12 Detection of Dynamical Complexity Changes in *Dst* Time Series Using Entropy Concepts and Rescaled Range Analysis

Georgios Balasis, Ioannis A. Daglis,
Anastasios Anastasiadis, and Konstantinos Eftaxias

Abstract
Using an array of diagnostic tools including entropy concepts and rescaled range analysis, we establish that the *Dst* index time series exhibits long-range correlations, and that the underlying stochastic process can be modeled as fractional Brownian motion. We show the emergence of two distinct patterns in the geomagnetic variability of the terrestrial magnetosphere: (1) a pattern associated with intense magnetic storms, which is characterized by a higher degree of organization (i.e., lower complexity or higher predictability for the system) and persistent behavior, and (2) a pattern associated with normal periods, which is characterized by a lower degree of organization (i.e., higher complexity or lower predictability for the system) and anti-persistent behavior.

12.1 Introduction

Studies of entropy provide physical insight into space plasma transport, intermittency, turbulence, and information flow in the heliosphere and magnetosphere (Wing and Johnson 2010). Moreover, entropy-based information theory can be used to characterize the dynamics of complex magnetospheric phenomena (Balasis et al. 2008, 2009; Balasis and Eftaxias 2009).

Here we present a detailed investigation of the *Dst* index time variations using entropy measures and rescaled range analysis (via the Hurst exponent) to search for long-time correlations and dynamical changes in the complex system of the magnetosphere. In this context, we seek a "good" complexity measure, i.e., a statistic quantifying regularity and complexity, which has application to relatively brief and noisy data.

We consider 1 year of *Dst* data (2001) including two intense magnetic storms, which occurred on 31 March 2001 and 6 November 2001 with minimum *Dst* values -387 nT and -292 nT respectively, as well as a number of weaker events (e.g. May and August 2001 with $Dst \sim -100$ nT in both cases). More precisely, the temporal evolution of nonlinear characteristics is studied by applying a variety of recently proposed entropy techniques: the original *Dst* time series is projected to a symbolic sequence and then analyses in terms of the classical Shannon entropy, dynamical (Shannon-like) block entropy, *T*-complexity and non-extensive Tsallis entropy follow. For the purpose of comparison we also analyze the original *Dst* data by means of approximate entropy. Finally, for the first time, the rescaled range analysis method is applied to the *Dst* data to calculate the values of the Hurst exponent. This analysis verifies

G. Balasis (✉)
Institute for Space Applications and Remote Sensing, National Observatory of Athens, Athens, Greece
e-mail: gbalasis@space.noa.gr

W. Liu, M. Fujimoto (eds.), *The Dynamic Magnetosphere*, IAGA Special Sopron Book Series 3,
DOI 10.1007/978-94-007-0501-2_12,

the results of an earlier fractal spectral analysis of the *Dst* index based on wavelet transforms (Balasis et al. 2006).

12.2 Theoretical Background

In this section we briefly introduce concepts of entropy and tools of information theory which will be used in the present study.

12.2.1 Fundamentals of Symbolic Dynamics

For the sake of completeness and for later use, we compile here the basic points of symbolic dynamics. Symbolic time series analysis is a useful tool for modelling and characterization of nonlinear dynamical systems. It provides a rigorous way of looking at "real" dynamics with finite precision (Hao 1989; Karamanos and Nicolis 1999). Briefly, it is a way of coarse-graining or simplifying the description.

The basic idea is quite simple. One divides the phase space into a finite number of partitions and labels each partition with a symbol (e.g. a letter from some alphabet). Instead of representing the trajectories by infinite sequences of numbers-iterates from a discrete map or sampled points along the trajectories of a continuous flow, one watches the alteration of symbols. Of course, in so doing one loses an amount of detailed information, but some of the invariant, robust properties of the dynamics may be kept, e.g. periodicity, symmetry, or the chaotic nature of an orbit (Hao 1989).

In the framework of symbolic dynamics, time series are transformed into a series of symbols by using an appropriate partition which results in relatively few symbols. After symbolization, the next step is the construction of "symbol sequences" ("words" in the language symbolic dynamics) from the symbol series by collecting groups of symbols together in temporal order.

To be more precise, the simplest possible coarse-graining of a time series is given by choosing a threshold C (usually the mean value of the data considered) and assigning the symbols "1" and "0" to the signal, depending on whether it is above or below the threshold (binary partition). Thus, we generate a symbolic time series from a 2-letter ($\lambda = 2$) alphabet (0, 1), e.g. 0110100110010110.... We usually read this symbolic sequence in terms of distinct consecutive "blocks" (words) of length $n = 2$. In this case one obtains 01/10/10/01/10/01/01/10/.... We call this reading procedure "lumping".

The number of all possible kinds of words is $\lambda^n = 2^2 = 4$, namely 00, 01, 10, 11. The required probabilities for the estimation of an entropy, p_{00}, p_{01}, p_{10}, p_{11} are the fractions of the blocks (words) 00, 01, 10, 11 in the symbolic time series, namely, 0, 4/16, 4/16, and 0, correspondingly. Based on these probabilities we can estimate, for example, the probabilistic entropy measure H_S introduced by Shannon (1948)

$$H_S = -\sum p_i \ln p_i \qquad (12.1)$$

where p_i are the probabilities associated with the microscopic configurations.

