



HAL
open science

Prediction of magnetic storm events using the Dst index

V. V. Anh, Z. G. Yu, J. A. Wanliss, S. M. Watson

► **To cite this version:**

V. V. Anh, Z. G. Yu, J. A. Wanliss, S. M. Watson. Prediction of magnetic storm events using the Dst index. *Nonlinear Processes in Geophysics*, 2005, 12 (6), pp.799-806. hal-00302592

HAL Id: hal-00302592

<https://hal.science/hal-00302592>

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prediction of magnetic storm events using the D_{st} index

V. V. Anh^{1,2}, Z. G. Yu¹, J. A. Wanliss³, and S. M. Watson²

¹Program in Statistics and Operations Research, Queensland University of Technology, GPO Box 2434, Brisbane, Q4001, Australia

²Florida Space Institute, University of Central Florida, Orlando, Florida 32816-2370, USA

³Embry-Riddle Aeronautical University, 600 S. Clyde Morris Blvd., Daytona Beach, Florida, 32114, USA

Received: 15 April 2005 – Revised: 1 August 2005 – Accepted: 8 August 2005 – Published: 11 August 2005

Part of Special Issue “Nonlinear and multiscale phenomena in space plasmas”

Abstract. This paper provides a method to predict magnetic storm events based on the time series of the D_{st} index over the period 1981–2002. This method is based on the multiple scaling of the measure representation of the D_{st} time series. This measure is modeled as a recurrent iterated function system, which leads to a method to predict storm patterns included in its attractor. Numerical results are provided to evaluate the performance of the method in outside-sample forecasts.

1 Introduction

A measure of the strength of a magnetic storm is the D_{st} index, which is supposed to reflect variations in the intensity of the symmetric part of the ring current at altitudes ranging from about 3–8 earth radii (Greenspan and Hamilton, 2000). The situation is more involved because a substantial portion of D_{st} may be a result of electromagnetic induction effects or other magnetospheric currents and may be asymmetric (Burton et al., 1975; Langel and Estes, 1985; Turner et al., 2001). The D_{st} index is calculated at hourly intervals from the horizontal magnetic field component at four observatories, namely, Hermanus (33.3° S 80.3°) in magnetic dipole latitude and longitude), Kakioka (26.0° N, 206.0°), Honolulu (21.0° N, 266.4°), and San Juan (29.9° N, 3.2°). These four observatories were chosen because they are close to the magnetic equator and thus are not strongly influenced by auroral current systems. They are therefore more likely to be an accurate gauge of the strength of the ring current perturbation.

Recent research (Wanliss, 2004, 2005) has found that D_{st} exhibits a power-law spectrum with the Hurst index varying over different stretches of the time series. This behavior indicates that D_{st} is a multifractional process. An important example of such processes is multifractional Brownian motion (Ayache and Lévy Véhel, 2000; Benassi et al., 2000).

Correspondence to: V. V. Anh
(v.anh@qut.edu.au)

Heavy-tailed Lévy-type behaviour, particularly that of stable distributions, has also been observed in the interplanetary magnetic field and the magnetosphere (Burlaga, 1991, 2001; Burlaga et al., 2003; Kabin and Papitashvili, 1998; Lui et al., 2000, 2003). It is known that, apart from Brownian motion and Poisson processes, all other Lévy processes display a form of multifractal scaling (Jaffard, 1999). Fractal and multifractal approaches have been quite successful in extracting salient features of physical processes responsible for the near-Earth magnetospheric phenomena (Lui, 2002). But it should be noted that these approaches, while characterizing the fractal/multifractal behavior of the process under study, do not yield a direct method for its prediction.

In this paper, we look at the multiple scaling of D_{st} from a different angle, namely from an iterated function system of its measure representation. This measure is a histogram-type probability measure of the patterns of the events extracted from the D_{st} time series (to be defined below). The measure has the characteristic of a multifractal measure, and will be modeled via a recurrent iterated function system (RIFS), which is considered as a dynamical system. The attractor of this dynamical system is in fact the support of the probability measure. A great advantage of this approach is that the probabilities of patterns of future events can be determined from the RIFS. Hence the method provides a mechanism for prediction of future events. It should be emphasized that this prediction is based on the multiple scaling inherent in the D_{st} , rather than its autocorrelation structure as in usual methods for prediction of stationary processes.

The next section will outline the concepts of measure representation, multifractal measures and RIFS. Section 3 develops RIFS models for the measure representations of the D_{st} time series at hourly and daily scales. Section 4 will look at the prediction of storm events via these models and evaluate the performance of this method through a number of accuracy indicators. Some concluding comments on the approach will be provided in Sect. 5.

