Wavelet–based index of magnetic storm activity [1]

March 22, 2006

A. Jach[2] P. Kokoszka[3] J. Sojka[4] L. Zhu[5]

Utah State University
Logan, UT 84322, USA.

Abstract

A wavelet–based method of computing an index of storm activity is put forward. The new index can be computed automatically using statistical procedures and does not require selecting quiet days and removing the secular component by polynomial fitting. This one-minute index is designed to facilitate the study of the fine structure of geomagnetic storm events and requires only the most recent magnetogram records, e.g. the two months including the storm event of interest. It can thus be computed over a moving window as soon as new magnetogram records become available. Averaged over one hour periods, it is practically indistinguishable from the standard Dst index.

[1]Research supported by NSF grant DMS-0413653.
[2]Department of Mathematics and Statistics. ajach@cc.usu.edu
[3]Department of Mathematics and Statistics. Piotr.Kokoszka@usu.edu
[4]Department of Physics and Center for Atmospheric and Space Science. fasojka@sojka.cass.usu.edu
[5]Department of Physics and Center for Atmospheric and Space Science. zhu@cc.usu.edu

1 Introduction

The currents flowing in the magnetosphere–ionosphere (M–I) form a complicated multi–scale geosystem that contains the temporal scales from seconds to days. Ground–based magnetometers have long been an important tool to observe the M–I current system and a number of indices based on magnetometer data have been introduced to characterize the variations of specific current components. Due to the nature of this current system, the magnetometer data are multi–scale, impulsive, and asynchronous with non–stationary frequency spectra. Based on the assumption that magnetometers in certain latitude bands are most sensitive to specific currents, a traditional way to separate the magnetic effects of different currents is to use magnetometer records from a specific region and combine them into an index characterizing the variation of a specific current. References to review papers on geomagnetic indices are given on p. 409 of Kivelson and Russell (1997) and a concise account of the main indices is presented in Appendix 13B on p. 451 of that monograph.

In this paper we propose an automatic wavelet–based statistical procedure designed to develop an index of storm activity associated with the intensification of the ring current. The Dst index, see Sugiura (1964), has long been used to characterize this variation and is produced from the magnetometer data recorded in the equatorial region. A main difficulty in producing this index lies in subtracting the quiet day variation from a magnetogram recorded at a given location. After this quiet day component and the long term component have been subtracted, the remainder is believed to describe the storm related magnetic activity. Constructing the quiet day component involves somewhat subjectively choosing several days in a month, during which the storm activity is believed to be absent, and averaging them. We show that using a wavelet technique it is possible to produce a one–minute index which, after averaging over one hour periods, is very close to the standard Dst. The unique strength of the technique we propose is that it is fully automatic and, in particular, does not involve selecting quiet days. In this technique, the daily and long–term variation, as well as inconsequential noise, are removed in the wavelet domain by means of statistical filtering. The procedure requires only the most recent stretch of data, and can be used to quickly produce a signature of a storm event.

We now introduce a wavelet–based representation for magnetometer data and formulate a time series model needed to explain the central ideas of the proposed approach. More details are presented in Section 2. For the purpose of this study, the specific magnetogram record to be used is the station's H component, i.e. the magnitude of the horizontal magnetic field. This is the same component as used in the standard Dst calculation, see

Sugiura (1964).

Suppose $\mathbf{X}_s = \{X_{s,t} : t = 0, 1, \dots, N-1\}$ is a magnetogram recorded at station s, the sample size N is the length of the record in minutes. The magnetogram $\mathbf{X}_s$ can be decomposed as

$$\mathbf{X}_s = \sum_{j=1}^{J} \mathbf{D}_{s,j} + \mathbf{S}_{s,J}, \tag{1.1}$$

where

$$\mathbf{D}_{s,j} = \{D_{s,j,0}, \dots, D_{s,j,N-1}\}, \qquad \mathbf{S}_{s,J} = \{S_{s,J,0}, \dots, S_{s,J,N-1}\}.$$

Decomposition (1.1) is known as the Multi Resolution Analysis (MRA). The *details* $D_{s,j,t}$ correspond to the component of the record at time of approximately $2^j t$ minutes and to frequencies in the range of 2^{-j-1} to 2^{-j} cycles per minute. This range corresponds to physical scales between $2^j/60$ and $2^{j+1}/60$ hours. The *smooths* $S_{s,J,t}$ correspond to time of approximately $2^J t$ minutes and to frequencies lower than 2^{-J-1} cycles per minute, or, alternatively, to broadly understood averages over intervals of $2^{J+1}/60$ hours.

As will be seen in the following, the different components of the magnetogram are most prominent at specific levels j of decomposition (1.1). By applying suitable statistical filters to each level j, we are able to isolate and remove the components which do not reflect the storm activity.

To justify our methodology, we assume the following approximate model for the time series $\{X_{s,t}\}$:

$$X_{s,t} = B_s + A_{s,t} + P_{s,t} + L_{s,t} + N_{s,t}, \tag{1.2}$$

with the terms on the right–hand side defined as follows:

B_s : Internal magnetic field of the Earth measured at a low latitude station s (assumed to be time independent).

$A_{s,t}$: The disturbance component attributable to storm activity.

$P_{s,t}$: Periodic component.

$L_{s,t}$: Slowly varying trend (assumed to be a low degree polynomial).

$N_{s,t}$: Noise component.

The components $A_{s,t}$ and $N_{s,t}$ should be thought of as random; $A_{s,t}$ is a "large" random component which becomes pronounced during a storm; $N_{s,t}$ is a "small" random component reflecting all kinds of random disturbances from the measurement error to irregular random disturbances of the M–I system. Small random variations in deterministic components B_s, $P_{s,t}$ and $L_{s,t}$ are thought of as being moved to $N_{s,t}$, so, for example, $P_{s,t}$ is considered a deterministic periodic component to be statistically estimated. By contrast, the component $A_{s,t}$, which we want to isolate, is random because we do not assume anything about the timing or signature of a storm event.

Admittedly, model (1.2) is a simplification needed to develop a usable statistical methodology. For example, $L_{s,t}$ evolves in a complex manner with the solar cycle and treating it as a polynomial is only an approximation. The periodicity of $P_{s,t}$ is only a mathematical assumption; $P_{s,t}$ should not be identified with the simple diurnal variability, a loose association is however helpful to understand the procedure. Without these assumptions, the components on the right–hand side would not be identifiable, as only the $X_{s,t}$ are observable and five components cannot be uniquely identified from their sum. The mathematical assumptions thus play a role of additional equations. A further caviat is that the present analysis does not make any corrections for the ground induction effects. This would require specific knowledge of the ground conducivity properties at each station. Our omission of this correction is in keeping with the existing procedures for calculating the Dst index.

