Temporal organization of magnetospheric fluctuations unveiled by recurrence patterns in the Dst index

Reik V. Donner, Veronika Stolbova, Georgios Balasis, Jonathan F. Donges, Marina Georgiou, Stelios M. Potirakis, and Jürgen Kurths

---

Articles you may be interested in

Deriving inferential statistics from recurrence plots: A recurrence-based test of differences between sample distributions and its comparison to the two-sample Kolmogorov-Smirnov test
Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 085712 (2018); 10.1063/1.5024915

Recurrence-based analysis of barrier breakup in the standard nontwist map
Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 085717 (2018); 10.1063/1.5021544

Recurrence threshold selection for obtaining robust recurrence characteristics in different embedding dimensions
Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 085720 (2018); 10.1063/1.5024914

Recurrence network modeling and analysis of spatial data
Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 085714 (2018); 10.1063/1.5024917

Extended recurrence plot and quantification for noisy continuous dynamical systems
Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 085722 (2018); 10.1063/1.5025485

Multivariate weighted recurrence network analysis of EEG signals from ERP-based smart home system
Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 085713 (2018); 10.1063/1.5018824

Temporal organization of magnetospheric fluctuations unveiled by recurrence patterns in the Dst index

Reik V. Donner,[1] Veronika Stolbova,[1,2,3,4,5] Georgios Balasis,[6] Jonathan F. Donges,[7,8] Marina Georgiou,[6,9,10] Stelios M. Potirakis,[11] and Jürgen Kurths[1,2,4,12]

[1]*Research Domain IV—Transdisciplinary Concepts and Methods, Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany*
[2]*Department of Physics, Humboldt University, Newtonstraße 15, 12489 Berlin, Germany*
[3]*Department of Banking and Finance, University of Zurich, Plattenstraße 14, 8032 Zurich, Switzerland*
[4]*Department of Control Theory, Lobachevsky State University of Nizhny Novgorod, Prospekt Gagarina 23, 603022 Nizhny Novgorod, Russia*
[5]*Center of Economic Research, ETH Zurich, Weinbergstraße 56/58, 8092 Zurich, Switzerland*
[6]*Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, I. Metaxa & Vas. Pavlou Street, 15236 Penteli, Greece*
[7]*Research Domain I—Earth System Analysis, Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany*
[8]*Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 11419 Stockholm, Sweden*
[9]*Department of Physics, Section of Astrophysics, Astronomy and Mechanics, National and Kapodistrian University of Athens, Panepistimioupoli Zografou, 15784 Athens, Greece*
[10]*Mullard Space Science Laboratory, Department of Space & Climate Physics, University College London, Holmbury Hill Road, Dorking RH5 6NT, Surrey, United Kingdom*
[11]*Department of Electrical and Electronics Engineering, University of West Attica, Campus 2, 250 Thivon and P. Ralli, Aigaleo, 12244 Athens, Greece*
[12]*Institute for Complex Systems and Mathematical Biology, University of Aberdeen, King's College, Aberdeen AB24 3FX, United Kingdom*

Magnetic storms constitute the most remarkable large-scale phenomena of nonlinear magnetospheric dynamics. Studying the dynamical organization of macroscopic variability in terms of geomagnetic activity index data by means of complexity measures provides a promising approach for identifying the underlying processes and associated time scales. Here, we apply a suite of characteristics from recurrence quantification analysis (RQA) and recurrence network analysis (RNA) in order to unveil some key nonlinear features of the hourly Disturbance storm-time (Dst) index during periods with magnetic storms and such of normal variability. Our results demonstrate that recurrence-based measures can serve as excellent tracers for changes in the dynamical complexity along non-stationary records of geomagnetic activity. In particular, trapping time (characterizing the typical length of "laminar phases" in the observed dynamics) and recurrence network transitivity (associated with the number of the system's effective dynamical degrees of freedom) allow for a very good discrimination between magnetic storm and quiescence phases. In general, some RQA and RNA characteristics distinguish between storm and non-storm times equally well or even better than other previously considered nonlinear characteristics like Hurst exponent or symbolic dynamics based entropy concepts. Our results point to future potentials of recurrence characteristics for unveiling temporal changes in the dynamical complexity of the magnetosphere. *Published by AIP Publishing.* https://doi.org/10.1063/1.5024792

Geomagnetic activity indices trace the temporal variability of the Earth's magnetic field across different spatial domains of the near-Earth environment. Among others, the Disturbance storm-time (Dst) index has been used in many previous studies as a diagnostic of the overall state of the magnetosphere. Here, we employ different techniques based upon the concept of recurrence plots to improve our understanding of the complex variability patterns exhibited by this index when the geomagnetic field undergoes a sequence of magnetic storm and quiescence periods triggered by non-stationary solar wind forcing. We demonstrate that recurrence characteristics provide unique tools for discriminating between the dynamical complexity properties of Dst during times with strong geomagnetic activity and quiescence phases.

Thus, future applications of these measures to other geomagnetic activity indices with higher temporal resolution may potentially allow us to identify characteristic signatures of complexity variations preceding intense magnetic storms, which could open new perspectives for space weather short-term forecasting.