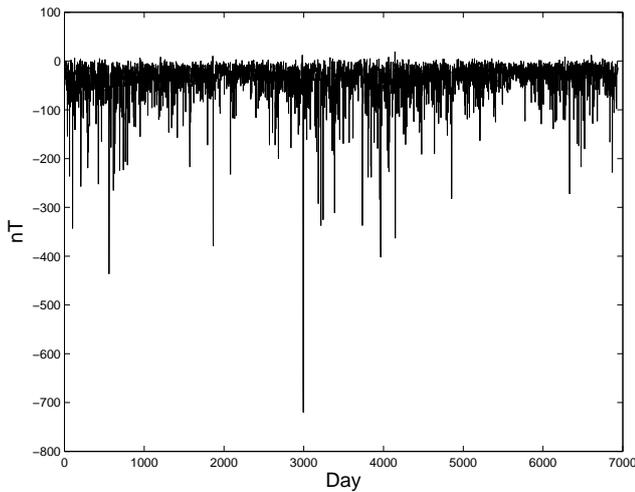


Fig. 1. The D_{st} series, measured in hourly intervals, from 1981 to 2002.

2 Iterated function systems for multifractal measures

In this section, we outline the concept of multifractal measure and present a class of models, namely recurrent iterated function systems, which will be used in this paper to represent a multifractal measure. These RIFS lead to an algorithm for prediction of the probability of patterns of storm events.

2.1 Multifractal measures

Magnetic storms are highly dynamic over many time scales. The D_{st} time series is apparently intermittent. This behavior is characterized by the generalized dimension of its measure, which is then known as a multifractal measure. The standard method to test for multifractality consists of calculating the moments of order q of a measure μ with support $S \subset \mathbb{R}$ (commonly normalized to have mass $\mu(S) = 1$):

$$M_l(q) = \sum_{\mu(B) \neq 0} (\mu(B))^q, \quad q \in \mathbb{R}, \quad (1)$$

where the summation runs over all different non-empty intervals $B = [kl, (k+1)l]$ of a given length l which cover the support S .

Multiple scaling holds if the moments of order q scale as

$$M_l(q) \sim l^{(q-1)D_q} \text{ as } l \rightarrow 0, \quad (2)$$

which defines the generalized dimensions D_q as

$$D_q = \begin{cases} \lim_{l \rightarrow 0} \frac{\log M_l(q)}{(q-1) \log l}, & q \neq 1, \\ \lim_{l \rightarrow 0} \frac{M_{1,l}}{\log l}, & q = 1, \end{cases} \quad (3)$$

where $M_{1,l} = \sum_{\mu(B) \neq 0} \mu(B) \log \mu(B)$ (Falconer, 1997). The value D_0 is known as the capacity dimension, D_1 the information dimension and D_2 the correlation dimension of the measure μ . A monofractal has all its dimensions identical: $D_q = \alpha$ and $M_l(q) \sim l^{(q-1)\alpha}$. It is seen that the sample values of the generalized dimensions D_q , denoted \overline{D}_q , can be

obtained through the linear regression of $(q-1)^{-1} \log M_l(q)$ against $\log l$ for $q \neq 1$, and through the linear regression of $M_{1,l}$ against $\log l$ for $q = 1$.

2.2 Measure representation

In this paper, we concentrate on developing models for the probability of occurrence of storm events. The proposed method examines the multiple scaling of a process via their measure representation. We first outline the method of Yu, Anh and Lau (2001) in deriving the measure representation of a time series. We assume that the time series can be grouped into a number of different categories. For example, each data point is classified according to $D_{st} \leq -50$ nT or $D_{st} > -50$ nT, which corresponds to the active or quiet category, respectively. Small storms, with D_{st} values above -50 nT, are placed in the quiet category since these are actually regarded to be substorms (Gonzalez et al., 1994, Wanliss et al., 2005). We then use the values $s=0, 1$ to indicate each category. We call any string made up of k numbers from the set $\{0, 1\}$ a k -string. For a given k there are in total 2^k different k -strings, and 2^k counters are needed to count the number of k -strings in a given time series. We divide the interval $[0, 1)$ into 2^k disjoint subintervals, and use each subinterval to represent a counter. Letting $s = s_1 \cdots s_k$, $s_i \in \{0, 1\}$, $i = 1, \dots, k$, be a substring with length k , we define

$$x_l(s) = \sum_{i=1}^k \frac{s_i}{2^i}, \quad x_r(s) = x_l(s) + \frac{1}{2^k}.$$

We then use the subinterval $[x_l(s), x_r(s))$ to represent the substring s . Let $N(s)$ be the number of times a substring s appears in the time series. If the time series has length L , we define $F(s) = N(s)/(L - k + 1)$ to be the frequency of substring s . It follows that $\sum_{\{s\}} F(s) = 1$. We can now view $F(s)$ as a function of x and define a measure μ on $[0, 1)$ by $\mu(x) = Y(x) dx$, where $Y(x) = 2^k F(s)$, $x \in [x_l(s), x_r(s))$. We call μ the “measure representation” of the given time series. It is noted that this histogram-type representation will have a different shape according to the order of the k -strings on the interval $[0, 1)$, but it is unique for each time series once this order is fixed (usually the dictionary order is used). This concept is an extension of that of the usual histogram, where each substring consists of a single value.