The paper is organized as follows. In Section 2 we present some background the wavelet analysis. Sections 3 and 4 focus, respectively, on the removal of the noise and periodic components. In Section 5, we formulate the procedure for computing our wavelet based index, and in Section 6 we compare it to the standard Dst index. We conclude with final remarks in Section 7.

2 MODWT of the magnetometer data

In this section we describe the Maximum Overlap Discrete Wavelet Transform (MODWT) and explain how it can be applied to a single magnetogram component (scalar data) from a single station. We will also point out the advantages of the MODWT over the Discrete Wavelet Transform (DWT). Finally, we will discuss the choice of the parameters of the wavelet analysis. Our choices are motivated by the task of isolating storm signatures in the presence of a strong periodic component.

Throughout this section, we follow closely the exposition and notation of Percival and Walden (2000). We focus on the aspects of the MODWT which are most relevant to our task, and cannot present all details. An interested reader is referred to Percival and Walden (2000), especially Chapter 5 of that monograph.

The MODWT is a non–orthogonal modification of the DWT which addresses some shortcomings of the latter, such as sample size restriction and sensitivity to the starting point of the series. Like the DWT, the MODWT produces a set of wavelet and scaling coefficients obtained by linear filtering of the signal (this is done by an efficient pyramid algorithm, although with a somewhat higher computational burden than that for the DWT). Unlike the DWT, there are N coefficients at each scale, where N is the number of observations. The MODWT details and smooths are associated with zero phase filters and, unlike details and smooths of the DWT, do not require any shifting to align time events.

Suppose $\mathbf{X}_s = \{X_{s,t} : t = 0, 1, \ldots, N-1\}$ is a scalar component of a magnetogram recorded at station s. For any integer $1 \leq J \leq \log_2(N)$, the MODWT wavelet and scaling coefficients for level j are defined as

$$W_{s,j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{s,t-l \bmod N} \tag{2.1}$$

and

$$V_{s,j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{s,t-l \bmod N} \tag{2.2}$$

where $t = 0, 1, \ldots, N-1$, and $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ are the jth level MODWT wavelet and scaling filters, both of length $L_j = (2^j-1)(L-1)+1$; L denotes the length of the underlying wavelet filter, for example, the Daubechies D(L) or the least asymmetric LA(L) filter, see Section 4.8 of Percival and Walden (2000). In practice, due to the application of the pyramid algorithm, these filters need not be computed.

The transfer function $\widetilde{H}_{j,L}(f)$ for the level j filter can be regarded as an approximation to a transfer function of a perfect band–pass filter with pass–band $[1/2^{j+1}, 1/2^j]$. The squared gain functions, defined as $\widetilde{\mathcal{H}}_{j,L}(f) = |\widetilde{H}_{j,L}(f)|^2$, are shown in Figure 8.1, which will be referred to later. Level j wavelet coefficients are thus associated with the portion of Discrete Fourier Transform (DFT) of $\mathbf{X}_s$ with frequencies in the interval $[1/2^{j+1}, 1/2^j]$.

The MODWT yields decomposition (1.1) with level j detail and smooth sequences $\mathbf{D}_{s,j}$ and $\mathbf{S}_{s,j}$ defined by

$$D_{s,j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} W_{s,j,t+l \bmod N}, \quad t = 0, 1, \ldots, N-1 \tag{2.3}$$

and

$$S_{s,j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} V_{s,j,t+l \bmod N}, \quad t = 0, 1, \ldots, N-1. \tag{2.4}$$

The circular (mod N) filtering in (2.1)–(2.2) and (2.3)–(2.4) generates the so–called boundary MODWT coefficients which do not have a physical interpretation, but are due to the fact that the observations are not part of an infinite sequence with period N. For long records, there are relatively few boundary coefficients, especially for small j. Several methods of reducing the impact of circular filtering have been devised. In this paper we use the *reflection rule*, see Percival and Walden (2000), pp. 140–141.

The application of the MODWT to the magnetometer data requires additional specifications, such as the choice of a wavelet filter and the maximum level J. With the goals of the analysis in mind, we discuss these issues below.

Wavelet filter: In our study, we use the LA(8) filter which has been extensively applied in several recent quantitative analyzes of geophysical data, see Whitcher *et al.* (2002), Overland *et al.* (2004), Cornish *et al.* (2004), among others. Using Daubechies filter D(8) produces the same results. LA filters are parametrized by even widths L. For a specific analysis, a width must be selected which would provide sufficient concentration in the octave pass–band and reduce the overlapping of frequencies from outside of it. This can be achieved by selecting a relatively long filter, but choosing L too large will result in many boundary coefficients, especially at higher levels j. Choosing $L = 8$ has been found to offer a reasonable balance between these two competing requirements. This choice also guarantees that the MODWT coefficients are uninfluenced by a polynomial of order $K \leq L/2 - 1 = 3$. Thus, a slowly varying trend (secular variation) in the magnetic field, which is typically modeled by a low order polynomial, see Kivelson and Russell (1997), p. 457, is practically not reflected in the details of the MODWT of magnetometer data. On the other hand, for one–year–long magnetometer data, this trend can be captured by a smooth $\mathbf{S}_{s,J}$ on the highest available level, and thus an appropriate adjustment of $\mathbf{S}_{s,J}$ will automatically remove it.

Number of levels J: The maximum level J in (1.1) is typically chosen to be smaller than its largest admissible value $\lfloor \log_2(N) \rfloor$. However, for one–year–long magnetometer data, we need to account for the secular variation. Decomposing the signal up to level $J = \lfloor \log_2(N) \rfloor = 19$ and replacing $\mathbf{S}_{19}$ by its average (this average is equal to the average of the magnetogram) allow to eliminate the annual trend. For shorter records, for example, two–month–long, the smooth corresponding to the highest available level is affected by the storm events and cannot be regarded as a good approximation to the annual trend. We assume that the long–term component is not visible in the records extending over the period of two months. This assumption is validated by comparing our index computed from two months of data to the standard Dst index. For such records we propose to set $J = 10$, because J should be large enough to allow the removal of the noise and periodic components. Since the latter is visible at higher scales, J will be dictated by the frequencies contributing to the periodic (Sq) variation. As we will see in Section 4, the lowest frequency corresponding to the periodic portion of the magnetogram is the daily frequency, 1/1440, that falls in $[2^{-j-1}, 2^{-j}]$, $j = 10$. Taking $J = 10$ ensures that this component can be effectively analyzed.

