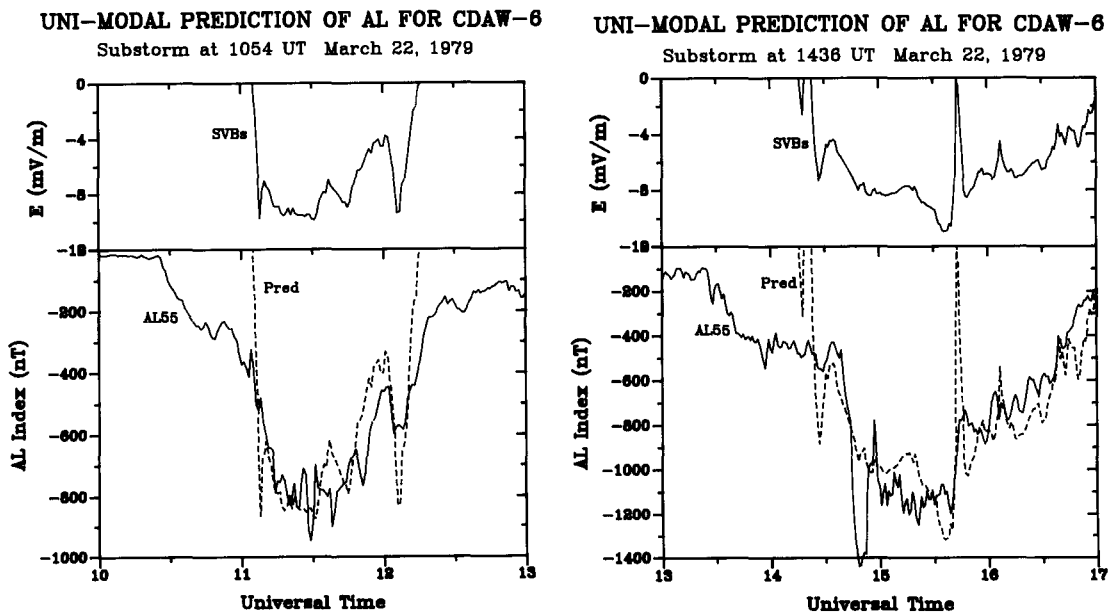

Fig. 7. The optimum fits of a uni-modal response function to two substorms during the CDAW-6 interval of 22 March 1979. The top panel in each frame shows the optimum shifted input. The bottom panel shows the AL index (solid line) and the optimum fit (dashed line).

The two panels of Figure 7 show results for two substorms on 22 March 1979 studied in the CDAW-6 workshop. These events were chosen for initial analysis because of the high quality auroral electrojet indices available for this day, and for their well established substorm chronologies /27/. In both cases it can be seen that the best the uni-modal can do is approximately fit the expansion phase of each substorm. In neither case does the model account for any of the variations of AL during the growth phases. This simple two parameter model accounts for 75% of the AL variance in the 1054 UT substorm, and 55% in the 1436 UT substorm.

One point concerning these results should be emphasized. The scale factors and time delays differ significantly for two substorms separated by only one hour on the same day. The differences suggest that the onset of the expansion phase is not directly caused by the solar wind electric field as assumed in the driven model, and that the strength of the current that flows during this phase is controlled by internal parameters (such as ionospheric conductivity produced by preceding events). However, once the expansion phase begins the form of the variations of the westward electrojet is roughly proportional to the rectified solar wind electric field.

Filters for a Bi-modal System

Results from the bi-modal model are similar to those from the uni-modal model although a somewhat different procedure was used to obtain them. Figure 8 summarizes some details of the procedure described earlier. The left panel presents a contour map of prediction efficiency as a function of assumed growth phase lag along the vertical axis, and expansion phase lag along the horizontal axis. Optimum lags of 27 and 57 minutes are easily indentified. Together these lags produce a prediction efficiency exceeding 90%. The right panel illustrates how the model approximates the observed AL index. The curve labeled FITGRO is a fit to the growth phase and FITEX is a fit to the expansion phase. The curve FITSUM is their sum. The curve FITTOT is the same as FITSUM except it has been displaced to give it the same mean as the observed AL index The curve labeled AL is included for comparison and shows the quality of the fit obtained.