As an example, we consider the D_{st} index, plotted in Fig. 1 in hourly resolution over the period 1981–2002 (further detail on this index is provided in Sect. 3.1). The measure representation of D_{st} for $k=12$ is given in Fig. 2, which has 2^{12} subintervals. Here, the D_{st} time series is clustered into two categories: $D_{st} \leq -50$ and $D_{st} > -50$ as noted above. Self-similarity is apparent in the D_{st} series via its measure representation.

2.3 Recurrent iterated function system for a multifractal measure

In this paper, we model the measure μ as constructed above by a recurrent iterated function system (Barnsley and Demko, 1985; Falconer, 1997). This technique has been applied successfully to fractal image construction (Barnsley and Demko, 1985; Vrscay, 1991) and genomics (Anh et al., 2001, 2002, Yu et al., 2001, 2003), for example. Consider a system of N contractive maps $S=\{S_1, S_2, \dots, S_N\}$ and the associated matrix of probabilities $\mathbf{P}=(p_{ij})$ such that $\sum_{j=1}^N p_{ij}=1, i=1, 2, \dots, N$. We consider a random sequence generated by the dynamical system $S: x_{t+1}=S_{\sigma_t}(x_t), t=0, 1, 2, \dots$, where x_0 is any starting point and σ_t is chosen among the set $\{1, 2, \dots, N\}$ with a probability that depends on the previous index $\sigma_{t-1}: P(\sigma_{t+1}=i)=p_{\sigma_t, i}$. Then (S, \mathbf{P}) is called a ‘‘recurrent iterated function system’’. The coefficients in the contractive maps and the probabilities in the RIFS are the parameters to be estimated for the measure that we want to simulate.

We now describe the method of moments to perform this estimation (Vrscay, 1991). If μ is the invariant measure and E the attractor of the RIFS in \mathbb{R} , the moments of order ν of μ are $g_\nu = \int_E x^\nu d\mu, g_0 = \int_E d\mu = 1$. If $S_i(x) = c_i x + d_i, i=1, \dots, N$, then $g_\nu = \sum_{j=1}^N p_{ji} g_\nu^{(j)}$, where $g_\nu^{(j)}, j=1, \dots, N$, are given by the solution of the following system of linear equations:

$$\sum_{j=1}^N (p_{ji} c_i^\nu - \delta_{ij}) g_\nu^{(j)} = - \sum_{k=0}^{\nu-1} \binom{\nu}{k} \left[\sum_{j=1}^N c_i^k d_i^{\nu-k} p_{ji} g_k^{(j)} \right],$$

$$i = 1, \dots, N, \nu \geq 1.$$

For $\nu=0$, we set $g_0^{(i)} = m_i$, where m_i are given by the solution of the linear equations

$$\sum_{j=1}^N p_{ji} m_j = m_i, \quad i = 1, 2, \dots, N, \quad \text{and} \quad g_0 = \sum_{i=1}^N m_i = 1.$$

If we denote by G_ν the moments obtained directly from a given measure, and g_ν the formal expression of the moments obtained from the above formulas, then solving the optimization problem

$$\min_{c_i, d_i, p_{ji}} \sum_{\nu=1}^m (g_\nu - G_\nu)^2$$

for some chosen m will provide the estimates of the parameters of the RIFS.

Once the RIFS $(S_i(x), p_{ji}, i, j=1, \dots, N)$ has been estimated, its invariant measure can be simulated in the following way (Anh, Lau and Yu, 2002): Generate the attractor E of the RIFS via the dynamic system described above. Let χ_B be the indicator function of a subset B of the attractor E :

$$\chi_B(x) = \begin{cases} 1, & \text{if } x \in B, \\ 0, & \text{if } x \notin B. \end{cases}$$

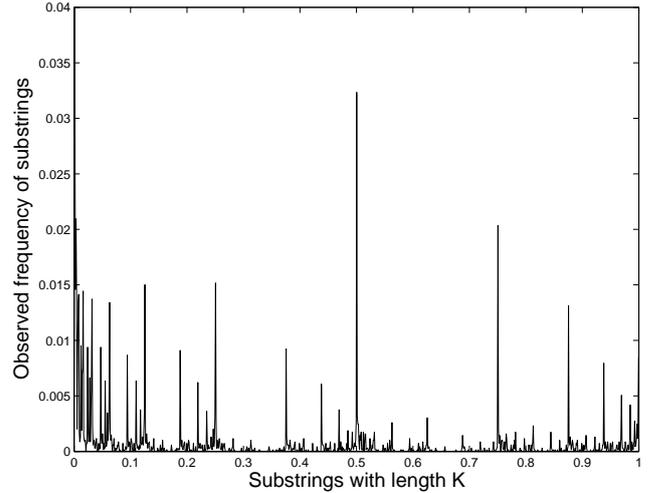


Fig. 2. The measure representation, constructed for $k=12$ with two levels, of the D_{st} time series of Fig. 1.