I. INTRODUCTION

Various effects of space weather present natural hazards to which the modern human civilization has become increasingly vulnerable, particularly through the use of ever more sophisticated technologies (see Ref. 1 and references therein). The magnetosphere of the Earth is generally a benign host

for satellite communication and global positioning systems but can change into a quite inhospitable environment. The miniaturization of electronic components with which spacecrafts are equipped renders them susceptible to damage by charged particles accelerated to high energies by impulsive geomagnetic field disturbances.[2] On the surface of the Earth, electrical currents induced during geospace storms, when absorbed, can damage long-line power networks connecting large geographic regions.[3–5]

Geospace magnetic storms occur as a disturbance in the Earth's magnetic field, driven by large-scale eruptions of plasma and magnetic fields from the solar corona launched onto a trajectory that impacts the Earth's magnetosphere.[6,7] Similar to severe weather phenomena, as well as other types of natural disasters associated with the Earth's internal dynamics like earthquakes and volcanic eruptions, they vary remarkably in the severity of disturbance. However, unlike the latter events, geospace storms have a global reach, and their effects can be seen simultaneously around vast parts of the Earth. The solar wind provides a continuous input in the form of mass, momentum, and energy. If not dissipated, it is stored in the magnetotail until, through a sequence of energy-loading and stress-developing processes, the magnetospheric system is reconfigured.[8] During magnetic storms, charged particles in the radiation belts are accelerated to high energies resulting in intensified electric current systems causing characteristic signatures in the Earth's magnetic field.[9,10]

As a consequence of the solar wind forcing, the magnetosphere is continuously far from equilibrium and undergoes complex variations at a broad range of temporal and spatial scales.[11] Its response is typically not proportional to the forcing and commonly changes abruptly rather than gradually. Taken together, there is strong evidence to consider the magnetosphere as a complex system with distinct, nonlinearly coupled regions, where multiple interlinked phenomena occur on a vast range of length and time scales.[12] Since the first pieces of evidence of large-scale coherence provided by observations of low-dimensional behavior in time series of auroral electrojet indices,[13] studying magnetospheric activity dynamics from the viewpoint of nonlinear dynamics and complexity has gained substantial new insights into the response of the magnetosphere to solar wind energy input.

Taking our understanding of the temporal variations of the magnetic field at the surface of the Earth and in the surrounding space a step further, additional evidence of a hierarchical multi-scale organization of magnetospheric activity has been found in the form of characteristic scaling laws in some dynamical properties of regional and global geomagnetic activity indices.[14,15] By making use of a phase transition approach,[16,17] global coherency was reconciled with scale-invariance. Specifically, it was demonstrated that some global features exhibit properties typical for phase transitions of first order, while the multi-scale properties appear compatible with second-order transitions.[18,19]

In this work, we utilize the powerful framework of recurrence analysis for studying some additional nonlinear properties of the Earth's magnetosphere. The fundamental idea of recurrence analysis is based on the long-known fact that many natural processes obey a distinct recurrent behavior in time, ranging from very regular diurnal or annual variability in meteorological variables over almost periodic oscillatory patterns (such as Milankovitch cycles in the Earth's orbital parameters and resulting climate history or the return intervals of extrema of cosmic-ray intensity measured at the surface of the Earth[20]) to more irregular climate modes such as the El Niño Southern Oscillation. From a dynamical system's perspective, the recurrence of states (i.e., finding new states arbitrarily close to previously assumed states if waiting long enough) is a fundamental property of both deterministic and stochastic dynamics.[21,22] Recently, it has been proven mathematically that the temporal pattern of such recurrences allows for reconstructing the dynamics of the underlying variable up to a monotonous transformation, which implies that recurrences contain fundamental information about the dynamical organization of the system under study.[23]

Here, we utilize a suite of measures from recurrence quantification analysis (RQA)[21] and recurrence network analysis (RNA)[24] to study the nonlinear dynamics exhibited by hourly values of the Disturbance storm-time (Dst) index (i.e., the average change of the horizontal component of the Earth's magnetic field recorded at four mid-latitude magnetic observatories) during one year of observations, specifically, the year 2001. On the one hand, one year of hourly resolution data provides a reasonable sample for the application of even complex statistical analysis techniques but does not yet pose any substantial restrictions regarding the available computational capacity. On the other hand, as already emphasized in various previous studies,[25,26] the year 2001 provides an ideal test case for such non-conventional methods since it was characterized by a succession of several distinct phases of enhanced geomagnetic activity (including strong magnetic storms) alternating with periods of relative quiescence (i.e., with an absence of relevant magnetic storms) and, thus, covers the general variety of observable geomagnetic variability patterns reasonably well.

It should be noted that both RQA and RNA have recently proven useful for quantifying dynamical complexity in nonstationary models as well as real-world time series,[27–29] including applications to geoscientific problems.[30,31] In this work, both frameworks are combined to study the complex signatures of magnetospheric fluctuations during non-storm and storm conditions as reflected in the Dst index. Our aim is to unveil non-trivial features of magnetospheric dynamics that are not captured by other linear and nonlinear methods of time series analysis. However, by focusing on a single geomagnetic activity index only, the primary goal of this work is to establish the applicability of quantitative recurrence-based characteristics in the context of geomagnetism and space weather, while follow-up studies[32] on a larger set of relevant variables allowing to obtain more detailed results and interpretations on specific geophysical processes and mechanisms are outlined as a subject of future work.

This paper is organized as follows: Sec. II describes the data and methods used in this study. The temporal variability of the dynamical complexity of the Dst index as uncovered by various recurrence-based characteristics is reported in Sec. III. Subsequently, we discuss the capability of the different measures to discriminate between storm and quiescence periods

with respect to a heuristic global-scale and a data-adaptive local-scale classification based on the Dst index in Secs. IV and V, respectively. Implications of our results are addressed in Sec. VI.