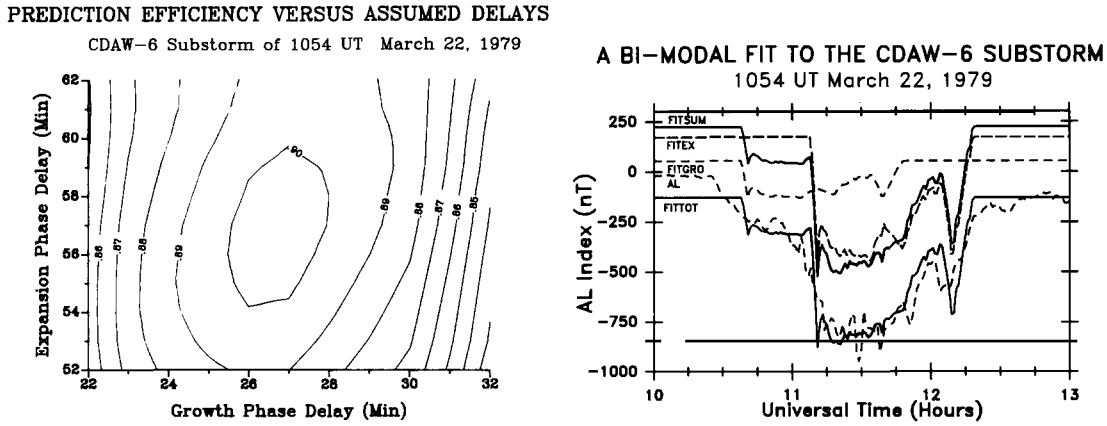

Fig.8. An example showing how the first CDAW-6 substorm is fit using the bi-modal response function. The left panel displays a contour map of the prediction efficiency as a function of the time delays chosen for the growth and expansion phase time delays. The right panel shows how two, time-delayed and scaled versions of the solar wind input are summed to approximate the observed AL index.

Fits of the bi-modal model to the two substorms of 22 March 1979 are presented in Figure 9. In both panels dashed lines are the observed AL index and solid lines are the bi-modal fits. It is interesting to note that in both substorms, a major change in the solar wind electric field appears to be reproduced in the AL index after nearly one hour's delay. For the 1054 UT substorm in the top panel such a change can be seen at about 1212 UT. For the 1436 UT substorm a major fluctuation reappears twice at 1450 and 1545 UT. This suggests that not only the gross structure of variations in the solar wind electric field is reproduced in the Al index, but some of the details may be as well.

A comparison of a number of significant time delays for the two substorms is made in Table 1. During the events the solar wind monitor IMP-8 was located on the dusk side of the earth-sun line about even with the bow shock. The dynamic pressure-D_{st} delay (5 minutes) is a good measure of solar wind propagation effects and should be used to correct other time delays calculated relative to the southward turning of the IMF. It should be noted that the corrected bi-modal growth delay is not equal to the time between southward turning and initial AL response, nor to the duration of the growth phase. It is the time required to

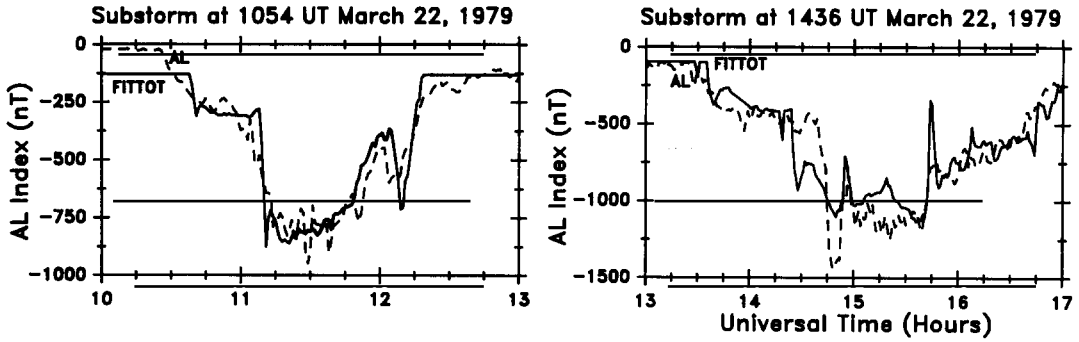

Fig. 9. A comparison of AL as predicted by the simplified bi-modal response model (solid line) to the measured AL index (dashed line) for the two CDAW-6 substorm events. Left panel shows a fit to the 1054 UT substorm. Right panel shows a fit to the 1436 UT substorm.

give the optimum fit of the waveform of the electric field to the observed AL index during the growth phase. The duration of the growth phase and the uni-modal and bi-modal expansion delays are quite similar, however. Exact agreement would be unlikely even if they are measures of the same phenomenon since the initial response of the AL index to an onset is often delayed due to substorm expansion over the finite network of AE stations. Since original magnetograms and other data are normally used to determine the substorm onset, the duration of the growth phase is somewhat shorter than times indicated by the AL index. It should also be noted that none of the substorm phases has a duration exactly equal to the duration of the southward IMF, so neither the substorm phases, nor the entirety of the substorm can be explained simply as a time delayed and scaled version of the input as assumed in the driven model of substorms.

The uni-modal scale factors are uniformly larger than the bi-modal factors by about 1.6. This is expected since the uni-modal model uses only one delayed version of the input while the bi-modal model superposes two at the time of maximum disturbance. The scale factors for the uni-modal model are very similar to results on the proportionality of VB_s and AL obtained by Holzer and Slavin /22/.