From the ergodic theorem for RIFS, the invariant measure is then given by

$$\mu(B) = \lim_{n \rightarrow \infty} \left[\frac{1}{n+1} \sum_{k=0}^n \chi_B(x_k) \right].$$

By definition, an RIFS describes the scale invariance of a measure. Hence a comparison of the given measure with the invariant measure simulated from the RIFS will confirm whether the given measure has this scaling behavior. This comparison can be undertaken by computing the cumulative walk of a measure, represented in the form of its histogram of k -strings, as $F_j = \sum_{i=1}^j (f_i - \bar{f}), j = 1, \dots, 2^k$, where f_i is the frequency of the i -th substring and \bar{f} is the average value of the histogram.

Returning to the D_{st} example of Sect. 2.2, an RIFS with 2 contractive maps $\left\{ S_1(x) = \frac{1}{2}x, S_2(x) = \frac{1}{2}x + \frac{1}{2} \right\}$ is fitted to the measure representation using the method of moments. Here, the interval $[0, 1)$ is divided into two non-overlapping subintervals $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$, hence it is natural to select $c_i = \frac{1}{2}$ and $d_i = 0$ or $\frac{1}{2}$ for the two maps. The optimization problem is run for moments up to order $m=15$, which is sufficiently large for the objective function to have negligible changes, hence for the estimates to converge. The estimates for the probabilities are

$$\mathbf{P} = \begin{pmatrix} 0.992068 & 0.007932 \\ 0.090440 & 0.909560 \end{pmatrix}.$$

The resulting invariant measure is plotted in Fig. 3. The cumulative walks of these two measures are reported in Fig. 4. It is seen that the fitted RIFS provides an excellent model of the scaling behavior of this D_{st} time series.

In order to check that the estimates of the parameters, hence the RIFS models, do not change over different solar cycles, we re-estimate the probabilities for two cycles inherent in the data:

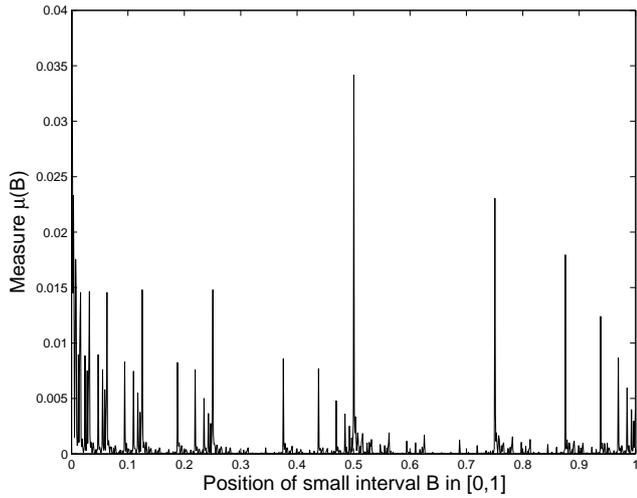


Fig. 3. Simulation of the fitted RIFS of the measure representation of Fig. 2.

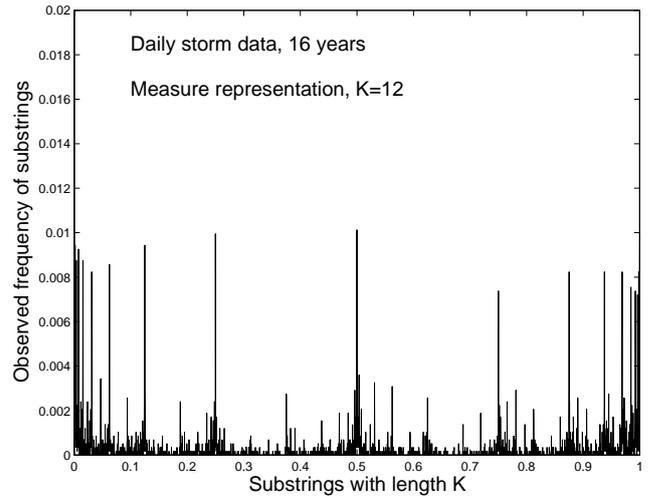


Fig. 5. The measure representation, constructed for $k=12$ with two levels, of the daily D_{St} time series over the period 1981–1996.

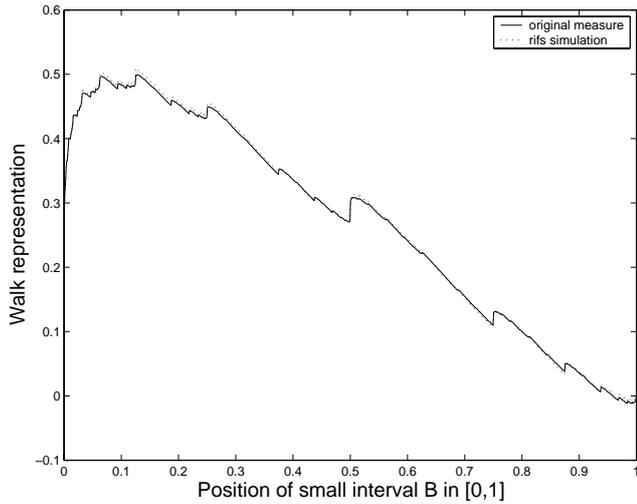


Fig. 4. The cumulative walks of the measure representation (Fig. 2) and its RIFS simulation (Fig. 3). The simulated model traces out closely the given measure.