3 Noise removal via wavelet thresholding

Wavelet thresholding is a nonparametric method of estimating a signal in the presence of additive noise, see Chapter 10 of Percival and Walden (2000) for a review, and has been developed both theoretically and practically over the last decade. The underlying assumption of this methodology is that the noise is "small" compared to the signal. We use thresholding to remove the component $N_{s,t}$ in decomposition (1.2). To justify this approach, note that by (1.2) and (2.1)

$$W_{s,j,t} = W^{(A)}_{s,j,t} + W^{(P)}_{s,j,t} + W^{(N)}_{s,j,t}, \tag{3.1}$$

where the $W^{(A)}_{s,j,t}$ are the MODWT coefficients of the storm component $A_{s,t}$, $W^{(P)}_{s,j,t}$ and $W^{(N)}_{s,j,t}$ being defined analogously. The wavelet coefficients of the components B_s and $L_{s,t}$ vanish because these components are assumed to be, respectively, a constant and a polynomial of degree not exceeding three. At low levels $j \le J_0$, where the value of J_0 is specified at the end of this section, the periodic coefficients $W^{(P)}_{s,j,t}$ are negligible, see Section 4, so in fact, we may assume that

$$W_{s,j,t} = W^{(A)}_{s,j,t} + W^{(N)}_{s,j,t}. \tag{3.2}$$

Moreover, for $j \le J_0$, the $W_{s,j,t}$ are mostly "small noise" with large coefficients during storm events. Thus, the goal of thresholding is to remove the $W^{(N)}_{s,j,t}$ which constitute the vast majority of the coefficients at low scales. By (3.2), each coefficient $W_{s,j,t}$ must be reduced by $W^{(N)}_{s,j,t}$. This can clearly be done only in some average statistical sense because the $W^{(N)}_{s,j,t}$ are not observable. We assume that the absolute values of the $W^{(N)}_{s,j,t}$ approximately do not exceed a level δ_j. The observed coefficients $W_{s,j,t}$ are thus replaced by zero if their absolute value is smaller than δ_j and their length is reduced by δ_j if their absolute value is greater than δ_j. The resulting coefficients are denoted by $\overline{W}_{s,j,t}$ and are thus given by the formula

$$\overline{W}_{s,j,t} = \mathrm{sign}\{W_{s,j,t}\}(|W_{s,j,t}| - \delta_j)_+, \quad t = 0, 1, \ldots, N-1 \quad (j \le J_0). \tag{3.3}$$

The $\overline{W}_{s,j,t}$ are statistical estimates of the $W^{(A)}_{s,j,t}$ in the absence of the $W^{(P)}_{s,j,t}$.

Formula (3.3) defines the so-called *soft level–dependent thresholding*. In *hard thresholding*, coefficients $W_{s,j,t}$ whose absolute value exceeds δ_j are not modified. Soft thresholding seems more appropriate in our context because formula (3.2) suggests that $W_{s,j,t}^{(N)}$ should be removed from all coefficients. Moreover, our exploratory analysis has shown that hard thresholding introduces occasional long spikes to the proposed index of storm activity which have no reasonable physical interpretation. In level–independent thresholding $\delta_j = \delta$ is the same for all levels and is typically determined by the statistical behavior of the coefficients at the finest level $j = 1$. This approach is suitable for the task of removing noise from a signal, but is not appropriate in our context because for magnetometer data the magnitude of the $W_{s,j,t}^{(N)}$ visibly changes with level j.

A natural and convenient way to select δ_j is to define it as the pth quantile of the distribution of the absolute values of the $W_{s,j,t}$. Thus, we define $\delta_j = \delta_{j,s}(p)$ by the formula

$$P(|W_{s,j,t}| \leq \delta_{j,s}(p)) = p. \tag{3.4}$$

The probability P in (3.4) is the empirical probability, i.e. the proportion of the $|W_{s,j,t}|, t = 0, 1, \ldots, N-1$ which exceed $\delta_{j,s}(p)$. Our exploratory analysis has shown that p should be taken relatively large (close to 1). For example, if $p = 0.98$, all coefficients, except the largest (in absolute terms) 2% of are set to zero. Note that this statistical approach guarantees that the thresholding can be automatically applied to every station no matter what the typical magnitude for the magnetic field is.

Recall that J_0 is the largest scale on which the effect of the periodic component is negligible. This definition is rather vague and implies merely that scales with $j \leq J_0$ are treated differently than scales with $j > J_0$. The periodic component is associated with the Sq variation which has a pronounced daily frequency (1/1440 cycles per minute), half–daily frequency (2/1440) and whose spectrum also contains higher frequencies, even though these are much less pronounced. Note that the half–daily frequency corresponds to the scale of $1440/2 = 720$ minutes which is contained in the scale range [512, 1024] corresponding to $j = 9$. Because of the presence of frequencies higher than 2/1440 in the spectrum of the Sq variation, the periodic component is also visible at level $j = 8$. We therefore choose $J_0 = 7$ based both on theoretical grounds discussed above and on an exploratory analysis of the wavelet decompositions of the magnetometer data.

4 Removal of the periodic component

This section proposes a statistical method of removing the periodic component $P_{s,t}$ in decomposition (1.2). We loosely identify the periodic component with the daily variation caused by the rotation of the Earth. As the position of the station relative to the M–I current system changes, various currents, including the wind currents on the dayside of the ionosphere, the magnetopause and tail currents, leave approximately periodic signatures in the magnetometer data.

The method will be applied to selected details $\mathbf{D}_{s,j}$ because, unlike the MODWT coefficients $W_{s,j,t}$, times t in the details $D_{s,j,t}$, are aligned with times t in the observations $X_{s,t}$ ($t = 0, 1, \ldots, N-1$), see Chapter 5 of Percival and Walden (2000). Such a time alignment is not necessary in the removal of the noise which does not have a regular pattern over time. Instead, the wavelet domain equation (3.2) was exploited.

Exploratory analysis of the details from four Dst stations, showed that the periodicity is clearly visible in the $\mathbf{D}_{s,j}$ for $j = 8, 9, 10$ and not visible by eye at other scales. This finding agrees with the known properties of the daily variation and the spectral properties of the LA(8) wavelet filter. Focusing first on the daily variation, see Section 4.4 of Parkinson (1983), recall that its spectrum is dominated by peaks at frequencies of 1, 2 and possibly 3 cycles per day. These frequencies correspond, respectively, to wavelet levels $j = 10, 9, 8$, as already discussed in Section 3. Figure 8.1 shows that the the squared gain functions $\widetilde{\mathcal{H}}_{j,L}(f),\ j = 8, 9, 10$, of the LA(8) filter together completely cover the interval [0.0005, 0.0040] which contains practically all frequencies present in the spectrum of the daily component, including the daily frequency of $1/1440 \approx 0.0007$ and the half–daily frequency of about 0.0014 (all frequencies are in cycles per minute).

As an example, Figure 8.2 shows the details $\mathbf{D}_j$ for levels 8 through 10 for station Kakioka for March–April of 2001. As we can see, the largest contribution to the periodic component comes from the "slow–oscillating" details capturing the daily frequency (bottom panel) and the smallest from the "fast–oscillating" $\mathbf{D}_8$.