TABLE 1 Model Parameters for CDAW-6 Substorms 22 March 1979

Model Parameter	1054 UT	1436 UT
Dynamic pressure delay	4 min	4 min
Duration of southward IMF	80 min	200 min
Initial response delay	13 min	13 min
Bi-modal growth delay	27 min	25 min
Duration of growth	34 min	76 min
Uni-modal expansion delay	54 min	72 min
Bi-modal expansion delay	56 min	72 min
Duration of expansion	56 min	54 min
Duration of recovery	70 min	90 min
Bimodal growth scale	17 nT/mV/m	37 nT/mV/m
Uni-modal expansion scale	88 nT/mV/m	118 nT/mV/m
Bi-modal expansion scale	62 nT/mV/m	73 nT/mV/m
Uni-modal efficiency	77%	58%
Bi-modal efficiency	90%	83%

SIMULATION OF THE BIMODAL RESPONSE

The preceding results demonstrate that in some substorms almost the entire variation of the AL index can be represented by a model having only four parameters. However, because these parameters are not known in advance and change with time, it is not possible to predict the index from solar wind data alone. On the other hand, our linear prediction results show that about 40% of the variance is predictable if one uses an average model (prediction filter) with about 100 fixed parameters. This suggests that there can not be too great a variation in the time delays or scale factors of the simple bi-modal model, for if there were, the average prediction filter would not represent the data as well as it does. To demonstrate this we have carried out a simulation of a bi-modal magnetosphere as described next.

The Stochastic Model

Our model assumes there are two types of processes which contribute to the AL index, driven processes and unloading processes, so that $AL(t) = AL_D(t) + AL_X(t)$. We assume the driven component, $AL_D(t)$, is entirely deterministic and characterized by an impulse response function, $g_D(t)$. Thus $AL_D(t) = g_D(t)*I(t)$. To be roughly consistent with empirical results we take $g_D(t)$ to be a Gaussian pulse with center time, amplitude, and width similar to the first peak in the moderate activity filter of Bargatze et al. /23/ (see Figure 6).

The unloading component of our model was represented by the convolution of a Gaussian function $g_X(t)$ with an impulse $\delta(t-t_d)$ located at a time corresponding to the end of the expansion phase. To simulate growth phases of various lengths the center time of the Gaussian, t_d, was changed randomly between 45 and 105 minutes while its height was held constant with amplitude comparable to that of the second peak in the results of Bargatze et al. The time delay t_d of the impulse function is initially the time from a southward turning of the IMF to the end of the first expansion. Subsequently, provided the IMF stays southward, it is the time from the end of the preceding expansion to the end of the subsequent one. To avoid a sequence of many successively smaller substorms during northward IMF only one expansion was allowed after a northward turning. To simulate expansion and recovery phases of different duration the width of this Gaussian was varied in a random manner (between 30-60 minutes). To make the size of an unloading event proportional to the IMF the impulse function was multiplied by an "energy" factor *E*. This was defined as the time integral of the input rectified electric field, diminished by any previous expansions. Since typical expansion phases do not unload all available energy only a fraction f of the energy factor was unloaded in a given event. The integration is continuous from the time of the first southward turning, but each time an expansion occurs *E* is reset to (1-f) of its value at the time of the onset. The factor f was varied randomly between 60-100% so that the most probable energy release was 80%. Thus $AL_X(t) = fEg_X(t)*\delta(t-t_d)$

To complete the simulation we generated a time history of the rectified solar wind electric field. Our model did this in two ways depending on the level of magnetic activity. For moderate activity we assumed that the input consisted of a sequence of intervals of square pulses separated by comparable intervals with zero input. The duration of the non-zero intervals was varied randomly between 1-3 hours, the amplitudes of successive pulses was varied between 2-6 mV/m, and the separation of pulses was varied between 2-6 hours. For disturbed times these parameters were doubled so that there were fewer intervals of stronger amplitude and considerably longer duration.

Given the stochastic model of the magnetospheric response described above and two time histories of the solar wind input, we generated corresponding histories for the AL index. This gave two sets of input and output data for a simulated magnetosphere. These data were then analyzed by the same programs used by Bargatze et al. to analyze actual observations. Prediction filters corresponding to moderate and disturbed activity were obtained and then used to predict the data from which they were derived. The results are described next.

Results of the Model

Prediction filters produced by analyzing the simulated magnetosphere are presented in the left panels of Figure 10. For comparison, the empirical filters of Bargatze et al., /23/ are shown on the right. The filter representing simulated moderate activity is bi-modal and similar in shape to the filter obtained from actual data. The simulated strong activity filter tends to be uni-modal and is also quite like the filter obtained from actual data. Also, the filters change in the same way as real filters do with a change in the level of activity. The second peak decreases in amplitude, while the first peak increases.