Various tools of information theory and entropy concepts are used to identify statistical patterns in the symbolic sequences, onto which the dynamics of the original system under analysis has been projected. For detection of an anomaly, it suffices that a detectable change in the pattern represents a deviation of the system from nominal behavior (Graben and Kurths 2003). Recent published work has reported novel methods for detection of anomalies in complex dynamical systems, which rely on symbolic time series analysis. Entropies depending on the word-frequency distribution in symbolic sequences are of special interest, extending Shannon's classical definition of the entropy and providing a link between dynamical systems and information theory. These entropies take a large/small value if there are many/few kinds of patterns, i.e., they decrease while the organization of patterns is increasing. In this way, these entropies can measure the complexity of a signal.

12.2.2 The Concept of Dynamical (Shannon-Like) Block Entropy

Block entropies, depending on the word-frequency distribution, are of special interest, extending Shannon's classical definition of the entropy of a single state to the entropy of a succession of states (Nicolis and Gaspard 1994).

Symbolic sequences, $\{A_1, \ldots, A_n, \ldots, A_L\}$ are composed of letters from an alphabet consisting of λ letters $\{A^{(1)}, A^{(2)}, \ldots, A^{(\lambda)}\}$. An English text, for example, is written on an alphabet consisting of 26 letters $\{A, B, C, \ldots, X, Y, Z\}$.

A word of length $n < L$, $\{A_1, \ldots, A_n\}$, is defined by a substring of length n taken from $\{A_1, \ldots, A_n, \ldots, A_L\}$. The total number of different words of length n which exists in the alphabet is $N_{\lambda n} = \lambda^n$.

We specify that the symbolic sequence is to be read in terms of distinct consecutive "blocks" (words) of length n,

$$\ldots, \underbrace{A_1, \ldots, A_n}_{B_1} \underbrace{A_{n+1}, \ldots, A_{2n}}_{B_2}, \ldots, \underbrace{A_{jn+1}, \ldots, A_{(j+1)n}}_{B_{j+1}}, \ldots \quad (12.2)$$

As stated previously, we call this reading procedure *lumping*. *Gliding* is the reading of the symbolic sequence using a *moving frame*. It has been suggested that, at least in some cases, the entropy analysis by lumping is much more sensitive than classical entropy analysis (gliding) (Karamanos 2000, 2001).

The probability $p^{(n)}(A_1, \ldots, A_n)$ of occurrence of a block $A_1, \ldots, A_n$ is defined by the fraction,

$$\frac{\text{No. of blocks, } A_1, \ldots, A_n, \text{ encountered when lumping}}{\text{Total no. of blocks}} \quad (12.3)$$

starting from the beginning of the sequence.

From the quantities characterize the information content of the symbolic sequence (Khinchin 1957; Ebeling and Nicolis 1992) we focus on the Shannon n-block entropy. Following Shannon's approach (Shannon 1948) the n-block entropy, $H(n)$, is given by

$$H(n) = -\sum p^{(n)}(A_1, \ldots, A_n) \ln p^{(n)}(A_1, \ldots, A_n) \quad (12.4)$$

The entropy $H(n)$ is a measure of uncertainty and gives the average amount of information necessary to predict a sub-sequence of length n.

12.2.3 T-complexity

T-entropy is a novel grammar-based complexity/information measure defined for finite strings of symbols (Ebeling et al. 2001; Titchener et al. 2005). It is a weighted count of the number of production steps required to construct a string from its alphabet. Briefly, it is based on the intellectual economy one makes when rewriting a string according to some rules.

An example of an actual calculation of the *T*-complexity for a finite string is given by Ebeling et al. (2001). We briefly describe how the *T*-complexity is computed for finite strings. The *T*-complexity of a string is defined by the use of one recursive hierarchical pattern copying (RHPC) algorithm. It computes the effective number of *T*-augmentation steps required to generate the string. The *T*-complexity may thus be computed effectively from any string and the resultant value is unique.

The string $x(n)$ is parsed to derive constituent patterns $p_i \in A^+$ and associated copy-exponents, $k_i \in N^+$, $i = 1, 2, \ldots, q$, where $q \in A^+$satisfying:

$$x = p_q^{k_q} p_{q-1}^{k_{q-1}} \ldots p_i^{k_i} \ldots p_1^{k_1} \alpha_0,\ \alpha_0 \in A \quad (12.5)$$

Each pattern p_i is further constrained to satisfy:

$$p_i = p_{i-1}^{m_{i,i-1}} p_{i-2}^{m_{i,i-2}}, \ldots, p_j^{m_{i,j}}, \ldots, p_1^{m_{i,1}} \alpha_i \quad (12.6)$$

$$\alpha_i \in A \text{ and } 0 \le m_{i,j} \le k_j \quad (12.7)$$

The *T*-complexity $C_T(x(n))$ is defined in terms of the copy-exponents k_i:

$$C(x(n)) = \sum_i^q \ln(k_i + 1) \quad (12.8)$$

One may verify that $C_T(x(n))$ is minimal for a string comprising a single repeating character.