Common filtering techniques involve differencing or averaging, but neither of them is advisable in the context of magnetometer data. Differencing the details will destroy the structure of the MRA, whilst average–based filtering will be greatly influenced by the extreme "stormy" observations superimposed on the periodic fluctuations. We therefore propose to use a robust median–based filter. It is analogous to the usual method of removing the periodic component, see e.g. Section 1.5 of Brockwell and Davis (2002), but uses the median instead of the average. Unlike the average, the median is not sensitive to unusually large or small observations. Moreover, the wavelet coefficients of magnetometer data have *heavy tails*, see Kokoszka *et al.* (2006). Discussing this property here would distract us from the focus of this paper, so we merely note that for heavy–tailed observations, median–based procedures generally work better than procedures based on averaging. The ultimate justification for using the median rather than the average comes from the comparison of the performance of both methods in our context.

To describe our method, suppose $Y_0, Y_1, \ldots, Y_{N-1}$ is a time series without a trend. In our context, it should be thought of as the time series of details at a given level j from which we wish to remove the periodic component. Recall from Section 2 that the MODWT coefficients and the details do not contain a trend because the slowly varying trend in the data is modeled as a polynomial of order not exceeding 3, and we use wavelet filter of length 8.

Given a series $Y_0, Y_1, \ldots, Y_{N-1}$ from which a periodic component with period d is to be removed, we follow these steps:

1: Construct $R = \lfloor N/d \rfloor$ sequences

$$Y_t, Y_{t+1d}, Y_{t+2d}, \ldots, Y_{t+(R-1)d}, \quad t = 0, 1, \ldots, d-1.$$

(R is the number of consecutive, non–overlapping sequences of length d that "fit into" the sequence of length N.)

2: For each such sequence compute the median m_t, $t = 0, 1, \ldots, d-1$.

3: For $t > d-1$, write $t = Kd + u$, for some integers $K \geq 1$ and $0 \leq u \leq d-1$, and set $m_t = m_u$. (In this step the sequence m_t, $t = 0, 1, \ldots, d-1$, is extended periodically to a sequence of length N.)

4: Compute the mean $\mu = N^{-1} \sum_{t=0}^{N-1} m_t$ and return sequence $\{Y_t - (m_t - \mu)\}$. (This step ensures that the estimated periodic component $m_t - \mu$, $t = 0, 1, \ldots, N-1$, has mean zero.)

In the estimation of the periodic component we include the boundary coefficients, but since the above procedure is robust to atypical observations, it is not affected by these coefficients.

In Section 5, to remove the daily variation, we apply the above procedure to the details $\mathbf{D}_{s,8}, \mathbf{D}_{s,9}, \mathbf{D}_{s,10}$ with periods, respectively, $d = 480, 770, 1440$.

5 Index algorithm

We now describe the full algorithm for obtaining an index of storm activity which incorporates the statistical procedures described in the previous sections. In light of the motivation and objectives outlined in Section 1, our primary focus is on an algorithm which uses two months worth of data, but we also show how to construct an index over a period of one year.

Recall that for $s = 1, 2, \ldots, S$, $\mathbf{X}_s = \{X_{s,t} : t = 0, 1, \ldots, N-1\}$ is a magnetogram recorded at station s.

1) If N corresponds to two months set $J = 10$, if N corresponds to one year, set $J = 19$. In both cases, set $J_0 = 7$. Compute the MODWT coefficients $W_{s,j,t}$, $j = 1, 2, \ldots, J$, and $V_{s,J,t}$;

2) Threshold the wavelet coefficients $W_{s,j,t}$, $j = 1, 2, \ldots J_0 = 7$, to obtain $\overline{W}_{s,j,t}$ according to Section 3, and using equation (2.3) compute the details $\overline{D}_{s,j,t}$ by replacing $W_{s,j,t}$ with $\overline{W}_{s,j,t}$;

3) Compute the details $\mathbf{D}_{s,j}$, $j = 8, \ldots, J$ and the smooth $\mathbf{S}_{s,J}$. To the details $\mathbf{D}_{s,8}$, $\mathbf{D}_{s,9}$, $\mathbf{D}_{s,10}$, apply median–based filter of Section 4 with periods $d = 480, 770, 1440$, respectively, and denote the resulting details without the periodic components by $\underline{\mathbf{D}}_{s,j}$, $j = 8, 9, 10$;

4a) If N corresponds to two months, compute the filtered magnetogram from which the periodic and noise components have been removed:

$$\begin{aligned} \widehat{A}_{s,t} + \widehat{B}_s &= \sum_{j=1}^{7} \overline{D}_{s,j,t} + \sum_{j=8}^{10} \underline{D}_{s,j,t} + S_{s,10,t} \\ &= X_{s,t} - \widehat{P}_{s,t} - \widehat{L}_{s,t} - \widehat{N}_{s,t}; \end{aligned}$$

4b) If N corresponds to one year ($J = 19$), replace the smooth $\mathbf{S}_{s,J}$, which reflects the long–term (secular) component, by its average $N^{-1}\sum_{t=0}^{N-1} S_{s,J,t} = N^{-1}\sum_{t=0}^{N-1} X_{s,t}$ and compute the filtered magnetogram from which the trend, periodic and noise components have been removed:

$$\begin{aligned} \widehat{A}_{s,t} + \widehat{B}_s &= \sum_{j=1}^{7} \overline{D}_{s,j,t} + \sum_{j=8}^{10} \underline{D}_{s,j,t} + \sum_{j=11}^{19} D_{s,j,t} + \frac{1}{N}\sum_{t=0}^{N-1} X_{s,t} \\ &= X_{s,t} - \widehat{P}_{s,t} - \widehat{L}_{s,t} - \widehat{N}_{s,t}; \end{aligned}$$

5) Compute the average of the filtered magnetogram

$$\widehat{M}_s = \frac{1}{N}\sum_{t=0}^{N-1}\left(\widehat{A}_{s,t} + \widehat{B}_s\right) = \left\{\frac{1}{N}\sum_{t=0}^{N-1}\widehat{A}_{s,t}\right\} + \widehat{B}_s$$

and center the filtered data to have mean zero:

$$\widehat{A}^c_{s,t} = \left(\widehat{A}_{s,t} + \widehat{B}_s\right) - \widehat{M}_s = \widehat{A}_{s,t} - \frac{1}{N}\sum_{t=0}^{N-1}\widehat{A}_{s,t}.$$

(This step eliminates, in particular, the constant field B_s.);

6) Adjust for the magnetic latitude:

$$\widehat{A}^{ca}_{s,t} = \widehat{A}^{c}_{s,t} / \cos(\theta_s)$$

(θ_s is the magnetic latitude of station s.);

7) Average over all stations to obtain a global index:

$$I_t = \frac{1}{S} \sum_{s=1}^{S} \widehat{A}^{ca}_{s,t}, \quad t = 0, 1, \ldots, N-1.$$

To facilitate the discussion below, we refer to the index I_t as the *Wavelet Index of Storm Activity* or WISA.