The ability of the two filters to model the simulated data from which they were generated is illustrated in Figure 11. The right panel shows a segment of moderate activity. The top trace is the simulated input including a small amount of random noise. The pulsed nature assumed for this type of activity is quite apparent. The second pair of traces shows the simulated output (solid line), and the prediction of this output using the filter derived from the data (dashed line). The efficiency of the prediction is comparable to that obtained with real data and filters. The second substorm is almost perfectly predicted. However, it is obvious that the filter cannot predict two expansions resulting from one pulse of southward IMF. The best that it can accomplish is to predict the average activity.

The failure of the filter to predict expansion phases during simulated strong activity in even more apparent in the left panel of Figure 11. Because we assumed that strong activity is produced by long pulses of southward IMF our prescription allows a number of expansions to occur during each such interval. The filter, however, can only approximately predict the first of these. This is because only the first expansion is even roughly correlated with a change in the input. After a time equal to the duration of the filter the output becomes

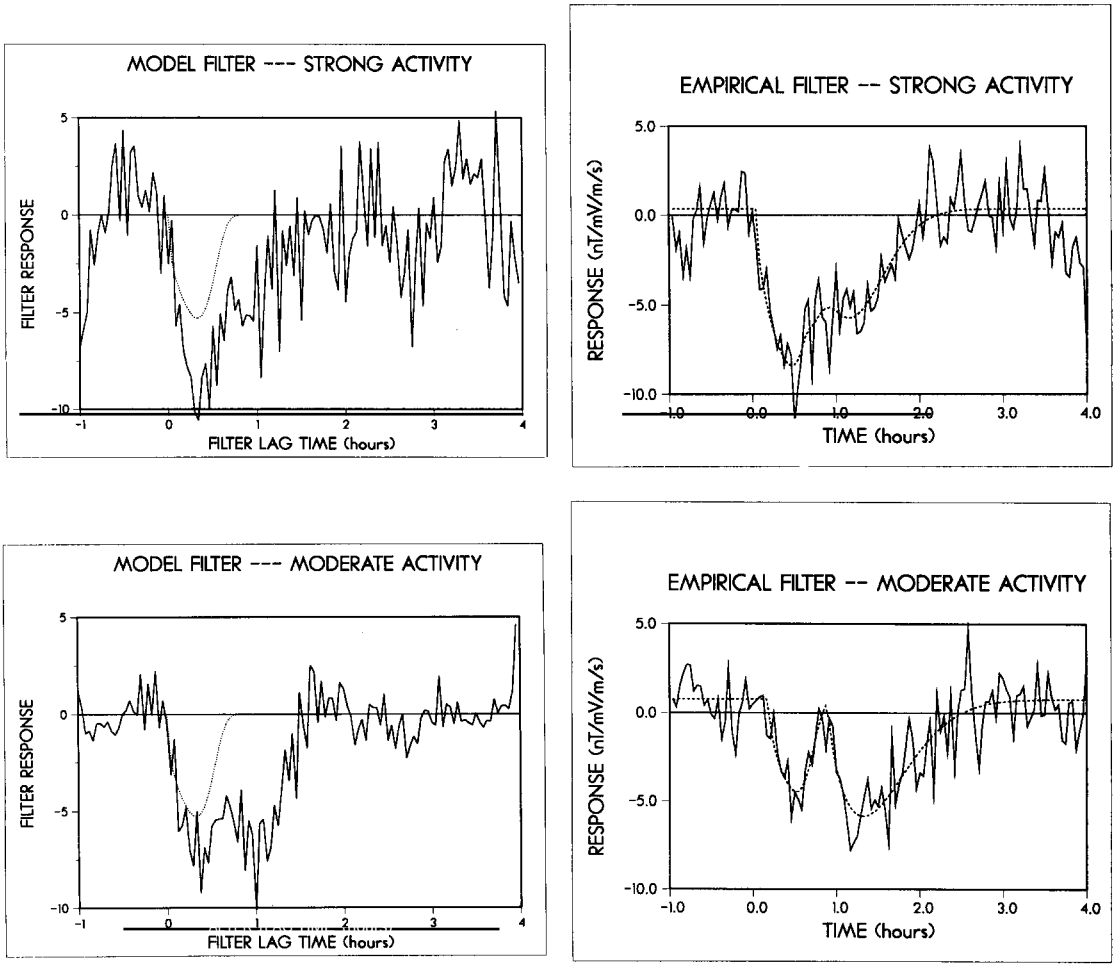

Fig. 10. A comparison of response functions calculated with data generated by the stochastic substorm simulation with response functions generated from actual observations.