For cycle 1 (1981–1991):

$$\mathbf{P} = \begin{pmatrix} 0.990986 & 0.009014 \\ 0.082972 & 0.917028 \end{pmatrix}$$

and for cycle 2 (1992–2002):

$$\mathbf{P} = \begin{pmatrix} 0.993075 & 0.006925 \\ 0.102738 & 0.897262 \end{pmatrix}.$$

It is seen that the estimates agree with each other over the two cycles and also over the entire period.

3 Prediction of storm events

3.1 Data analysis

The raw data set used in this work comes from the World Data Center (WDC-Kyoto) where an uninterrupted hourly time series is available from 1963 to the present time. We use the period 1981–2002 in this work. Such period provides sufficient information on intermittency of storm data for multifractal analyses. The hourly D_{St} time series from 1981 through 2002 is shown in Fig. 1. The time series appears stationary at this scale and a striking feature is its bursty negative excursions corresponding to intense storm events. In fact, zooming in on shorter time intervals shows the same pattern. This apparent scaling and intermittency of D_{St} suggests that multifractal techniques would be suitable for its analysis and prediction, which is what we follow in this paper.

In the next illustration, we construct a daily time series for the period 1981–1999 by taking the minimum value for each day of the original D_{St} series. We use the period 1981–1996 for modeling, leaving the last three years 1997–1999 for evaluation of our prediction method in the next subsection. Here, the D_{St} is clustered into two categories: $D_{St} > -30$ (corresponding to a no-storm event) and $D_{St} \leq -30$ (corresponding to a storm event). The measure obtained for 2 categories and $k=12$ is plotted in Fig. 5. Note that the self-similarity pattern of this measure is quite different from that presented in Fig. 2. An RIFS with 2 contractive maps $\{S_1, S_2\}$ is fitted to this measure using the method of moments. The resulting invariant measure is plotted in Fig. 6. The cumulative walks of these two measures are reported in Fig. 7. It is seen again that the fitted RIFS provides an excellent model of the scaling behavior of this daily time series.

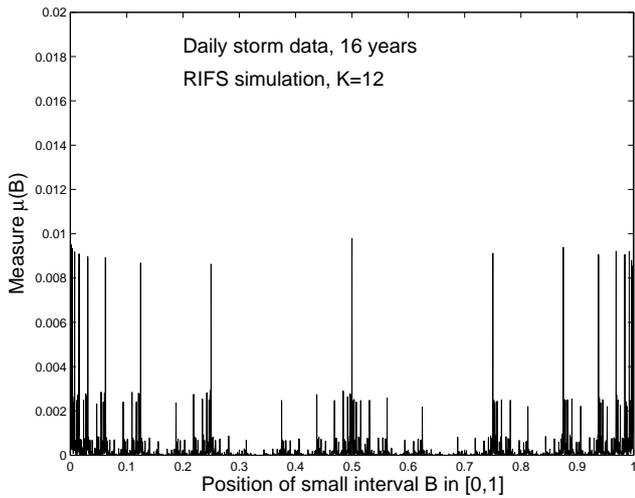


Fig. 6. Simulation of the fitted RIFS model of the measure representation of Fig. 5.

3.2 Prediction

In using measure representations, storm events are compounded into the patterns of the k -strings, and a probability is computed for each k -string. An RIFS is then fitted to this measure representation using the method of moments. The attractor of this RIFS is in fact a fractal set which describes the fractal behavior inherent in the measure representation. As described in Sect. 2.3, the contractive maps of the fitted RIFS can be used to simulate a measure, which is the invariant measure of the original D_{st} time series. This invariant measure gives the probability for each event in the attractor.

The fitted RIFS can also be used to predict future events according to their patterns and generated probabilities. For example, assume that we have observed 11 storm events in the past 11 days and we want to predict the occurrence of the next event. All the patterns, for example $D_{st} < -30$ nT for 11 consecutive days, of the 12-strings are known, together with their probabilities given by the fitted RIFS. We then select the 12-string (the first 11 symbols of which are known) with the highest probability. The last symbol/event of this 12-string will give us the desired prediction. To be concrete, assume that we use the map

$$f_1 = \begin{cases} 0, & \text{if } D_{st} > -30, \\ 1, & \text{if } D_{st} \leq -30 \end{cases}$$

to convert the D_{st} time series into a symbolic sequence of 0 (no-storm) and 1 (storm). Assume that the fitted RIFS gives the probabilities of 0.012 and 0.035 to the 12-strings 011011101111 and 011011101110, respectively. If it has been observed that the event 01101110111 occurred in the previous 11 days, then our method suggests to predict that there would be no storm in the next day as this event corresponds to a higher probability of 0.035 for the pattern to occur. It should be noted that this method of prediction is based on the invariant measure obtained from the model, which represents the scaling behavior of the underlying process.