Comments.

1) The last two steps of the algorithm are the same as for the standard Dst index, see Kivelson and Russell (1997), p. 457.

2) In order to obtain the WISA over a given period of time, only the data from that period are used. For example, to construct an index over a period of two months, no observations outside that period are needed. This feature offers the flexibility of constructing the index only from new data or of using a moving window of suitably chosen length.

3) The WISA reflects only the dynamic range of storm activity over a given period of time and does not provide any reference level relative to the past. If the WISA is constructed for a period of one year, its mean over that year is zero, but the means over any subperiods of that year need not be zero. If the index is constructed over two months, its mean over these two months is zero.

4) The WISA can be constructed practically over any period of time, but to obtain values similar to the standard Dst, at least two months worth of data are needed. The periods of two months and one year were chosen for illustration only. Note that if I_t^{1y} is an index constructed from 1 year worth of data and I_t^{2m} is an index constructed only from e.g. March–April data, then over the months of March and April the indices I_t^{1y} and I_t^{2m} will differ slightly.

5) Once the length of the period of time over which the WISA is to be computed is set, the procedure is fully automatic and requires merely one minute measurements of the horizontal intensity as input.

6) Even quality one–minute data from INTERMAGNET CDs which we used contain some missing values. We used linear interpolation to estimate these missing values. Linear interpolation gives slightly different values of the WISA, the difference is however not perceptible by eye on a graph showing both indices over a period of two months.

The data we worked with had only occasional stretches of a few missing values. The presence of long stretches of missing values may be a serious problem. This issue and the use of data from non-Dst stations will be explored in future work.

7) The WISA is a one minute index, so in order to compare it to the standard Dst, some averaging is needed. This issue is taken up in greater detail in Section 6.

8) The algorithm described in this section can be modified in several ways while preserving the general idea. For example, different thresholding levels could be used and the removal of the periodic component could be extended to level $j = 7$. At levels $j = 8, 9$, a component with period of 1 day could be removed rather than components with periods of, respectively, 8 and 12 hours. All these modifications would produce only a slightly different index. The algorithm we settled on seems most logical.

9) Naive removal of the details containing the noise and periodic component rather than statistical filtering produces highly oversmoothed storm events because the information about the storms is contained in all levels and must be extracted with care.

10) For one year data, the long–term trend at a given station is estimated by the smooth $\mathbf{S}_{s,J}$, see step 4b) of the algorithm. Since our goal is to produce an index of storm activity, $\mathbf{S}_{s,J}$ was replaced by its average.

11) It is well–established that the pattern of the daily variation changes with season and undergoes other long term changes. We therefore see the primary application of the proposed method in computing the index over a window of two months. Assuming daily periodicity of $P_{s,t}$ over the period of one year is more questionable, but the WISA computed over one year is still remarkably close to the standard Dst. A detailed comparison is presented in Section 6.

6 Comparison to Dst

The objective of this section is to compare the new WISA index to the Dst index. There are several sources of differences:

1) WISA and Dst are calculated in fundamentally different way.

2) Dst is an hourly index and WISA is a one–minute index. Averaging of high frequency magnetometer records could possibly be done differently to produce hourly Dst than to obtain average values of WISA over one hour periods. To obtain hourly values of WISA we used one minute averages from INTERMAGNET's CDs (which are themselves averages of raw data recorded with 5 s frequency) and then averaged the sixty values of one–minute WISA centered at the time for which Dst is reported. Details of how raw magnetometer measurements are used to produce the Dst are not readily available.

3) Even the best quality records contain missing values. We interpolated the missing via linear interpolation. A different interpolation technique would produce slightly different WISA values. We do not know how missing values are handled to produce Dst.

Ideally, we would like to compare WISA to Dst with respect to defferences arising from the different methodology, i.e. from source 1). This source cannot however be separated from sources 2) and 3) and keeping this in mind we will argue that the contribution of source 1) is negligible, i.e. over a period of one year our technique can automatically reproduce Dst up to the acuracy determined by the preprocessing of the raw data.

In the comparison below we use the final one–hour Dst index and WISA computed from one–minute values of the H–component for Hermanus, Honolulu, Kakioka and San Juan (Dst stations). We consider the time period January–December, 2001, which contains a few very strong storms with the dynamic range approaching 500 nT. By construction, the average of the WISA is zero, cf. comment 3) in Section 5. The Dst is computed over longer periods of time and its average is not zero for any subperiod (but is close to zero). To make the comparison possible, we substracted the average value of the Dst for the periods over which we compare it to WISA. Specifically, when using the Dst over the whole year 2001, we added 18 nT to it to make its average for 2001 zero. For the subperiods of two months, different constants had to be added. For example, for the period of March–April, which contains a very strong storm, we added 31 nT to the Dst to make its average over this two month period zero.

Figure 8.3 shows the Dst index (bottom panel) and the hourly averages of WISA (top panel). The two indices look very similar. In Figure 8.4, the differences between Dst and WISA are plotted and are seen to be generally smaller than 10 nT. These differences tend to be slightly larger during disturbed period (Dst < -30 nT) than during quiet periods (Dst ≥ -30 nT). This point is illustrated in Figure 8.5. The physical interpretation of the larger spread of differences for disturbed conditions requires additional study with larger data sets, but it can be conjectured that our approximation of the diurnal component is more accurate during quiet periods; during disturbed periods this component appears less stable. The differences for both the disturbed and calm periods have an error–type distribution, indicating that they may be attributable to chance errors rather than a systematic bias.

As explained above, the averaging of WISA to obtain an hourly index introduces some arbitrary variability. To assess this variability, we computed the differences

$$\Delta I_t = I_t - I_{t-1}, \quad t = 1, 2, \ldots, \lfloor N/60 \rfloor - 1.$$

The sample standard deviation of these differences, denoted $s_{\Delta I}$ is a measure of variability due to averaging over one hour intervals. Since the differences are approximately normally distributed, the $100(1-\alpha)\%$ confidence interval for the hourly WISA can be defined as

$$(I_t - z_{1-\alpha/2} s_{\Delta I}, I_t + z_{1-\alpha/2} s_{\Delta I}), \tag{6.1}$$

where z_β denotes βth quantile of the standard normal distribution. For the whole of 2001, $s_{\Delta I} = 5.20$ nT, and 97.72% of the Dst values are within the 95% confidence limits. Figure 8.6 shows the 95% confidence intervals together with the two indices for the first week of January. The graphs have a similar appearance for the remaining weeks of 2001.