II. DATA AND METHODS

A. Description of the data

The previous solar cycle 23 (May 1996 to January 2008) peaked in 2000–2003 with many strong solar events, which in turn caused extended periods of strong magnetospheric activity. In particular, the year 2001 saw the occurrence of two intense magnetic storms on 31 March 2001 and 6 November 2001 (when Dst reached minimum values of -387 nT and -292 nT, respectively), which were associated with two large coronal mass ejections (CMEs) on 29 March 2001 and 4 November 2001, respectively.

Figure 1 shows the corresponding Dst time series. It can be seen that the data can be rather naturally divided into five shorter time series (I-V in Fig. 1). The second and fourth time windows (II and IV in Fig. 1, highlighted by gray shading) include the Dst variations related to the periods around the aforementioned intense magnetic storms of March and November 2001, respectively. The same time intervals were previously shown to be compatible with the emergence of two distinct patterns in the Earth's magnetosphere at time scales between hours and a few weeks (i.e., which are relatively long in comparison to the short-term bursty dynamics of the magnetosphere): (i) a pattern associated with intense magnetic storms ("pathological" states of the magnetosphere) in windows II and IV, which have been characterized by a higher degree of organization, and (ii) a pattern associated with "normal" (non-storm) periods ("physiological" states of the magnetosphere) in windows I, III, and V, which are characterized by a lower degree of organization, i.e., more random fluctuations.[19,25,26,33,34]

We note that this heuristic global distinction into storm and non-storm periods does not lead to a clear separation

FIG. 2. Probability density functions (Gaussian kernel estimates) of hourly Dst values during storm (gray area) and quiescence (white area) periods as considered in this paper (see text).

between the Dst values recorded within the different periods. As Fig. 2 shows, there is a considerable overlap between the probability density functions (PDFs) of hourly Dst values observed during both types of periods, which is mainly due to the fact that the actual individual storms (strongly negative Dst values) have a much shorter duration than the considered overall storm periods (note that geomagnetic storms typically appear clustered in time) so that there are lots of close to zero Dst values even during storm periods. In turn, there have also been some weaker storms during the supposed non-storm periods. This observation does not contradict our paradigm of separating the time series under study into distinct periods, since the dynamical characteristics to be studied will be computed for running windows instead of individual points in time (see below). In this respect, we still expect that these characteristics (relating to time-scales from days to weeks) exhibit marked differences between time periods prone to severe magnetic storms and those with relative quiescence of the magnetosphere.

B. Recurrence analysis

The methods used in this work are based on the concept of phase space in deterministic dynamical systems, which is commonly spanned by a multitude of complementary variables that—taken together—uniquely determine the system's state. In our case, however, we have access to only a single variable (Dst), the dynamics of which contains contributions due to both the "internal" dynamics of the geomagnetic field and the external forcing by the solar wind and related extraterrestrial processes, which are mutually entangled in some complex, nonlinear way. In order to qualitatively reconstruct unobserved variables from this time series, we employ time-delay embedding[35,36] and construct a multivariate time series $X(t)$ from the original data $x(t)$ by considering appropriate time-shifts,
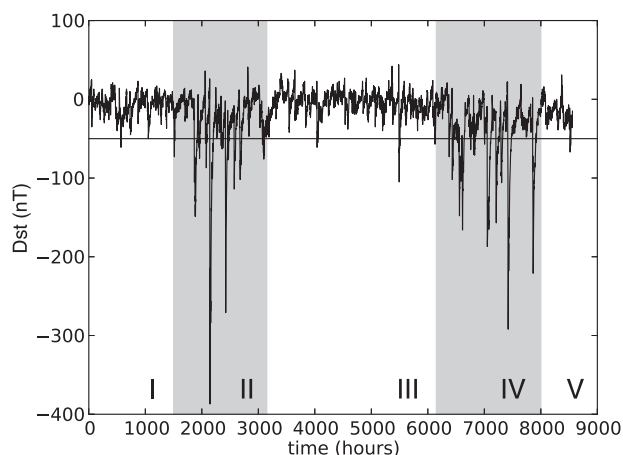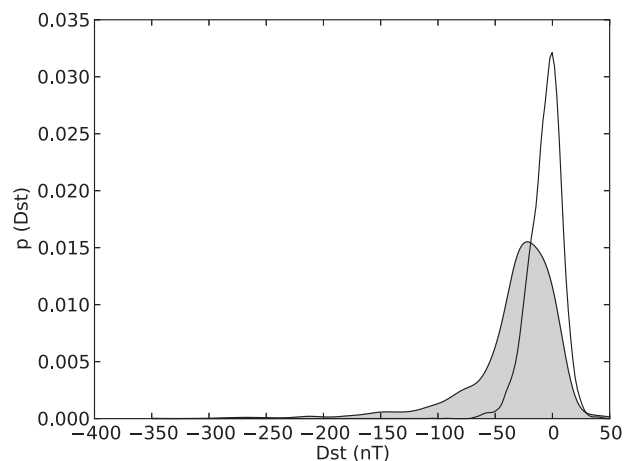