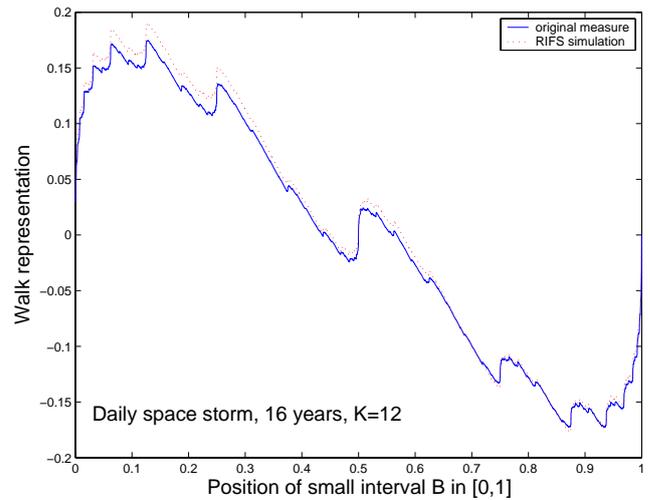


Fig. 7. The cumulative walks of the measure representation (Fig. 5) and its RIFS simulation (Fig. 6).

To evaluate the above method of prediction, we use three symbolic maps: f_1 as defined above with 1 meaning storm and 0 meaning no storm,

$$f_2 = \begin{cases} 0, & \text{if } D_{st} > -80, \\ 1, & \text{if } D_{st} \leq -80, \end{cases}$$

with 1 meaning big storm and 0 meaning no big storm, and

$$f_3 = \begin{cases} 0, & \text{if } D_{st} > -30, \\ 1, & \text{if } -100 < D_{st} \leq -30, \\ 2, & \text{if } D_{st} \leq -100, \end{cases}$$

with 2 meaning big storm, 1 meaning storm and 0 meaning no storm.

We will evaluate the prediction based on both the daily and hourly D_{st} data series over the period 1981–1996. Note that we will make true predictions as we attempt to predict D_{st} during 1997–1999, the data of which was not used in the estimation of the RIFS. After converting the daily or hourly data to symbolic sequences, the fitted RIFS are used to simulate their measure representations (we take $k=12$ for 2-symbol representations and $k=8$ for 3-symbol representations). We found from the cumulative walks that, for daily data, RIFS works well for both 2-symbol and 3-symbol representations; while for hourly data, RIFS works well only for the 2-symbol representations.

We obtain the predicted events for up to eight hours ahead, or up to three days ahead. This prediction is performed for each time point and repeated recursively over the full three years 1997–1999. For example, for s -hours ahead predictions based on f_1 and 12-strings, we start with the first 12- s known events in the time series (i.e. starting with the symbols from 1 to 12- s), then predict the next s events according to the 12-string with the highest probability (there are 2^s such 12-strings). The next prediction is then based on the known symbols from 2 to 13- s for the next s events. The prediction is repeated until the last time point, which is $T-s$, where T is the number of points of the time series.

Table 1. Prediction of hourly data using map f_1 with $k=12$.

Hours ahead	r_1	r_2	r_3
1	25094/26269=95.53%	3811/4540=83.94%	3811/4540=83.94%
2	24328/26269=92.61%	3484/5047=69.03%	3942/5047=78.24%
3	23711/26269=90.26%	3229/5464=59.10%	4048/5464=74.08%
4	23186/26269=88.26%	3052/5814=52.03%	4110/5814=70.69%
5	22709/26269=86.45%	2840/6131=46.32%	4155/6131=67.77%
6	22268/26269=84.77%	2674/6418=41.66%	4218/6418=65.72%
7	21877/26269=83.28%	2530/6680=37.87%	4284/6680=64.13%
8	21506/26269=81.87%	2393/6923=34.57%	4307/6923=62.21%

Table 2. Prediction of hourly data using map f_2 with $k=12$.

Hours ahead	r_1	r_2	r_3
1	26130/26269=99.47%	310/401=77.31%	310/401=77.31%
2	26039/26269=99.12%	271/457=59.30%	319/457=69.80%
3	25957/26269=98.80%	236/508=46.46%	328/508=64.57%
4	25890/26269=98.56%	213/555=38.38%	339/555=61.08%
5	25823/26269=98.30%	193/599=32.22%	347/599=57.93%
6	25763/26269=98.07%	175/642=27.26%	357/642=55.61%
7	25705/26269=97.85%	157/682=23.02%	360/682=52.79%
8	25648/26269=97.64%	140/722=19.39%	357/722=49.45%

Table 3. Prediction of daily data using map f_1 with $k=12$.