The *T*-information $I_T(x(n))$ of the string $x(n)$ is defined as the inverse logarithmic integral, l_i^{-1}, of the *T*-complexity divided by a scaling constant ln 2:

$$I_T(x(n)) = l_i^{-1} \ln\left(\frac{C_T(x(n))}{\ln 2}\right) \quad (12.9)$$

In the limit $n \to \infty$ we have that $I_T(x(n)) \le \ln(\#A^n)$.

The form of the right-hand side may be recognizable as the maximum possible n-block entropy of Shannon's definition. The Neperian logarithm implicitly gives to the *T*-information the units of nats. $I_T(x(n))$ is the *T*-information of string $x(n)$. The *average T-information rate per symbol*, referred to here as the average T-entropy of $x(n)$ and denoted by $h_T(x(n))$, is defined along similar lines,

$$h_T\left(x\left(n\right)\right) = \frac{I_T\left(x\left(n\right)\right)}{n}\ (\text{nats/symbol}) \tag{12.10}$$

12.2.4 Principles of Non-extensive Tsallis Entropy

It has been established that physical systems which are characterized by longrange interactions or long-term memories, or are of a multi-fractal nature, are best described by a generalized statistical-mechanical formalism proposed by Tsallis (1988, 2009). More precisely, inspired by multifractals concepts, Tsallis introduced an entropic expression characterized by an index q which leads to non-extensive statistics (1988, 2009):

$$S_q = k\frac{1}{q-1}\left(1 - \sum_{i=1}^{W} p_i^q\right) \tag{12.11}$$

where p_i are probabilities associated with the microscopic configurations, W is their total number, q is a real number and k is Boltzmann's constant. The entropic index q describes the deviation of Tsallis entropy from the standard Boltzmann-Gibbs entropy. Indeed, using $p_i^{(q-1)} = e^{(q-1)\ln(p_i)} \sim 1 + (q-1)\ln\left(p_i\right)$ in the limit $q \to 1$, we recover the usual Boltzmann-Gibbs entropy

$$S_1 = -k\sum_{i=1}^{W} p_i \ln\left(p_i\right) \tag{12.12}$$

The entropic index q characterizes the degree of non-extensivity reflected in the following pseudo-additivity rule:

$$S_q\left(A+B\right) = S_a\left(A\right) + S_q\left(B\right) + \frac{1-q}{k}S_q\left(A\right)S_q\left(B\right) \tag{12.13}$$

For subsystems that have special probability correlations, extensivity

$$S_{B-G} = S_{B-G}\left(A\right) + S_{B-G}\left(B\right) \tag{12.14}$$

is not valid for $SB{-}G$, but may occur for Sq with a particular value of the index q. Such systems are sometimes referred to as non-extensive (Tsallis 1988, 2009).

The cases $q > 1$ and $q < 1$, correspond to sub-additivity, or super-additivity, respectively. We may think of q as a bias-parameter: $q < 1$ privileges rare events, while $q > 1$ privileges prominent events (Zunino et al. 2008).

We clarify that the parameter q itself is not a measure of the complexity of the system but measures the degree of non-extensivity of the system. It is the time variations of the Tsallis entropy for a given q, (S_q), that quantify the dynamic changes of the complexity of the system. Lower S_q values characterize the portions of the signal with lower complexity.

In terms of symbolic dynamics the Tsallis entropy for the word length n is (Balasis et al. 2008):

$$S_q\left(n\right) = k\frac{1}{q-1}\left(1 - \sum_{(A_1,A_2,\ldots,A_n)} \left[p(n)_{A_1,A_2,\ldots,A_n}\right]^q\right) \tag{12.15}$$

12.2.5 Approximate Entropy

Approximate entropy (*ApEn*), has been introduced as a quantification of regularity in time series data, motivated by applications to a wide variety of relatively brief, noisy data sets. *ApEn* could serve as a valuable tool for dynamically monitoring "health" status in a wide range of non-stationary systems. Therefore, it is here adopted for examining the dynamic system of the Earth's magnetosphere.

Related to time series analysis, *ApEn* provides a measure of the degree of irregularity or randomness within a series of data (of length N). *ApEn* was pioneered by Pincus as a measure of system complexity (Pincus 1991). It was introduced as a quantification of regularity in relatively brief and noisy data. It is rooted in the work of Grassberger and Procaccia (1983) and has been widely applied to biological systems (Pincus and Goldberger 1994; Pincus and Singer 1996 and references therein).

The *ApEn* examines time series for similar epochs: more similar and more frequent epochs lead to lower values of *ApEn*.

For a qualitative point of view, given N points, the *ApEn*-like statistics is approximately equal to the negative logarithm of the conditional probability that two sequences that are similar for m points remain similar, that is, within a tolerance r, at the next point. Smaller *ApEn*-values indicate a greater chance that a set of data will be followed by similar data (regularity), thus, smaller values indicate greater regularity. Conversely, a greater value for *ApEn* signifies a lesser chance of similar data being repeated (irregularity), hence, greater values convey more disorder, randomness and system complexity. Thus a low/high value of *ApEn* reflects a high/low degree of regularity. Notably, *ApEn* detects changes in underlying episodic behavior not reflected in peak occurrences or amplitudes (Pincus and Keefe 1992).