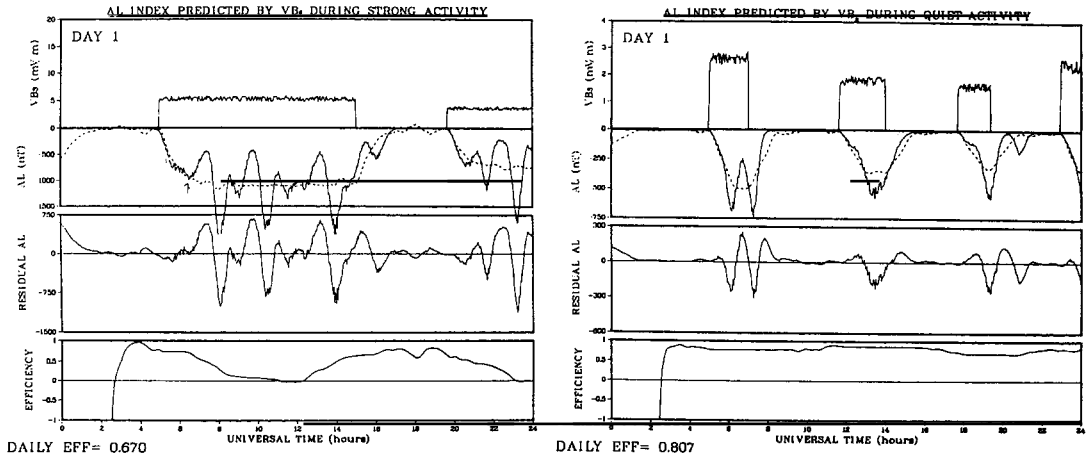

Fig. 11. A comparison of the AL index predicted by filters generated from simulated data with the data used to generate the filters. The right panel shows moderate activity defined as a sequence of isolated substorms. The left panel shows disturbed activity consisting of long intervals of continuously southward IMF. Solid lines are the model data and dashed lines are the predictions.

constant. This constant value is approximately the mean of the variations produced by all of the expansion phases, plus the effect of the driven processes. Apparently, in the calculation of the prediction filter by least square techniques, the additional offset required to approximate the average effect of the sequence of expansions, appears as an enhancement of the first peak in the prediction filter. Thus when the strong activity filter is convolved with the input it predicts an AL index stronger than expected on a basis of the driven component as determined during times of more moderate activity.

DISCUSSION

The work described here demonstrates that substorms are produced by the superposition of two main processes. Using terminology suggested in several recent papers /19,24,28/ these are the driven and unloading processes. Driven processes are those associated with direct coupling of the solar wind to ionospheric currents. Physically, the driven processes are caused by reconnection of the solar wind magnetic field with the earth's magnetic field on the dayside magnetopause. Reconnection drives magnetospheric convection transporting magnetic flux from the dayside to the lobes of the magnetotail. It also stimulates a return flow of closed magnetic flux from the magnetotail plasma sheet. The electric field produced by the convecting flux is projected onto the conducting ionosphere by the nearly infinitely conductive magnetic field lines. Because of spatial gradients in the electric field applied to the ionosphere, and gradients in ionospheric conductivity, ionospheric currents diverge along magnetic field lines. Thus, after some time delay due to induction, fluctuations in the rate of dayside reconnection appear as variations in the strength of the ionospheric currents. These currents appear to be directly driven by the solar wind dynamo.

The onset of unloading processes is a natural consequence of the driven processes. In a magnetosphere driven by dayside reconnection, closed magnetic flux is only available from within the nightside plasma sheet. As time progresses this shrinks as flux is transported to the day side. Observations of changes associated with this shrinkage are called the growth phase of a magnetospheric substorm. Eventually, new flux must be added to the plasma sheet from the reservoir of closed flux in the tail lobes. This is done by nightside reconnection. Significant nightside reconnection does not begin immediately after the onset of dayside reconnection. Typically about one hour of growth phase changes are required before the magnetosphere becomes sufficiently unstable to allow the process to become significant in the near-earth plasma sheet. The duration of the growth phase is probably controlled by the initial amount of flux in the plasma sheet, the rate at which it is withdrawn, and external factors such as triggering pulses. Reconnection apparently begins slowly on closed field lines within the plasma sheet. Eventually the last closed field lines bounding the plasma sheet are reconnected and open field lines in the opposing lobes drift together at the center of the plasma sheet at a rate dependent on the electric field across the tail lobes.

Once nightside reconnection begins it "unloads" the magnetic energy stored in the lobes converting it to kinetic and thermal energy of plasma sheet particles. The electric field induced by reconnection drives plasma towards the day side, and additional currents into the nightside ionosphere. Since nightside reconnection is initially confined within a small region of the nightside plasma sheet, its effects are projected into a small sector of the midnight auroral ionosphere. Observations of these effects are called the expansion phase of the substorm. If the rate of nightside reconnection is proportional to the electric field in the lobes, then it might be expected that the ionospheric currents driven by the unloading process are also proportional to the solar wind electric field. These currents are superimposed on those directly driven by the solar wind.