Days ahead	r_1	r_2	r_3
1	822/1084=75.83%	273/406=67.24%	273/406=67.24%
2	630/1084=58.12%	187/536=34.89%	299/536=55.78%
3	474/1084=43.73%	125/642=19.47%	311/642=48.44%

We then compare with real events as known from the data, and determine the accuracy of the prediction according to the following three indicators:

$$r_1 = \frac{\text{number of correct predictions}}{\text{total number of predictions}};$$

$$r_2 = \frac{Num_1}{Num_2},$$

where Num_1 denotes the number of correct predictions on strings whose last s events contain a storm event, and Num_2 the total number of strings whose last s events contain a storm event; and

$$r_3 = \frac{Num_3}{Num_4}.$$

where Num_3 denotes the number of predictions of storm events on strings whose last s events contain a storm event and Num_4 the total number of strings whose last s events contain a storm event. In these indicators, s is the number of hours or days ahead. In r_3 , the predicted string is not required

Table 4. Prediction of daily data using map f_3 with $k=8$.

Days ahead	r_1	r_2	r_3
1	780/1088=71.69%	230/406=56.65%	230/406=56.65%
2	593/1088=54.50%	150/536=27.99%	270/536=50.37%
3	449/1088=41.27%	98/642=15.26%	282/642=43.93%

Table 5. s -hours ahead prediction of hourly data using map f_1 with $k = 12$ based on a prefix 00 ... 01 of length $12-s$.

Hours ahead	r_1	r_2	r_3
1	170/212=80.19%	170/170=100%	170/170=100%
2	124/223=55.60%	124/185=67.03%	124/124=100%
3	110/233=47.21%	110/200=55.00%	110/110=100%
4	100/243=41.15%	100/213=46.95%	213/213=100%

Table 6. s -days ahead prediction of daily data using map f_1 with $k=12$ based on a prefix 00 ... 01 of length $12-s$.

Days ahead	r_1	r_2	r_3
1	14/21=66.67%	14/14=100%	14/14=100%
2	11/27=40.74%	11/18=61.11%	18/18=100%
3	11/32=34.38%	11/24=45.83%	24/24=100%
4	9/36=25.00%	9/28=32.14%	28/28=100%

to be the same as the observed string. Hence it is expected that $r_3 > r_2$. Also, since most of the values of the time series are larger than -30 nT (corresponding to a no-storm event), it is more difficult to predict a storm event than a no-storm event. And since r_1 includes the predictions on those strings whose last s events contain a no-storm event, it is expected that $r_1 > r_3 > r_2$. The results are reported in Tables 1–4.

3.3 Remark

The method of this paper is not suitable to predict a new storm onset given that there was no storm in the previous $k-1$ periods in a k -string setting, for example, the event 000000000001 in a 12-string. This situation arises because the probability of the event 000000000000 is always larger than that of 000000000001, hence the method always predicts the occurrence of 000000000000 even though 000000000001 would occur. A system for prediction of storm onsets such as 000000000001 would normally require incorporation of an internal magnetospheric mechanism involving solar wind as a driver. Our method requires the occurrence of at least one storm event in the previous $k-s$ periods to be able to predict the pattern in the next s periods. To confirm this point, we provide the following examples using the map f_1 on 12-strings of both hourly and daily D_{st} data. These are reported in Tables 5 and 6.

4 Conclusions

In this paper, we pay attention to the prediction of storm patterns up to three days ahead using daily data, or up to eight hours ahead using hourly data. Based on the values recorded, the D_{st} is clustered into events such as {storm, no storm} or {intense storm, moderate storm, small storm, no storm}. Some previous works have suggested the values to distinguish these events. For example, storms with $D_{st} < -50$ nT are classified as moderate or intense, and those in the range -50 nT $\leq D_{st} < -30$ nT classified as small storms (Gonzalez et al., 1994; Wanliss et al., 2005). In this way, the D_{st} time series is converted into a sequence of symbols {0, 1}, or {0, 1, 2, 3} accordingly. The events are then grouped into strings of k symbols, yielding k -strings. Going through the symbolic sequence, the probabilities of these k -strings can be obtained, yielding a probability measure for all possible k -strings. We call each of these measures a measure representation of the D_{st} time series.

The work of this paper indicates that each of the above probability measures is in fact a multifractal measure and can be modeled by a set of contractive maps known as a recurrent iterated function system. The excellent fit of this RIFS to data suggests that the attractor/fractal set of the RIFS is the set of k -strings started out. The fitted RIFS represents the multifractal scaling of storm events. This scaling is the key element in our method for prediction of storm events.

It should be noted that these are outside-sample forecasts, hence are quite meaningful. The method works reasonably well in predicting storm patterns for hourly data when we only pay attention to 2-symbol scenarios such as {storm, no storm}, or {big storm, no big storm}. A further point to note is that this performance is achieved from the scaling behavior of the D_{st} series captured in its measure representation. This distinguishes our approach from the usual approach based on the correlation structure of D_{st} . As the method does not rely on additional information such as a solar wind driver, it is simple to implement and its prediction can be used as a benchmark to evaluate more elaborate systems.

Acknowledgements. This work is partially supported by the NSF grant DMS-0417676 and the ARC grant DP0559807. The authors wish to thank the referees for their detailed comments and suggestions.