We also compared hourly WISA computed over 2 month periods in 2001 to the Dst. Graphs for the most stormy period, March–April, are presented in Figures 8.7– 8.10. For this period, all Dst values are within the 95% confidence limits. The increasing spread of differences visible in Figure 8.8 is due to a very strong storm in the middle of that period. As for the WISA computed from one year worth of data, the variability due to the averaging of the one-minute WISA is statistically greater than the differences between the hourly WISA and the Dst. Perhaps more importantly, the difference between the WISA and the Dst, which is typically about 5 nT, is negligible compared to the dynamic range of a storm which is several hundred nT.

7 Conclusions

We propose an automatic procedure for calculating an index of geomagnetic storm activity which uses scale time–dependent decomposition provided by the Maximum Overlap Discrete Wavelet Transform and statistical filtering techniques. We use the wavelet thresholding to remove the noise and a median–based filter to remove the periodic component. The procedure produces a one–minute index and requires only a recent short stretch of data. It replaces the Dst determination of quiet days involving subjective human selection by robust statistical estimation of the diurnal component. When the H–components of the four Dst stations are provided as input, the hourly averages of the new index are statistically indistinguishable from the Dst values.

The procedure can take magnetometer data from different stations as input thus allowing the study of the effect of the choice of the stations on the index of storm activity. Such a study is however beyond the scope of this work. The proposed procedure is flexible and amenable to various modifications and can potentially serve as a useful tool in the study of the magnetic stroms. Its full automation has a substantial potential for operational purposes.

Research is ongoing to investigate both the short term storm related wavelet decompositions and the long term seasonal and solar cycle dependencies of the other components in equation (1.2).

Acknowledgement. This research was supported by NSF Grant DMS0413653 and NASA Grant NNG05GJ48G. The magnetometer data was provided by INTERMAGNET.

References

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*, Second edn. Springer-Verlag, New York.

Cornish, C. R., Bretherton, C. S. and Percival, D. B. (2004). Wavelet analysis of marine atmospheric boundary layer turbulence during epic. Technical Report. University of Washington. available at `http://faculty.washington.edu/dbp/`.

Kivelson, M. G. and Russell, C. T. (1997) (eds). *Introduction to Space Physics.* Cambridge University Press.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2006). Probability tails of wavelet coefficients of magnetometer records. *Journal of Geophysical Research*; under review.

Overland, J. E., Spillane, M. C., Percival, D. B., Wang, M. and Mofjeld, H. O. (2004). Seasonal and regional variation of pan-arctic surface air temperature over the instrumental record. *Journal of Climate*, **17,** 3263–82.

Parkinson, W. D. (1983). *Introduction to Geomagnetism.* Scottish Academic Press.

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge University Press, Cambridge.

Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY. *Ann. Int. Geophysical Year*, **35,** number 9; Pergamon Press, Oxford.

Whitcher, B., Byers, S., Guttorp, P. and Percival, D. (2002). Testing for homogeneity of variance in time series: long memory, wavelets and the Nile River. *Water Resources Research*, **38,**; DOI 10.1029/2001WR000509.

8 Graphs

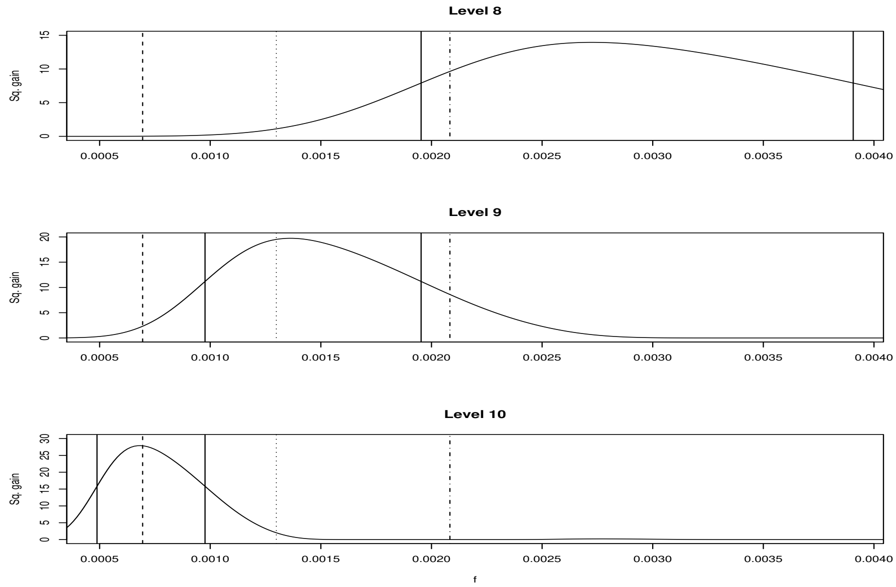

FIGURE 8.1 Squared gain functions $\widetilde{\mathcal{H}}_{j,L}(f)$ for LA(8) filter, levels $j = 8, 9, 10$. Nominal pass–bands are indicated by solid vertical lines. Periodic frequencies 1/1440, 2/1440, 3/1440 are marked by dashed, dotted and dashed–dotted lines, respectively.

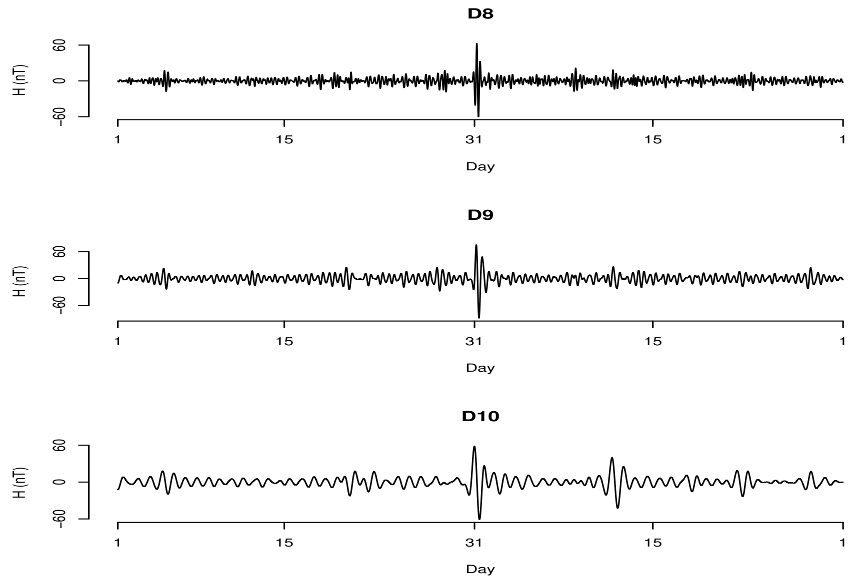

FIGURE 8.2 Details from levels $j = 8, 9, 10$ for station Kakioka, March–April, 2001.