This model suggests that there should be a close relation between the solar wind electric field and currents in the ionosphere. During the substorm growth phase ionospheric currents are driven by dayside reconnection. Since the solar wind electric field controls the rate of reconnection we expect that measures of these currents such as the AL index are roughly proportional to the strength of the solar wind electric field. On the night side the situation is more complex. Magnetic flux cannot drift across the opposing tail lobes until reconnection begins. Since this is delayed until appropriate conditions develop in the plasma sheet, magnetic flux must accumulate in the tail lobes. Once reconnection begins magnetic flux drifts into the reconnection region as a result of the electric field across the tail. Since this is generated by the connection of the lobe field lines to the solar wind, we expect the drift rate in the lobes, and hence the rate of reconnection, is again proportional to the solar wind electric field. Plasma ejected earthward from nightside reconnection creates an electric field which is also projected to the ionosphere driving currents. Sharp gradients in this electric field combined with sharp gradients in conductivity produced by precipitating particles result in a field aligned current connected to the reconnection region. It is not unreasonable to assume this current, or AL which monitors the ionospheric portion of this current, is also proportional to the rate of reconnection and hence to the solar wind electric field.

A convenient technique for investigating the validity of this model is linear prediction filtering. This technique provides a simple means whereby empirical data establishes the most general linear relation between the input to, and output from, a physical system. This technique calculates the impulse response of the system defined as the Fourier transform of the system transfer function. The transfer function can be determined empirically by dividing the Fourier transform of the output by the transform of the input. Once the impulse response is known it may be convolved with an arbitrary input signal to predict the output of the system, hence the name prediction filtering.

We have calculated the impulse response relating the rectified solar wind electric field (VB_s) to the AL index measuring the strength of the westward electrojet. We found that the gross structure of this response function is a miniature negative bay peaking at 1 hour and having duration of 2.5 hours. We demonstrated that different coupling functions produce similar response functions, but the noise in the calculated results is much higher for parameters known to be non-linearly related to AL.

In an investigation of possible non-linear relations between the assumed coupling function and the AL index we found changes in the response function with the level of activity, but the changes were not continuous. There seems to be two modes of magnetospheric response mixed together in the data. Moderate activity made up of a sequence of isolated substorms has an impulse response function which is bi-modal. The first peak in the response function occurs at about 20 minutes and the second at 60 minutes. The two peaks have comparable areas and hence predict roughly comparable levels of activity. Because of the timing of these pulses, and the known structure of isolated substorms, we attribute the two peaks to growth phase and expansion phase currents respectively. When this impulse response is convolved with the solar wind electric field it predicts that substorm currents consist of the superposition of two parts each proportional to the rectified solar wind electric field. This result supports the superposition model outlined above.

Encouraged by this result we investigated how well individual substorms can be characterized by a bi-modal response function. We assumed that the impulse response consists of only two non-zero terms at delays corresponding to the growth phase and expansion phase. This model contains only four parameters, the two time delays, and the scale factors relating currents at these delays to the solar wind. We then developed an iterative procedure for determining the optimum delays and the corresponding scale factors. Our results are very surprising! Between 80% and 90% of the variance in the AL index is predicted by our model. This should be contrasted to the 40% predictability achieved with the average bi-modal response function. This result provides even stronger support for the superposition model, and if fact, lends credence to the its fundamental assumption that both dayside and nightside reconnection rates are directly proportional to the solar wind electric field, although with different time delays.

An examination of the bi-modal response parameters obtained for two successive substorms during the CDAW-6 interval explains the difference in predictability between average and individual events. The parameters for successive substorms are different. Reasons for this difference are easy to suggest. Precipitation from the first substorm increases ionospheric conductivity thereby altering the scale factors between the solar wind and ionospheric currents. The time required for nightside reconnection to begin is longer in the second substorm because flux from the preceding one has filled the plasma sheet. Regardless of the cause, an average prediction filter is not able to do as well as an individual filter because it must try to account for changes between substorms by averaging over all observed delays and scale factors.

To assess the validity of this explanation we performed a simple simulation of the electrojet response to the solar wind. Our model assumed that two types of processes contribute to the AL index, directly driven processes and unloading processes. The directly driven part was taken to be completely deterministic and given by convolution of a Gaussian pulse with the input. For the unloading part we assumed that the output was related to the time integral of the input through several stochastic parameters controlling the fraction of available energy unloaded, the time between unloading events, and the duration of the unloading process. In this way the relation between the simulation input and output was made partially random. Two essential features of this simulation are the assumption that moderate activity is driven by a sequence of electric field pulses with duration shorter than the system response function, and the constraint that no more than one unloading can occur after coupling ceases. When the simulated moderate activity data were analyzed by the prediction filter technique we obtained a bi-modal response function as we did for real moderate activity data. This result demonstrates that the bi-modal response function obtained from real magnetospheric data can be produced by the superposition of two processes as we have suggested.