Edited by: D. Vassiliadis

Reviewed by: two referees

References

- Anh, V. V., Lau, K.-S., and Yu, Z. G.: Multifractal characterisation of complete genomes, *J. Phys. A, Math. Gen.*, 34, 7127–7139, 2001.
- Anh, V. V., Lau, K. S., and Yu, Z. G.: Recognition of an organism from fragments of its complete genome, *Phys. Rev. E*, 66, 031910, 1–9, 2002.
- Ayache, A. and Lévy Véhel, J.: The generalized multifractal Brownian motion, *Statistical Inference for Stochastic Processes*, 3, 7–18, 2000.
- Barnsley, M. F. and Demko, S.: Iterated function systems and the global construction of fractals, *Proc. R. Soc. London, Ser. A*, 399, 243–275, 1985.
- Benassi, A., Bertrand, P., Cohen, S., and Istas, J.: Identification of the Hurst exponent of a step multifractal Brownian motion, *Statistical Inference for Stochastic Processes* 3, 101–110, 2000.
- Burlaga, L. F.: Multifractal structure of the interplanetary magnetic field: Voyager 2 observations near 25 AU, 1987–1988, *Geophys. Res. Lett.*, 18, 69–72, 1991.
- Burlaga, L. F.: Lognormal and multifractal distributions of the heliospheric magnetic field, *J. Geophys. Res.*, 106, 15 917–15 927, 2001.
- Burlaga, L. F., Wang, C., and Ness, N. F.: A model and observations of the multifractal spectrum of the heliospheric magnetic field strength fluctuations near 40 AU, *Geophys. Res. Lett.*, 30, doi:10.1029/2003GL016903, 2003.
- Burton, R. K., McPherron, R. L., and Russell, C. T.: An empirical relationship between interplanetary conditions and Dst, *J. Geophys. Res.*, 80, 4204–4214, 1975.
- Falconer, K.: *Techniques in Fractal Geometry*, Wiley, 1997.
- Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rosotok, G., Tsurutani, B. T., and Vasyliunas, V. N.: What is a geomagnetic storm?, *J. Geophys. Res.*, 99, 5771–5792, 1994.
- Greenspan, M. E. and Hamilton, D. C.: A test of the Dessler-Parker-Sckopke relation during magnetic storms, *J. Geophys. Res.*, 72, 105, 5419–5430, 2000.
- Jaffard, S.: The multifractal nature of Lévy processes, *Prob. Th. Rel. Fields*, 114, 207–227, 1999.
- Kabin, K. and Papitashvili, V. O.: Fractal properties of the IMF and the Earth's magnetotail field, *Earth Planets Space*, 50, 87–90, 1998.
- Langel, R. K. and Estes, R. H.: large-scale, near-field magnetic fields from external sources and the corresponding induced internal field, *J. Geophys. Res.*, 90, 2487–2494, 1985.
- Lui, A. T. Y., Chapman, S. C., Liou, K., Newell, P. T., Meng, C. I., Brittnacher, M., and Parks, G. K.: Is the dynamic magnetosphere an avalanching system?, *Geophys. Res. Lett.*, 27, 911–914, 2000.
- Lui, A. T. Y.: Multiscale phenomena in the near-Earth magnetosphere, *J. Atmos. Sol.-Terr. Phys.*, 64, 125–143, 2002.
- Lui, A. T. Y., Lai, W. W., Liou, K., and Meng, C. I.: A new technique for short-term forecast of auroral activity, *Geophys. Res. Lett.*, 30, 1258, doi:10.1029/2002GL016505, 2003.
- Metzler, R. and Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamic approach, *Physical Reports*, 339, 1–72, 2000.
- Turner, N. E., Baker, D. N., Pulkkinen, T. I., and McPherron, R. L.: Evaluation of the tail current contribution to DST, *J. Geophys. Res.*, 105, 5431–5439, 2000.
- Vrscay, E. R.: Iterated function systems: theory, applications and inverse problem, in: *Fractal Geometry and Analysis*, edited by: Belair, J. and Dubuc, S., Kluwer, Dordrecht, 405–468, 1991.
- Wanliss, J. A.: Nonlinear variability of SYM-H over two solar cycles, *Earth Planets Space*, 56, e13–e16, 2004.
- Wanliss, J. A.: Fractal properties of SYM-H during quiet and active times, *J. Geophys. Res.*, 110, A03202, doi:10.1029/2004JA010544, 2005.
- Wanliss, J. A., Anh, V. V., Yu, Z. G., and Watson, S.: Multifractal modelling of magnetic storms via symbolic dynamics analysis, *J. Geophys. Res.*, accepted, 2005.

Yu, Z. G., Anh, V. V., and Lau, K. S.: Measure representation and multifractal analysis of complete genomes, *Phys. Rev. E*, 64, 031903, 1–9, 2001.

Yu, Z. G., Anh, V. V., and Lau, K. S.: Iterated function system and multifractal analysis of biological sequences, *International J. Modern Physics B* 17, 4367–4375, 2003.