FIG. 1. Dst index time series for the period from 1 January 2001 to 31 January 2002. Gray shaded areas highlight the two time intervals (II, IV) heuristically classified as storm periods, whereas the remaining periods (I, III, V) are considered as episodes of quiescence. The horizontal line depicts a value of Dst $= -50$ nT, which is commonly considered as a criterion for defining a magnetic storm.

$$X(t) = \{x(t), x(t-\tau), x(t-2\tau), \ldots, x[t-(m-1)\tau]\}$$
$$= [X^{(1)}(t), \ldots, X^{(m)}(t)]. \tag{1}$$

To make the individual components of the embedding vector approximately independent, we choose an embedding delay of $\tau = 100$ h to account for the relatively slow decay of auto-correlations of Dst during storm periods. The embedding dimension is set to $m = 3$ to cope with the necessity of studying short time series segments individually (see below). Further discussion on the choice of the embedding parameters is provided in the supplementary material accompanying this paper.

Recurrence plots have been originally introduced as a simple means of visualizing the temporal succession of close states in phase space.[21,37] Given a time series containing the relevant system variables (either from multivariate observations or time-delay embedding of univariate signals), we define the recurrence matrix $\mathbf{R} = (R_{ij})$ as a binary matrix encoding whether or not each pair of state vectors on the sampled trajectory is mutually close. One possible formulation, which we will adopt in this work, is making this decision by comparing all pairwise distances with some threshold distance ε,

$$R_{ij}(\varepsilon) = \Theta(\varepsilon - \|X(t_i) - X(t_j)\|_\infty), \qquad (2)$$

where $\Theta(\cdot)$ denotes the Heaviside function and $\|\cdot\|_\infty$ the maximum norm defined as

$$\|X(t_i) - X(t_j)\|_\infty = \max_k \left| X^{(k)}(t_i) - X^{(k)}(t_j) \right|. \qquad (3)$$

In Fig. 3, we display the recurrence plot of the Dst data (i.e., the visual representation of the binary recurrence matrix $\mathbf{R}$). For illustrative purposes, we use here a recurrence threshold ε chosen such as to yield a recurrence rate of $RR = 0.05$, i.e., the 5% closest pairs of state vectors are considered as recurrences.[38] This value represents a typical choice in many applications, while different authors have provided different—sometimes conflicting—rules of thumb for selecting the recurrence rate based on different types of considerations.[21,39]
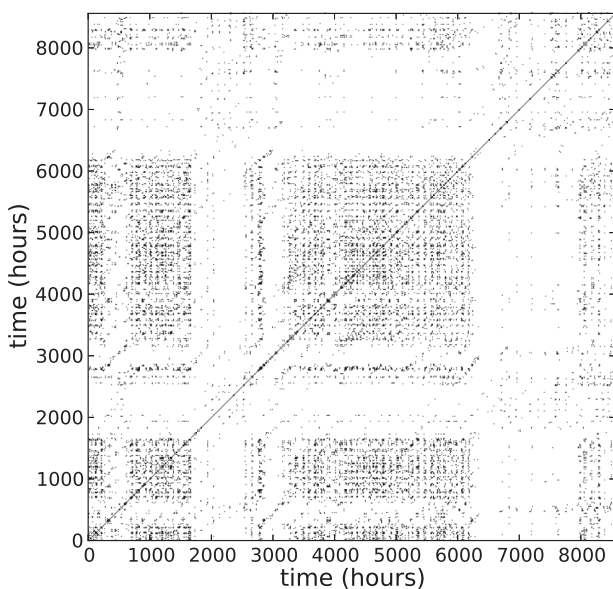

FIG. 3. Recurrence plot for the Dst index time series computed with the embedding parameters $m = 3$ and $\tau = 100$, using a global recurrence rate of $RR = 0.05$.

Looking at the resulting recurrence plot, one immediately recognizes that the density of recurrences is much higher during non-storm periods than during storm times, resulting in some distinct block structure of the recurrence plot. This is mainly due to the large negative Dst amplitudes associated with magnetic storms. Specifically, when considering the same absolute differences in Dst values for defining the closeness of two values during storm and quiescence periods, it is easy to see that storm conditions (characterized by a larger variance of Dst values than periods without magnetic storms) tend to exhibit a lower number of recurrences.

In order to unveil changes in the dynamical characteristics of the Dst index during storm and non-storm periods, in the following we will consider sliding windows in time (with width w and mutual offset Δw, see below). For this purpose, we will fix the recurrence rate $RR = 0.05$ and compute recurrence plots individually for each time window with this value of RR, thereby allowing for some quantitative comparison between properties calculated for different time slices.

For characterizing the dynamical complexity as encoded in the resulting recurrence structures and its temporal variability, a variety of statistical characteristics can be estimated. Here, we restrict our attention to representative measures from two particularly useful frameworks, recurrence quantification analysis (RQA)[21,22] and recurrence network analysis (RNA).[24,40,41] RQA is based on the idea that deterministic dynamics gives rise to the emergence of line structures in recurrence plots. Notably, diagonal lines in the recurrence plot formed by recurrent states only (i.e., off-diagonal structures in the recurrence matrix containing exclusively values of one) correspond to a similar evolution of the system within two time slices.[42,43] The longer such diagonal lines, the more predictable is the observed dynamics. In a similar spirit, vertical lines (and, due to the symmetry of the recurrence plot, also horizontal lines) in the recurrence plot emerge whenever the system's state does not change much within a certain time window.[44] The length-frequency distributions of such lines therefore encode fundamental dynamical characteristics. Here, we statistically characterize these properties by means of three selected measures:

- the degree of *determinism*,

$$DET = \frac{\sum_{d=d_{min}}^{d_{max}} d \, p(d)}{\sum_{d=1}^{d_{max}} d \, p(d)}, \qquad (4)$$

- the degree of *laminarity*,

$$LAM = \frac{\sum_{v=v_{min}}^{v_{max}} v \, p(v)}{\sum_{v=1}^{v_{max}} v \, p(v)}, \qquad (5)$$

and
- the mean length of vertical lines referred to as the *trapping time*,

$$TT = \frac{\sum_{v=v_{min}}^{v_{max}} v \, p(v)}{\sum_{v=v_{min}}^{v_{max}} p(v)}. \qquad (6)$$

Here, d and v denote the lengths of diagonal and vertical lines in the recurrence plot, respectively, $p(d)$ and $p(v)$ the associated relative frequencies, and d_{max} and v_{max} the maximal diagonal and vertical line lengths, respectively (excluding the

main diagonal in the plot). In what follows, we will consider a lower bound of $d_{min} = v_{min} = 2$ to define line structures. In order to highlight the presence of episodes of extraordinary dynamical characteristics, we adopt a randomization procedure,[45] which consists of determining the expected distribution of RQA measures from bootstrapping the respective line length distribution from the recurrence plot obtained from the entire time series (see the supplementary material for details).

As a complementary analysis framework, we also consider some characteristics of RNA, for we identify each state vector with a node and define the adjacency (connectivity) matrix of the associated graph as

$$A_{ij}(\varepsilon) = R_{ij}(\varepsilon) - \delta_{ij}, \qquad (7)$$

where δ_{ij} denotes the Kronecker delta. According to our previous experience in studying dynamical transitions from paradigmatic model systems as well as real-world geoscientific time series,[30,31,40,46,47] we select the following three quantifiers that characterize the resulting recurrence networks on a global scale:

- the *network transitivity*[48]

$$\mathcal{T} = \frac{\sum_{i,j,k} A_{ij} A_{ik} A_{jk}}{\sum_{i,j,k} A_{ij} A_{ik}}, \qquad (8)$$

- the *global clustering* coefficient

$$\mathcal{C} = \frac{1}{N} \sum_i \frac{\sum_{j,k} A_{ij} A_{ik} A_{jk}}{\sum_{j,k} A_{ij} A_{ik}}, \qquad (9)$$

and

- the *average path length* $\mathcal{L}$, i.e., the mean minimum number of edges separating all pairs of nodes in the recurrence network from each other.

Similar as for the RQA measures, we are interested in whether or not the aforementioned RNA characteristics deviate significantly from the properties of a recurrence network sampled from state vectors belonging to arbitrary points in time. This is achieved by randomly drawing individual samples from the full set of state vectors corresponding to the entire Dst time series and creating recurrence networks from these random samples with the same overall recurrence rate as in our windowed analysis.[30]

III. RESULTS

A. Temporal variations of recurrence characteristics

Figure 4 shows the variability of the six selected recurrence properties for a representative window size of $w = 256$ h (i.e., about 10 days, capturing the essential time scales of magnetic storms[34,49]). According to the relatively large embedding delay, this implies that we effectively take information from Dst values into account that spread over a considerably larger time window of $w + (m-1)\tau = 456$ h. By doing so, we utilize variations on time scales between hours and a few weeks for our analysis, disregarding any higher or lower frequency variability. As shown in the supplementary material, our results discussed in the following do not change

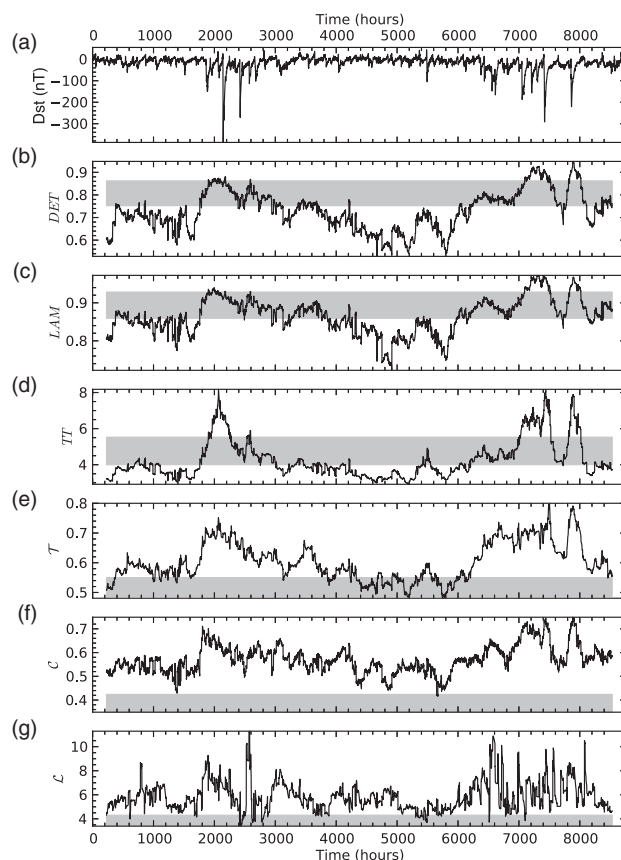FIG. 4. Dst time series (a) and RQA and RNA measures for the year 2001: (b) determinism *DET*, (c) laminarity *LAM*, (d) trapping time *TT*, (e) transitivity $\mathcal{T}$, (f) global clustering coefficient $\mathcal{C}$, (g) average path length $\mathcal{L}$ (computed for running windows of size $w = 256$ h and mutual offset $\Delta w = 1$ h). Gray shades indicate the expected 90% range provided by the bootstrapping approaches introduced in Sec. II. Note that the tests for RQA and RNA measures are based on different rationales, so the outcomes have essentially different interpretations (see the supplementary material for some more detailed discussion).

qualitatively under different choices of the recurrence rate and embedding delay, respectively.