The simulation provides additional support for our superposition model in the results obtained from the simulation of strong activity. For the purposes of the simulation we

defined strong activity as that produced by intervals of solar wind coupling with duration substantially longer than the duration of the observed impulse response function, and magnitude larger than that during moderate activity. Analysis of simulated strong activity data produced an impulse response with only one peak at about 20 minutes. This peak was substantially larger than the first peak of the moderate activity response function. These results are the same as found with real data.

Our explanation of the difference between the moderate and strong activity filters in the simulation is the following. Moderate activity was simulated by numerous short coupling pulses. Part of the output was directly related to the input by convolution. Another part was produced by a pulse whose area was roughly proportional to the time integral of the input, and whose time of onset was roughly fixed relative to the beginning of coupling. Occasionally a second expansion occurred after coupling ceased, but the model constraint allowed at most one such event. Thus, overall most of the output was roughly correlated with the input.

In contrast, we assumed disturbed activity is driven by long intervals of continuous coupling. During such intervals our simulation produced many unloading events which on the average had strength proportional to the input. Only the first of these unloadings occurred at a time even roughly correlated with the beginning of solar wind coupling. Subsequent unloadings occurred with delays which were randomly distributed about an average delay from the time of the preceding unloading. This caused later unloadings to be almost entirely uncorrelated with the beginning of coupling. Consequently the filter analysis approximates the effects of these later unloadings by their average. This appears as an enhancement of the first peak in the response function.

We believe that the assumptions of our simulation may be true for the real magnetosphere. If this is the case the single strong peak obtained in our filter analysis of real magnetospheric data does not imply that the magnetosphere is directly driven. Instead it indicates that unloading events are occurring so randomly in relation to changes in the solar wind input that the best that filter analysis can do is predict the average of their effects. Examination of real data during magnetic storms suggests that this is usually the case. It is rare that a long interval of relatively steady solar wind coupling produces a steady AL index as expected for a directly driven system.

We note, however, that occasionally a situation occurs in which many of the usual indicators of unloading are not observed. Such intervals have been called convection bays by Pytte and co-workers /29/. It is possible that such events could be caused by a stable, near-earth neutral line which reconnects open flux on the night side nearly as fast as it is created on the dayside. Such a situation would truly be a "driven system" with storage and release of energy relegated to relatively insignificant intervals at the beginning and ending of such events.

As stated at the beginning, our application of the technique of linear prediction filtering to solar wind-magnetosphere coupling was motivated by our belief that the magnetosphere is not directly driven by solar wind coupling in the sense expected for a deterministic system. Instead, we assert that the typical interaction consists of the superposition of two types of processes. The first type are directly driven by solar wind coupling and hence appear to be well correlated with the input. These processes are responsible for the storage of energy in the magnetotail. The second type are of internal origin and are consequences of the unloading of the stored energy. Since the amount of energy unloaded is loosely related in time and magnitude to the solar wind input they also appear to some extent to be related to the input. However, because these processes are constrained by internal magnetospheric parameters which change with time, it is not possible to predict the temporal behavior of the output of the system from solar wind data alone. We note, however, once unloading begins the rate of reconnection seems to be controlled (or driven) by the solar wind electric field. Thus, in conclusion our results suggest that magnetic activity consists of both directly driven processes (dayside reconnection) and indirectly driven unloading processes (near-earth nightside reconnection).

ACKNOWLEDGEMENTS

The work described in this paper has been supported by several agencies including the National Science Foundation ATM-87-2194, the Office of Naval Research N00014-85-K-0556, the National Aeronautics and Space Administration NGL-05-007-004, the National Oceanic and Atmospheric Administration 40RANR708458, and the Institute of Geophysics and Planetary Physics of the Los Alamos National Laboratory.

REFERENCES

1. Dungey, J.W., Interplanetary magnetic field and the auroral zones, Phys. Res. Letters, 6, 47-48, 1961.

2. Akasofu, S.-I, Physics of Magnetospheric Substorms, 599 pp., D. Reidel Publishing, Co., Dordrecht, 1977.

3. Akasofu, S.-I, Polar and Magnetospheric Substorms, 280 pp., D. Reidel Pub. Co., Dordrecth, Holland, 1968.

4. McPherron, R.L., Substorm related changes in the geomagnetic tail: The growth phase, Planet. Space Sci., 20, 1521-1539, 1972.

5. Russell, C.T., and R.L. McPherron, The magnetotail and substorms, Space Sci. Rev., 11, 111-122, 1972.

6. Rostoker, G., S.-I. Akasofu, J. Foster, R.A. Greenwald, Y. Kamide, K. Kawasaki, A.T.Y. Lui, R.L. McPherron, and C.T. Russell, Magnetospheric substorms - definition and signatures, J. Geophys. Res., 85(A4), 1663-1668, 1980.

7. McPherron, R.L., C.T. Russell, and M. Aubry, Satellite studies of magnetospheric substorms on August 15, 1978, 9. Phenomenological model for substorms, J. Geophys. Res., 78(16), 3131-3149, 1973.