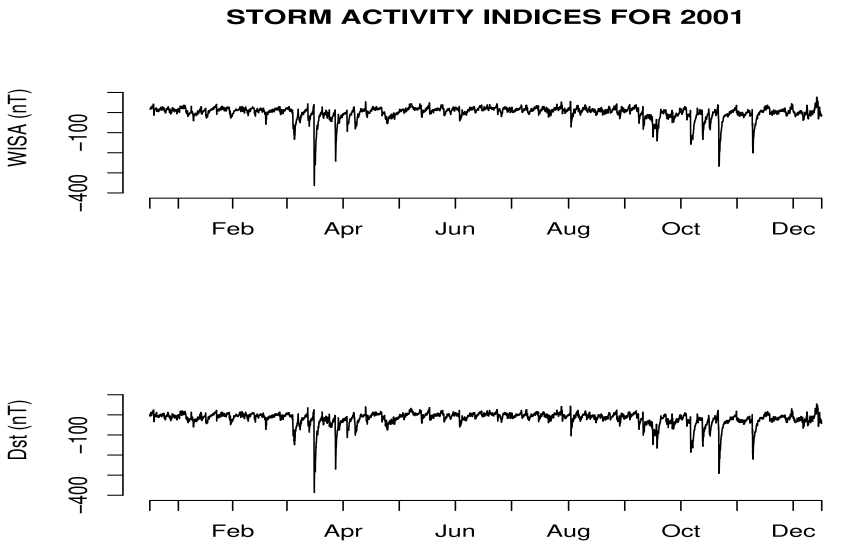

FIGURE 8.3 Hourly values of WISA and Dst, January–December, 2001.

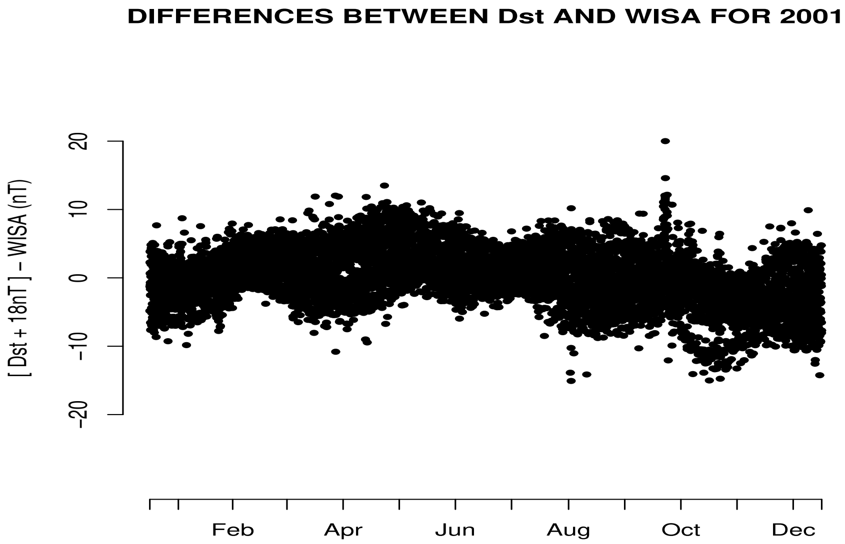

FIGURE 8.4 Differences between Dst (+18nT) and WISA, January–December, 2001.

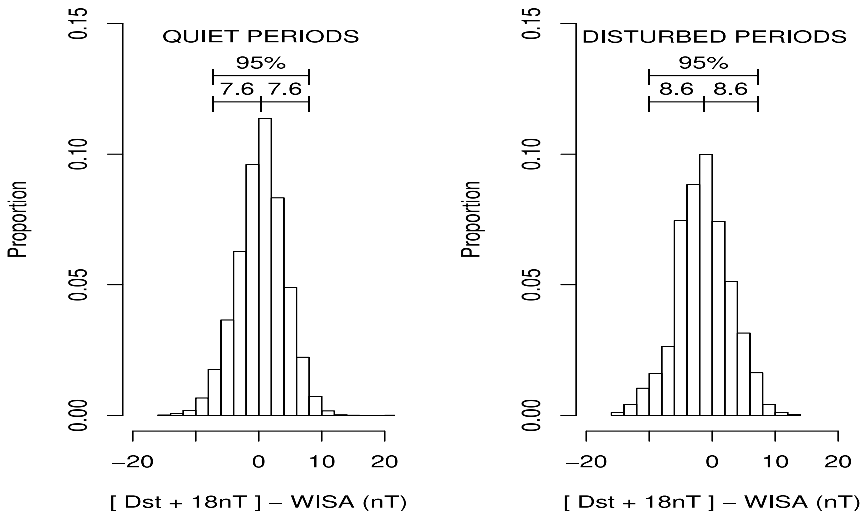

FIGURE 8.5 Histograms of differences between Dst (+18nT) and WISA for "quiet" and "disturbed" periods, January–December, 2001.

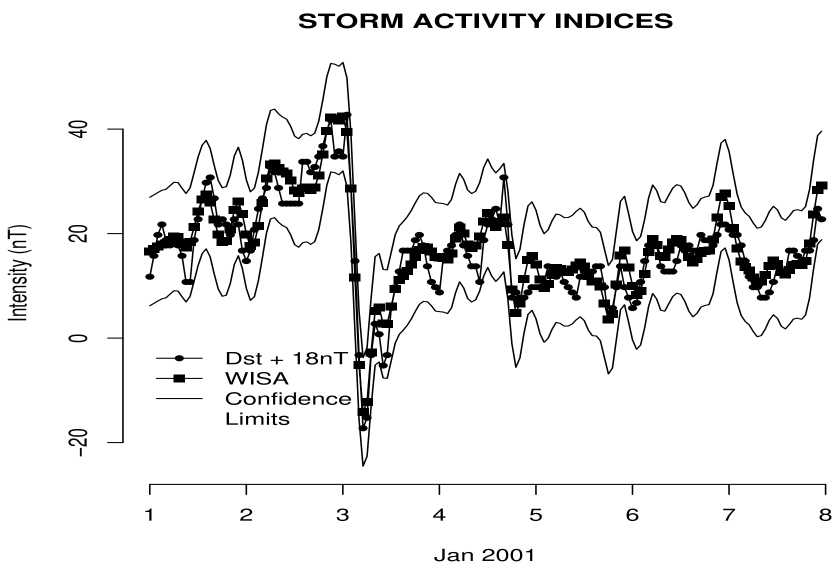

FIGURE 8.6 Dst, WISA and 95% confidence intervals for hourly WISA for the first week of 2001.