As can be seen, both RQA and RNA characteristics exhibit marked temporal variability corresponding to the alternation between storm and non-storm periods. Specifically, during storm periods, all six considered measures take distinct maxima. For the RQA parameters *DET*, *LAM* and *TT* as well as the RNA measures $\mathcal{T}$ and $\mathcal{C}$, this is consistent with a higher degree of organization in the system as proposed earlier,[19] leading to a more regular (and thus more predictable) dynamics. In the Dst index, this behavior is typically reflected by gradual trends corresponding to the emergence and subsequent disappearance of the magnetic field perturbation over a time scale comparable to the considered window size.[34] In turn, the maxima of $\mathcal{L}$ during storm periods are rather indicative of persistent qualitative changes of dynamical complexity, which trace the succession of individual storms interrupted by short periods of quiescence in magnetospheric fluctuations during these activity phases.

With the exception of $\mathcal{L}$, the considered measures characterize related but conceptually different aspects of dynamical complexity of the geomagnetic field fluctuations captured by Dst. In particular, *DET* and *LAM* show an extremely strong

similarity. This indicates that in the Dst index, periods characterized by a higher degree of dynamical organization (i.e., "smoother" variability with respect to the full temporary range of states, commonly emerging during storm periods) coincide with a tendency towards more laminar dynamics (i.e., values "not changing much" over certain periods of time). The latter is probably related to generally weaker fluctuations superimposed to the gradual Dst trends during magnetic storms, given that the adaptive recurrence threshold ε follows the general variance of Dst within the considered time intervals (not shown). In other words, we observe a separation of "normal" background fluctuations from larger-scale temporary trends in the Dst index, and both *DET* and *LAM* appear to trace primarily the variation of the level of detail provided by our methods.

In a similar way, we observe that *TT* and $\mathcal{T}$ present closely related temporal variability profiles. We interpret this observation as both characteristics being associated with the regularity of fluctuations, one (*TT*) capturing this property from the dynamical, the other ($\mathcal{T}$) from a geometric perspective. We particularly observe maxima of both measures during storm periods, which are more distinct than in case of *DET* and *LAM*, suggesting that *TT* and $\mathcal{T}$ are more sensitive tracers of dynamical changes in the magnetospheric variability during storm phases than *DET* or *LAM*.

Moreover, we find that $\mathcal{T}$ and $\mathcal{C}$ also present a similar evolution, with $\mathcal{C}$ exhibiting less distinct maxima. The latter behavior is not uncommon in applications of RNA, since both properties are based on similar considerations. However, by construction $\mathcal{C}$ exhibits some sensitivity to changes in the heterogeneity of connectivity distributions in the recurrence networks, whereas $\mathcal{T}$ does not. Recalling the close relationship of $\mathcal{T}$ with a generalized notion of fractal dimension,[50] it is not surprising that the transitivity presents some more distinct variability profile sensitively tracing changes in the degree of regularity of the observed dynamics.

Notably, in comparison with the other five measures, $\mathcal{L}$ shows a less clear yet still visible signature of transitions between storm and non-storm periods. As mentioned above, we relate this observation to the fact that this measure primarily traces the succession of different dynamical "states" (individual magnetic storms interrupted by short periods of quiescence) during storm periods consisting of several events, while the remaining two RNA measures as well as the

three RQA characteristics quantify the overall complexity of fluctuations.

The aforementioned interdependences are confirmed and further quantified by Table I displaying Pearson's correlation coefficients[51] between the individual measures taken over the respective running windows. For the sake of simplicity, we restrict this analysis to the consideration of linear correlations between the different measures, ignoring possibly more complex (nonlinear) interrelations as well as effects due to strong deviations of the associated probability distribution functions from Gaussianity.

B. Amplitude-complexity relationship

The qualitative assessment provided above points to a close relationship between the amplitude of the Dst index and the complexity of its temporal variations. Notably, this aspect goes beyond classical mean-variance interrelationships often observed in heteroscedastic signals (for which Dst might be a good example) but describes a specific feature of magnetospheric dynamics. In the following, we further quantify these interrelationships between recurrence characteristics and the window-wise mean Dst index.

As already expected from Fig. 4, Table I reveals a general anti-correlation between all measures and the mean Dst index in each window, which underlines the fact that maxima of the considered recurrence characteristics commonly coincide with magnetic storms (strongly negative Dst values). Here, we have shifted the time axis of the recurrence characteristics such that each window indexed by some value of t represents the embedding vectors $X(t + \tau), \ldots, X(t + \tau + w)$, i.e., effectively takes into account information from Dst values from the window $[t - \tau, t + \tau + w]$. For a discussion of this choice, see the supplementary material.