8. Nishida, A., and N. Nagayama, Synoptic survey for the neutral line in the magnetotail during the substorm expansion phase, J. Geophys. Res., 78(19), 3782-3798, 1973.

9. Hones, E.W., Jr., J.R. Asbridge, S.J. Bame, and S. Singer, Substorm variations of the magnetotail plasma sheet from Xsm = -6Re to Xsm = -60 Re, J. of Geophys. Res., 78(1), 109-132, 1973.

10. Hones, E.W., Jr., Plasma flow in the magnetotail and its implications for substorm theories, in Dynamics of Magnetosphere, edited by S.-I. Akasofu, pp. 542-562, D. Reidel, Hingam, 1979.

11. Akasofu, S.-I, What is a magnetospheric substorm?, in Dynamics of the Mangetosphere, edited by S.-I. Akasofu, pp. 447-460, D. Reidel, Hingam, MA, 1979.

12. Akasofu, S.-I, Energy coupling between the solar wind and the magnetosphere, Space Sci. Rev., 28, 121-190, 1981.

13. Kamide, Y. and J.A. Slavin, (eds.), Solar Wind-Magnetosphere Coupling, 807 pp., Terra Scientific Publishing Company and D. Reidel Publishing Company, Tokyo, Japan and Dordrecht, Holland, 1986.

14. Akasofu, S.-I., Magnetospheric substorms: A newly emerging model, Planet. Space Sci., 29(10), 1069-1078, 1981.

15. Clauer, C.R., R.L. McPherron, C. Searls, M.G. Kivelson, Solar wind control of auroral zone geomagnetic activity, Geophys. Res. Letts., 8(8), 915-918, 1981.

16. Bargatze, L.F., Baker, D.N., and McPherron, R.L., Magnetospheric response to solar wind variations, in Solar Wind-Magnetosphere Coupling, edited by Kamide, Y. and Slavin, J.A., pp. 93-100, Terra Scientific Publishing Co. (Tokyo), 1986.

17. McPherron, R.L., Magnetospheric substorms, Rev. Geophys. Space Phys., 17(4), 657-681, 1979.

18. Rostoker, G., S.-I. Akasofu, W. Baumjohann, Y. Kamide, and R.L. McPherron, The roles of direct input of energy from the solar wind and unloading of stored magnetotail energy in driving magnetospheric substorms, Space Science Reviews, 46, 93-111, 1987.

19. Baumjohann, W., Some recent progress in substorm studies, J. Geomagn. Geolelectr., 38, 633-651, 1986.

20. Otnes, R.K., and L. Enochson, Applied Time Series Analysis, 449 pp., John Wiley and Sons, New York, 1978.

21. Maezawa, K., Dependence of geomagnetic activity on solar wind parameters: a statistical approach, Solar Terrestrial Environ. Res., 2(103), 103-125, 1978.

22. Holzer, R.E. and Slavin, J.A., An evaluation of three predictors of geomagnetic activity , J. of Geophys. Res., 87(A4), 2558-2562, 1982.

23. Bargatze, L.F., D.N. Baker, R.L., McPherron, and E.W. Hones, Magnetospheric impulse response for many levels of geomagnetic activity, J. Geophys. Res., 90(A7), 6387-6394, 1985.

24. Rostoker, G., S.-I. Akasofu, W. Baumjohann, Y. Kamide, and R.L. McPherron, The roles of direct input of energy from the solar wind and unloading of stored magnetotail energy in driving magnetospheric substorms, Space Science Reviews, 46, 93-111, 1987.

25. Clauer, C.R., The technique of linear prediction filters applied to studies of solar wind-magnetosphere coupling, in Solar Wind Magnetosphere Coupling, edited by Y. Kamide and J. Slavin, pp. 39-57, Terra Scientific Publishing Company and D. Reidel Publishing Company, Tokyo and Dordrecht, 1986.

26. Akasofu, S.-I, Some critical issues on magnetospheric substorms, Planet. Space Sci., 34(6), 563-570, 1986.

27. McPherron, R.L. and R.H. Manka, Dynamics of the 1054 UT, March 22, 1979 substorm event: CDAW-6, J. Geophys. Res., 90(A2), 1175-1190, 1985.

28. Baker, D.N., S.-I. Akasofu, W. Baumjohann, J.W. Bieber, D.H. Fairfield, E.W. Hones, Jr., B. Mauk, R.L. McPherron, T.E. Moore, Substorms in the magnetosphere, Chapter 8, Solar Terrestrial Physics: Present and Future, NASA Scientific and Technical Information Branch, NASA Reference Pub. 1120, 1984.

29. Pytte, T., R.L. McPherron, E.W. Hones, Jr., and H.I. West, Jr., Multiple-satellite studies of magnetospheric substorms, III. Distinction between polar substorms and convection-driven negative bays, J. Geophys. Res., 83(A2), 663-679, 1978.