The following is a brief description of the calculation of *ApEn*. A more comprehensive description of *ApEn* may be found in (Pincus 1991; Pincus and Goldberger 1994; Pincus and Singer 1996).

Given any sequence of data points $u(i)$ from $i = 1$ to N, it is possible to define vector sequences $x(i)$, which consists of length m and are made up of consecutive $u(i)$, specifically defined by the following:

$$x(i) = (u[i], u[i+1], \ldots, u[i+m-1]) \quad (12.16)$$

In order to estimate the frequency that vectors $x(i)$ repeat themselves throughout the data set within a tolerance r, the distance $d(x[i], x[j])$ is defined as the maximum difference between the scalar components $x(i)$ and $x(j)$. Explicitly, two vectors $x(i)$ and $x(j)$ are "similar" within the tolerance or filter r, namely $d\left(x[i], x[j]\right) \leq r$, if the difference between any two values for $u(i)$ and $u(j)$ within runs of length m does not exceed r (i.e. $|u(i+k) - u(j+k)| \leq r\{\wr\nabla\ 0 \leq k \leq m)$. Subsequently, the correlation sum of vector $x(i)$ is

$$C_i^m = \frac{\left[\text{number of } j \text{ such that } d\left(x[i], x[j]\right) \leq r\right]}{(N-m+1)}$$

where $j \leq (N-m+1)$.

The $C_i^m(r)$ values measure, within a tolerance r, the regularity (frequency) of patterns similar to a given one of window length m. The parameter r acts like a filter value: within resolution r, the numerator count the number of vectors that are approximately the same as a given vector $x(i)$. The quantity $C_i^m(r)$ is called the correlation sum because it quantifies the summed (or global) correlation of vector $x(i)$ with all other vectors.

Taking the natural logarithm of $C_i^m(r)$, the mean logarithmic correlation sum of all vectors is defined as:

$$\Phi^m(r) = \sum_i \ln C_i^m(r)/(N-m+1) \quad (12.17)$$

where Σ_i is a sum from $i = 1$ to $(N-m+1)$. $\Phi^m(r)$ is a measure of the prevalence of repetitive patterns of length m within the filter r. Briefly, $\Phi_m(r)$ represents the average frequency of all the m-point patterns in the sequence remain close to each other.

Finally, $ApEn(m, r, N)$, is defined as the natural logarithm of the relative prevalence of repetitive patterns of length m as compared with those of length $m+1$:

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (12.18)$$

Thus, $ApEn(m, r, N)$ measures the logarithmic frequency that similar runs (within the filter r) of length m also remain similar when the length of the run is increased by 1. Small values of *ApEn* indicate regularity, given that increasing run length m by 1 does not decrease the value of $\Phi^m(r)$ significantly (i.e., regularity connotes that $\Phi^m[r] \approx \Phi^{m+1}(r)$). $ApEn(m, r, N)$ is expressed as a difference, but in essence it represents a ratio; note that $\Phi^m(r)$ is a logarithm of the averaged $C_i^m(r)$, and the ratio of logarithms is equivalent to their difference.

In summary, *ApEn* is a "regularity statistics" that quantifies the unpredictability of fluctuations in a time series. The presence of repetitive patterns of fluctuation in a time series renders it more predictable than a time series in which such patterns are absent. A time series containing many repetitive patterns has a relatively small *ApEn*; a less predictable (i.e., more complex) process has a higher *ApEn*.

12.2.6 Rescaled Range Analysis

Following the original work of Hurst (1951), the rescaled range (R/S)method is used to calculate the scaling exponent (Hurst exponent), H, to give quantitative measure of the persistence of a signal. First, when $0 < H < 1$ a signal can be modeled by fractional Brownian motion (fBm) (Henegham and McDarby 2000). $0.5 < H < 1$ is taken to indicate persistence, while $H = 0.5$ indicates an uncorrelated process. Persistence means that if the amplitude of the fluctuations increases in a time interval it is likely to continue increasing in the next interval. $0 < H < 0.5$ indicates anticorrelation or anti-persistence. Anti-persistence implies a set of fluctuations tending to induce stability within the system.