Among the six considered recurrence characteristics, $\mathcal{T}$ and *TT* generally exhibit the strongest linear (anti-) correlations with Dst, followed by *DET*, *LAM*, and $\mathcal{C}$, whereas $\mathcal{L}$ shows significantly weaker correlations with the Dst index values. This finding allows formulating a preliminary hypothesis regarding the appropriateness of different recurrence parameters for tracing signatures of magnetic storms. Specifically, we expect that $\mathcal{T}$ and *TT* better discriminate the dynamical state of the magnetosphere during storm and non-storm periods than the other measures. We will further investigate this hypothesis in Secs. IV and V. Note again that this simple linear correlation analysis disregards any effects due to non-normality and possible nonlinearity of the relationships between the different recurrence properties and Dst.

C. Estimated vs. baseline recurrence characteristics

The intervals obtained from the bootstrapping approaches (gray bars in Fig. 4—corresponding to the upper and lower 5% percentiles of the test distribution) display the expected behavior (see our corresponding discussion in the supplementary material). Specifically, for all three RNA measures, the values within the individual windows are mostly above the upper 5% quantiles of the distributions obtained by bootstrapping. This indicates that in comparison with the window-wise recurrence networks, bootstrapping state

TABLE I. Pearson's correlation coefficients among the different recurrence parameters and between these characteristics and the window-wise mean Dst index for a window size of $w = 256$ h, offset $\Delta w = 1$ h. The last column presents the area under the ROC curve (AUC) indicating the performance of the six recurrence measures as discriminators between the (heuristically classified) storm and non-storm periods (see Sec. IV).

r^2	*LAM*	*TT*	$\mathcal{T}$	$\mathcal{C}$	$\mathcal{L}$	$\langle$Dst$\rangle$	AUC
DET	0.97	0.90	0.89	0.85	0.50	−0.76	0.956
LAM	—	0.85	0.86	0.87	0.47	−0.72	0.938
TT	—	—	0.87	0.80	0.49	−0.80	0.940
$\mathcal{T}$	—	—	—	0.82	0.58	−0.81	0.959
$\mathcal{C}$	—	—	—	—	0.53	−0.70	0.918
$\mathcal{L}$	—	—	—	—	—	−0.53	0.806

vectors from the entire time series introduces a considerable degree of randomness to the resulting networks, resulting in lower values of $\mathcal{T}$ and $\mathcal{C}$. This is because state vectors representing essentially different states of the system are mixed in the considered bootstrap samples. Since such states typically persist in time (cf. the time scales of several days at which the ACF decays, see the supplementary material), one has to expect a much higher degree of regularity in the networks obtained within running windows, expressed by higher values of both $\mathcal{T}$ and $\mathcal{C}$. In turn, if understanding the individual bootstrap samples as representatives of some stationary process that covers the full variety of dynamical patterns of Dst index variations, our results would be compatible with an almost constantly non-stationary (i.e., out-of-equilibrium) system, which is consistent with the modern view on the magnetosphere as described in the Introduction of this paper. In this context, it is interesting to note that the distributions of RNA measures from the bootstrapped samples are considerably more narrow than the ranges of values obtained for the sliding windows, which is particularly well visible for $\mathcal{L}$.

Unlike for the RNA measures, the window-wise RQA characteristics scatter more symmetrically around the values obtained from the bootstrapped line length distributions, displaying both higher and lower values as time proceeds. Notably, positive "anomalies" exclusively arise during periods with strong magnetic storms, whereas values below the lower 5% quantile of the line bootstrapping-based distributions are found only during periods without strong perturbations of the magnetosphere where the recorded Dst variations are dominated by short-term fluctuations. This general observation again underlines the fundamentally different conceptual foundations of the randomization procedures used for obtaining significance bounds for possible indications of regime changes in RQA and RNA.

IV. DISCRIMINATION BETWEEN STORM AND NON-STORM PERIODS

In the following, we investigate in some detail how well the values of the different recurrence characteristics discriminate between storm and non-storm periods. For this purpose, we employ the heuristic classification already used in Sec. II A. Specifically, for running windows of width $w = 256$ h and mutual offset $\Delta w = 1$ h, we consider the windows with midpoints between 8 March and 16 May 2001 and 17 September and 4 December, respectively, as storm periods, the others as non-storm periods.

A. Recurrence characteristics

Figure 5 presents the PDFs of all six characteristics estimated separately for both types of conditions. As for the hourly Dst index itself (Fig. 2), we observe considerable overlaps between the empirical distributions for both cases, but the mean values appear clearly separated given the respective variances within the two subsamples. In order to characterize the significance of the mutual differences between the distributions for storm and non-storm periods, a variety of statistical approaches could be used (see Ref. 46 for a corresponding discussion and one example regarding RQA and
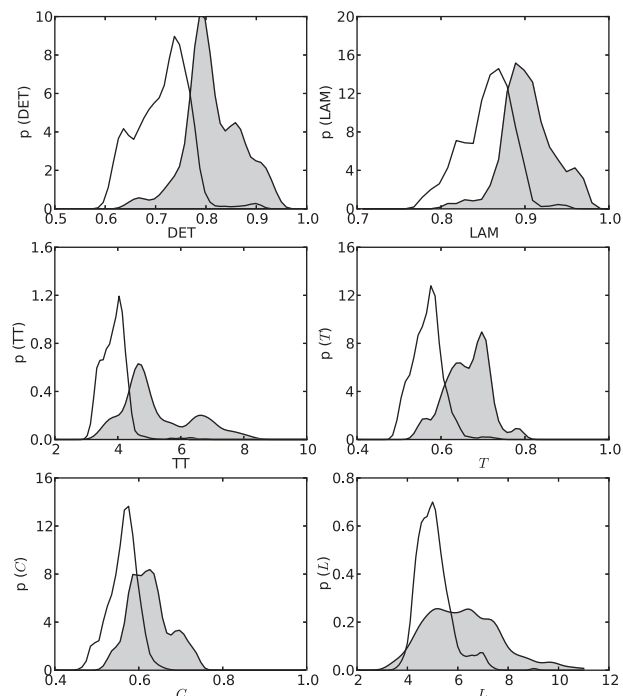