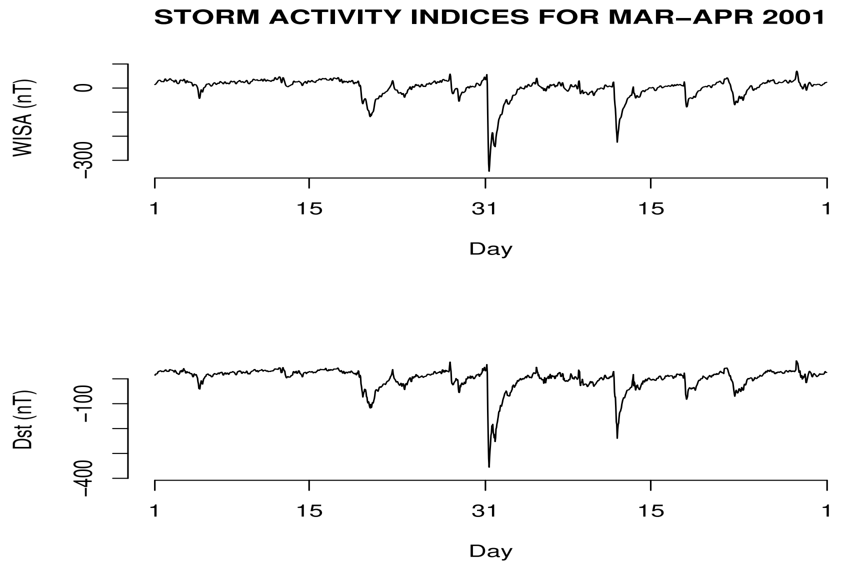

FIGURE 8.7 Hourly values of WISA and Dst, March–April, 2001.

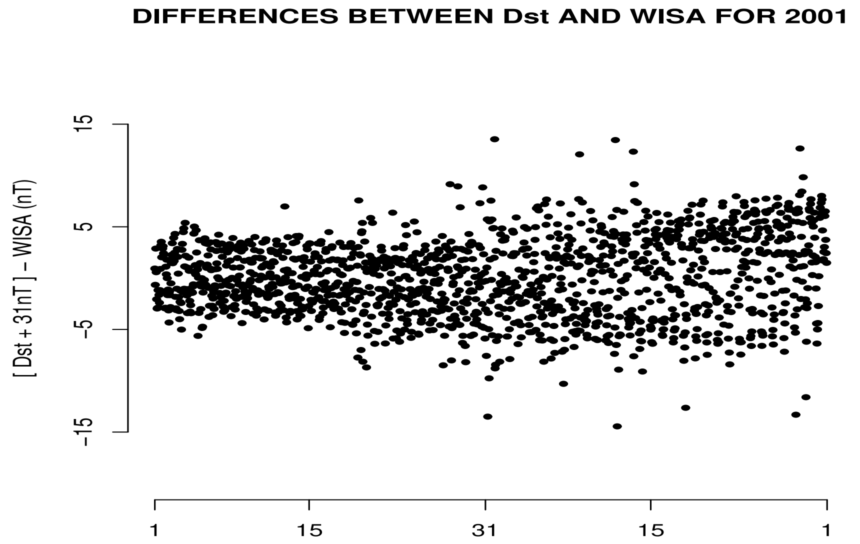

FIGURE 8.8 Differences between Dst (+31nT) and WISA, March–April, 2001.

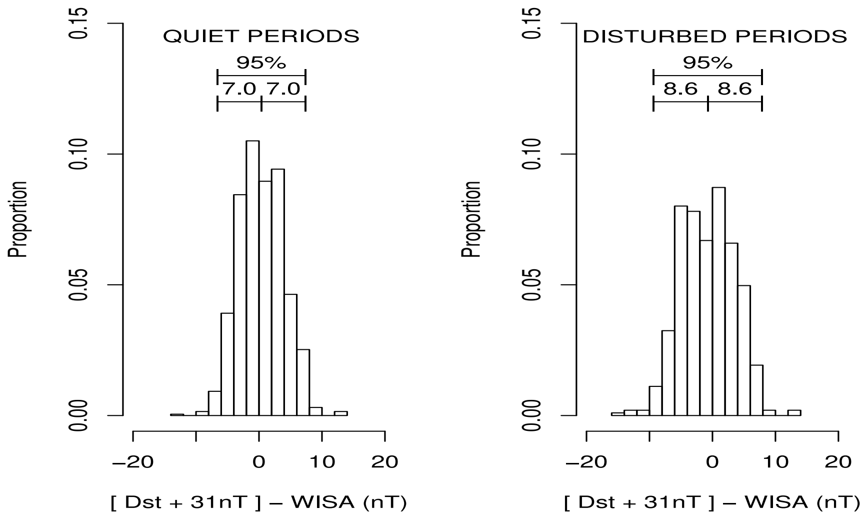

FIGURE 8.9 Histograms of differences between Dst (+31nT) and WISA for "quiet" and "disturbed" periods, March–April, 2001.

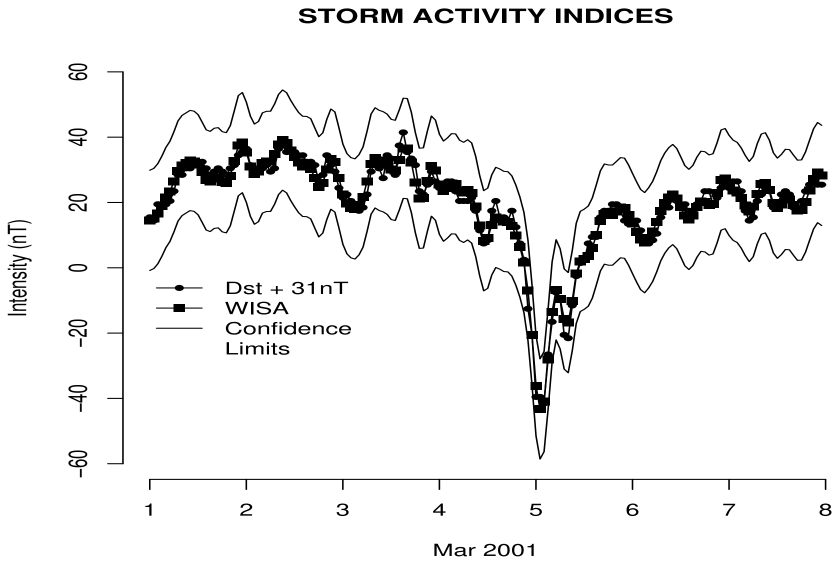

FIGURE 8.10 Dst, WISA and 95% confidence intervals for hourly WISA for the first week of March.