The R/S analysis is a statistical method to analyse long records of natural phenomena (Vanouplines 1995). There are two factors used in this analysis: firstly the range R, this is the difference between the minimum and maximum "accumulated" values or cumulative sum of $X(t,\tau)$ of the natural phenomenon at discrete integervalued time t over a time span τ, and secondly the standard deviation S, estimated from the observed values $Xi(t)$. Hurst found that the ratio R/S is very well described for a large number of natural phenomena by the following empirical relation:

$$R/S = (c\tau)^H \tag{12.19}$$

where τ is the time span, and H the Hurst exponent. The coefficient c was taken equal to 0.5 by Hurst. R is defined as:

$$R(\tau) = \max(X(t,\tau)) - \min(X(t,\tau)) \tag{12.20}$$

where t takes values from $[1\ \tau]$. S is given by:

$$S = \left\{ \frac{1}{\tau} \sum_{t=1}^{\tau} \left[\xi(t) - \langle\xi\rangle_\tau\right]^2 \right\}^{1/2} \tag{12.21}$$

where $\langle\xi\rangle_\tau = \sum_{t-1}^{\tau} \xi(t)/\tau$ and $X(t,\tau) = \sum_{u=1}^{t} \left[\xi(u) - \langle\xi\rangle_\tau\right]$. This method handles observations in time. The graphical representation uses time in the abscissa, and the observed value in the ordinate.

12.3 Results

In Fig. 12.1 the *Dst* time series is presented. The 1 year *Dst* data (2001) are divided into 5 shorter time series (see triangles denoting 5 distinct time intervals in Fig. 12.1). The 2nd and 4th time windows include the *Dst* variations associated to the 2 intense magnetic storms of 31/3/2001 and 6/11/2001, respectively. In Fig. 12.1 the values of the H parameter calculated by two different methods (i.e., fractal spectral analysis using wavelets (see Balasis et al. 2006, 2008, 2009) and R/S analysis) are also shown for the *Dst* index data. Figure 12.1 shows that *Dst* variations follow the fBm model (H always lies between 0 and 1) and exhibit persistent properties ($0.5 < H < 1$) around 31 March and 6 November 2001 magnetic storms (c.f. parts of H plot calculated by fractal spectral analysis marked in red represent persistency). In general, there is good agreement with the results for the H exponent

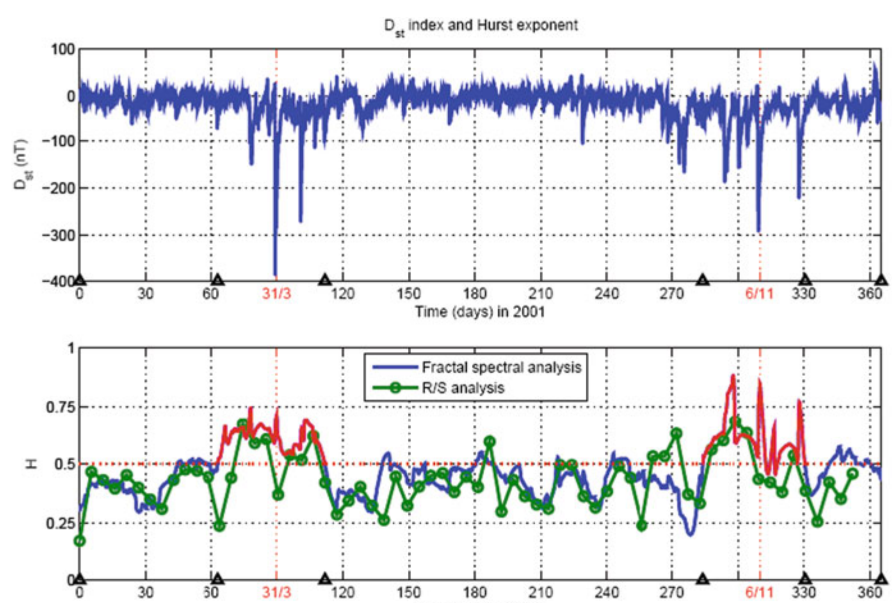

Fig. 12.1 *Dst* time series (*upper panel*) and Hurst exponents H (*lower panel*) calculated by fractal spectral analysis (*blue color*) and R/S analysis (*green color*). The 31 March and 6 November 2001 magnetic storms are marked with *red*. The *red dashed line* in H plot marks the transition between anti-persistent and persistent behavior. The triangles denote 5 time intervals corresponding to windows related to intense magnetic storms (2nd and 4th) and windows related to normal times (1st, 3rd and 5th)

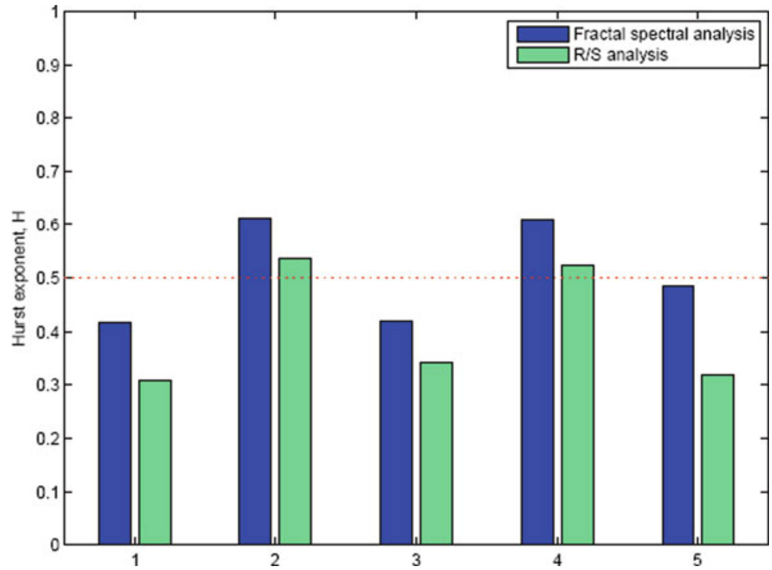

Fig. 12.2 Average values of the Hurst exponents H calculated by fractal spectral analysis (*blue color*) and R/S analysis (*green color*) at the 5 time windows indicated in Fig. 12.1. The *red dashed* line marks the transition between anti-persistent and persistent behavior. It is evident that time windows 1, 3 and 5 correspond to *Dst* index data with anti-persistent behavior, whereas, time windows 2 and 4 represent *Dst* time series with persistent behavior

from the R/S method. Thus, it is evident that the onset and development of the magnetic storms of 31/3/2001 and 6/11/2001 are associated with persistent behavior, which suggests that the underlying dynamics is governed by a positive feedback mechanism.

In Fig. 12.2 we show the average values of the Hurst exponents H calculated by fractal spectral analysis (blue color) and R/S analysis (green color) at the 5 time windows of Fig. 12.1. We stress that as both methods state the anti-persistent epochs ($0 < H < 0.5$)

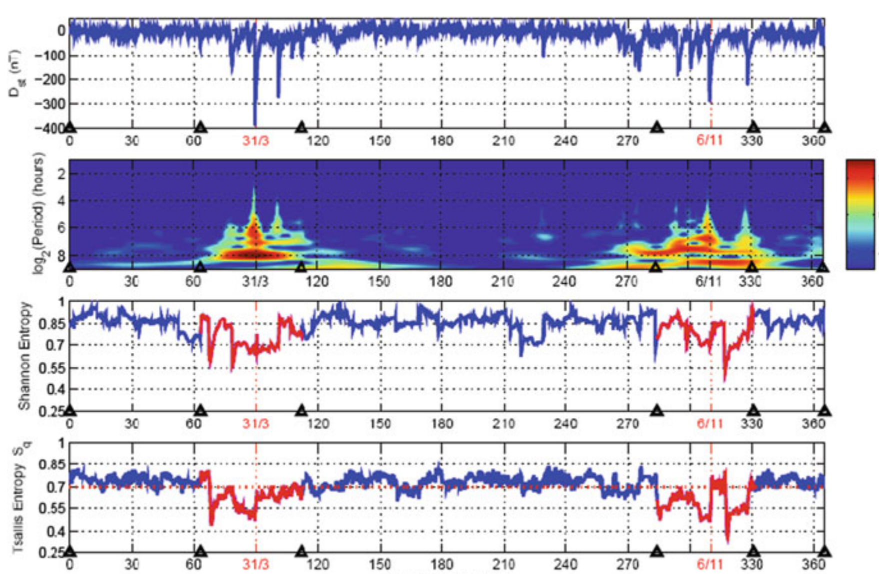

Fig. 12.3 From *top* to *bottom*: *Dst* time series along with its wavelet power spectrum for 2001 and time variations of Shannon entropies and Tsallis entropies, S_q. The 31 March and 6 November 2001 magnetic storms are marked with *red*. The *red dashed line* in S_q plot marks a possible boundary value for the transition to the lower complexity characterizing the different state of the magnetosphere. The triangles denote 5 time intervals (as in Fig. 12.1) in which: 1st, 3rd and 5th time windows correspond to anti-persistent ($0 < H < 0.5$) or high Tsallis entropies epochs; second and fourth time windows correspond to persistent ($0.5 < H < 1$) or lower Tsallis entropies epochs

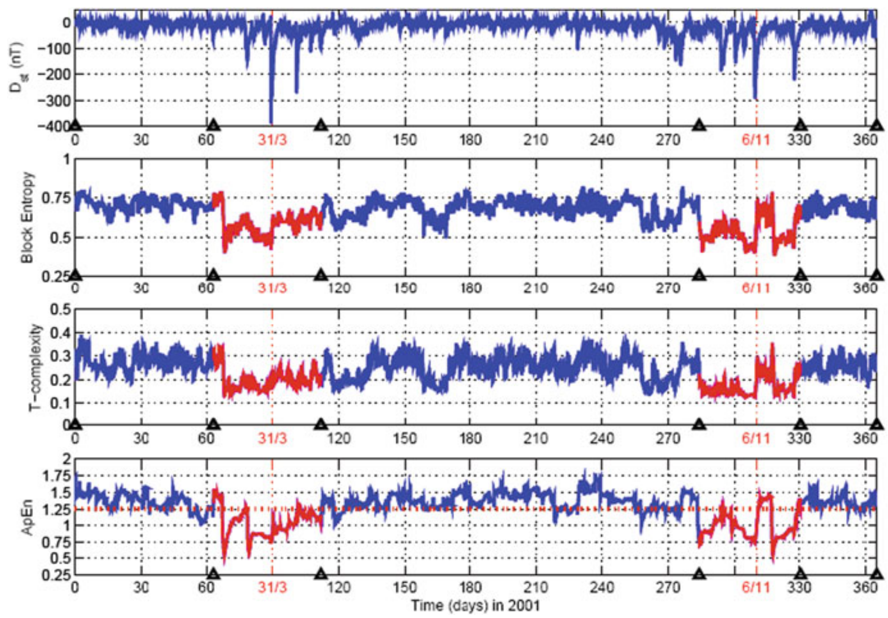

Fig. 12.4 From *top* to *bottom*: *Dst* time series along with time variations of block entropies, approximate entropies *ApEn* and *T*-complexities. The 31March and 6 November 2001 magnetic storms are marked with *red*. The *red dashed line* in *ApEn* plot marks a possible boundary value for the transition to the lower complexity characterizing the different state of the magnetosphere. The triangles denote 5 time intervals (as in Figs. 12.1 and 12.3)

correspond to the normal (quiet-time) magnetosphere (1st, 3rd and 5th time windows where H is below 0.5), while, the persistent epochs ($0.5 < H < 1$) correspond to the abnormal (storm-time) magnetosphere (2nd and 4th time windows where H is above 0.5). This finding supports the existence of two different epochs referring to two distinct states of the magnetic storm evolution.

Figure 12.3 gives the temporal evolution of *Dst* along with its wavelet power spectrum (Balasis et al. 2006) and corresponding time variations of Shannon entropy and Tsallis entropy for the whole year of 2001. In terms of entropy measures, we see how nicely Tsallis entropy variations identify the different complexity regimes in the *Dst* time series (c.f. red part of the corresponding plot). Figure 12.3 further demonstrates that Tsallis entropy yields superior results in comparison to Shannon entropy regarding the detection of dynamical complexity in the Earth's magnetosphere (i.e., offer a clearer picture of the transition from normal state to magnetic storms). A possible explanation for this is that Tsallis is an entropy obeying a non-extensive statistical theory, which is different from the usual Boltzmann-Gibbs statistical mechanics obeyed by Shannon entropy. Therefore, it is expected to better describe the dynamics of the magnetosphere, which is a nonequilibrium physical system with large variability.

Figure 12.4 gives the temporal evolution of *Dst* along with corresponding time variations of the block entropy, the *T*-complexity and, the *ApEn* for the whole year of 2001. We see how nicely the entropy measures identify the different complexity regimes in the *Dst* time series (c.f. red part of the corresponding plots). Figure 12.4 further demonstrates that the *ApEn* entropy yields superior results in comparison to the other entropy measures regarding the detection of dynamical complexity in the Earth's magnetosphere (i.e., offer a clearer picture of the transition). A possible explanation for this is that *ApEn* is more stable when dealing with nonstationary signals of dynamical systems (such the magnetospheric signal) than the rest of the entropy measures presented in Fig. 12.4.

12.4 Discussion and Conclusions

Entropy can provide the basis for information theory, which can be used to analyze the dynamics of a complex system. An example of such an application in the terrestrial magnetosphere is provided here. Our study uses several complexity measures (e.g. Tsallis entropy, approximate entropy, and Hurst exponent) to analyze *Dst* during storm and nonstorm ("normal") times. Moreover, Hurst exponent can be used to characterize the persistence of a system, e.g., whether the trend of the fluctuations will continue ("persistent") or differ ("anti-persistent") in the next time interval. We also find that the nonstorm intervals have higher entropies and lower persistence than storm intervals. Tsallis entropy, which is based on a non-extensive statistical theory, rather than Boltzmann-Gibbs statistical mechanics, can better describe the dynamics of the magnetosphere because the latter is a nonequilibrium system with large variabilities.

We stress that the anti-persistent time windows correspond to the time windows of higher entropies, while the persistent time windows correspond to the time windows of lower entropies. Importantly, a recent analysis presented by Carbone and Stanley (2007) shows that anti-correlated time series, with Hurst exponent $0.5 < H < 1$, are characterized by entropies greater than correlated time series having $0.5 < H < 1$. This suggestion is in agreement with our results. Wanliss (2005) and Wanliss and Dobias (2007) analyzed fluctuations of the *SYM-H* index around magnetic storms and found that there was a rapid and unidirectional change in the Hurst scaling exponent at the time of storm onset indicating a nonequilibrium dynamical phase transition.

An important remark is the agreement of the results between the linear analysis in terms of the Hurst exponent and nonlinear entropy analyses. A combination of linear and nonlinear analysis techniques can offer a firm warning that the onset of an intense magnetic storm is imminent.

As an extension of this application, we will consider in the near future a basic space weather challenge (Daglis et al. 2001, 2003, 2009), i.e., the problem of continuous monitoring of the magnetospheric condition, where the time series is not a fixed and complete set, but is "streaming". If we can associate a change in signal complexity with a change in the condition of the system, then we can hope that an entropy-like measure will be able to detect a developing storm and potential problem for space systems (and possibly provide some warning before system failure).

This would be the ultimate task of our research efforts.

Acknowledgements The *Dst* data are provided by the World Data Center for Geomagnetism, Kyoto (http://swdcwww.kugi.kyoto-u.ac.jp/).

References

Balasis G, Eftaxias K (2009) A study of non-extensivity in the Earth's magnetosphere. Eur Phys J Special Topics 174:219–225

Balasis G, Daglis IA, Kapiris P, Mandea M, Vassiliadis D, Eftaxias K (2006) From prestorm activity to magnetic storms: a transition described in terms of fractal dynamics. Ann Geophys 24:3557–3567

Balasis G, Daglis IA, Papadimitriou C, Kalimeri M, Anastasiadis A, Eftaxias K (2008) Dynamical complexity in Dst time series using non-extensive Tsallis entropy. Geophys Res Lett. doi:10.1029/2008GL034743

Balasis G, Daglis IA, Papadimitriou C, Kalimeri M, Anastasiadis A, Eftaxias K (2009) Investigating dynamical complexity in the magnetosphere using various entropy measures. J Geophys Res. doi:10.1029/2008JA014035

Carbone A, Stanley H (2007) Scaling properties and entropy of long-range correlated time series. Physica A 384:267–271

Daglis IA, Baker DN, Galperin Y, Kappenman JG, Lanzerotti LJ (2001) Technological impacts of space storms: outstanding issues. Eos Trans AGU. doi:10.1029/01EO00340

Daglis IA, Kozyra J, Kamide Y, Vassiliadis D, Sharma A, Liemohn M, Gonzalez W, Tsurutani B, Lu G (2003) Intense space storms: critical issues and open disputes. J Geophys Res. doi:10.1029/2002JA009722

Daglis IA, Balasis G, Ganushkina N, Metallinou F-A, Palmroth M, Pirjola R, Tsagouri IA (2009) Investigating dynamic coupling in geospace through the combined use of modeling, simulations and data analysis. Acta Geophys. doi:10.2478/s11600-008-0055-5

Ebeling W, Nicolis G (1992) Word frequency and entropy of symbolic sequences: a dynamical Perspective. Chaos Solitons Fractals 2:635–650

Ebeling W, Steuer R, Titchener M (2001) Partition-based entropies of deterministic and stochastic maps. Stochast Dyn 1:45–61

Graben P, Kurths J (2003) Detecting subthreshold events in noisy data by symbolic dynamics. Phys Rev Lett 90:100602(1–4).

Grassberger P, Procaccia I (1983) Estimation of the Kolmogorov entropy from a chaotic signal. Phys Rev A 28:2591–2593

Hao B-L (1989) Elementary symbolic dynamics and chaos in dissipative systems. World Scientific, Singapore

Henegham C, McDarby G (2000) Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. Phys Rev E 62:6103–6110

Hurst HE (1951) Long-term storage of reservoirs: an experimental study. Trans Am Soc Civ Eng 116:770–799

Karamanos K (2000) From symbolic dynamics to a digital approach: chaos and transcendence. Lect Notes Phys 550:357–371

Karamanos K (2001) Entropy analysis of substitutive sequences revisited. J Phys A: Math Gen 34:9231–9241

Karamanos K, Nicolis G (1999) Symbolic dynamics and entropy analysis of Feigenbaum limit sets. Chaos Solitons Fractals 10(7):1135–1150

Khinchin AI (1957) Mathematical foundations of information theory. Dover, New York, NY

Nicolis G, Gaspard P (1994) Toward a probabilistic approach to complex systems. Chaos Solitons Fractals 4(1):41–57

Pincus S (1991) Approximate entropy: a complexity measure for biologic time series data. In: Proceedings of IEEE 17th annual northeast bioengineering conference, IEE Press, New York, NY, p 35–36

Pincus S, Keefe D (1992) Quantification of hormone pulsatility via an approximate entropy algorithm. Am J Physiol (Endocrinol Metab) 262: E741–E754

Pincus S, Goldberger A (1994) Physiological time-series analysis: what does regularity quantify? Am J Physiol 266:H1643–H1656

Pincus S, Singer B (1996) Randomness and degree of irregularity. Proc Natl Acad Sci USA 93:2083–2088

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Titchener M, Nicolescu R, Staiger L, Gulliver A, Speidel U (2005) Deterministic complexity and entropy. Fund Inform 64:443–461

Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. J Stat Phys 52:479–487

Tsallis C (2009) Introduction to nonextensive statistical mechanics, approaching a complex word. Springer, Berlin

Vanouplines P (1995) Rescaled range analysis and the fractal dimension of pi. University library, Free University Brussels, Brussels, Belgium. http://ftp.vub.ac.be/_pvouplin/pi/rswhat.htm

Wanliss JA (2005) Fractal properties of SYM-H during quiet and active times. J Geophys Res. doi:10.1029/2004JA010544.

Wanliss JA, Dobias P (2007) Space storm as a dynamic phase transition. J Atmos Sol Terr Phys 69:675–684

Wing S, Johnson J.R. (2010) Introduction to special section on entropy properties and constraints related to space plasma transport. J Geophys Res. doi:10.1029/2009JA014911

Zunino L, Perez D, Kowalski A, Martin M, Garavaglia M, Plastino A, Rosso O (2008) Fractional Brownian motion, fractional Gaussian noise and Tsallis permutation entropy. Physica A 387:6057–6068