FIG. 5. Probability density functions (Gaussian kernel estimates) of the different RQA and RNA measures (obtained for running windows of width $w = 256$ h and mutual offset $\Delta w = 1$ h) during storm (gray area) and non-storm (white area) periods (heuristic classification as detailed in the text).

RNA for discriminating between intertwined periodic and chaotic windows in a paradigmatic model system). Regarding a classical one-way ANOVA (analysis of variance, quantifying the difference between the mean values of two samples given their respective variances) approach,[51] all six recurrence measures clearly do not pass the F-test for equality of the means and the Wilcoxon rank-sum test[52] for equality of the medians at very high confidence levels, indicating that the distributions of all RQA and RNA measures for storm and non-storm periods are significantly different.

One particularly instructive approach for quantitatively studying the discriminative skills of statistical methods based on a given data classification is the so-called receiver-operating characteristics (ROC) analysis.[53] Here, a variable threshold is applied to each of the six measures, and for each threshold value, the classification of values (above/below the threshold) is compared with the given classification of the data into storm and non-storm periods. The rates of true and false "positive" classifications (TPR and FPR, respectively) of a storm period based on each RQA or RNA measure are continuously monitored and provide a characteristic closed curve (the ROC curve) in the (FPR,TPR) plane.

Figure 6 shows the ROC curves of all six measures. The closer the ROC curve lies to the upper left corner or, more precisely, the stronger it deviates from the main diagonal TPR = FPR corresponding to a random "prediction," the better are the skills of the respective measure to discriminate between both types of (longer-term) magnetospheric "states." Consequently, this skill can be measured in terms of the area under the ROC curve, AUC. We find that $\mathcal{T}$ generally provides the highest AUC values, followed by *DET*, *TT*, and *LAM*, all of
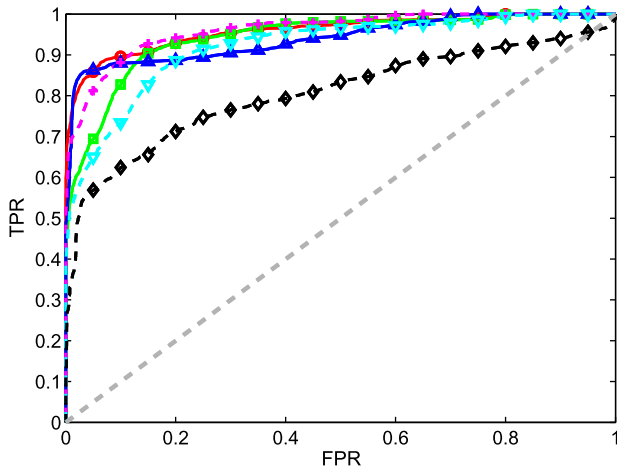
FIG. 6. Receiver-operating characteristics (ROC) curves for the RQA and RNA measures using windows of 256 h and mutual offset of $\Delta w = 1$ h, given the heuristic splitting of the Dst time series into five storm and non-storm periods (see text) as a reference. The dashed diagonal line indicates the expected performance of a random discrimination (TPR = FPR). Solid lines: RQA measures *DET* (red, ○), *LAM* (green, □), and *TT* (blue, △); dashed lines: RNA measures $\mathcal{T}$ (magenta, +), $\mathcal{C}$ (cyan, ▽), and $\mathcal{L}$ (black, ◇).

which still provide a very good discrimination. To a certain degree, the AUC values of $\mathcal{C}$ also suggest a reasonable distinction between storm and non-storm periods, whereas the results for $\mathcal{L}$ are not satisfactory in this respect.

Notably, the ranking of the six recurrence measures according to AUC (cf. also Table I, last column) differs somewhat from that based upon the correlation analysis in Sec. III. Specifically, $\mathcal{T}$ still performs best among all characteristics, closely followed by *DET*, whereas *TT* provides lower AUC values than expected. This discrepancy can be explained by the fact that we have considered here a rather coarse-grained classification between storm and non-storm periods, whereas correlations directly rely on the explicit window-wise mean Dst values, which can be the same for a considerable number of windows as illustrated in Fig. 5. We will return to this aspect in Sec. V.

B. Comparison with other methods

In previous works, changes in the dynamical properties of the Dst index have been extensively studied and related to different levels of organization of magnetospheric dynamics during storms and non-storm periods. In the following, we will briefly review the main corresponding findings and compare them with the results of our recurrence-based analysis.

In Refs. 19 and 54, temporal changes in the persistence of the Dst data have been studied in terms of the associated Hurst exponent H estimated by some fractal wavelet spectral approach and rescaled-range analysis, respectively. Both studies reported a marked increase in the estimated exponents during storm periods, pointing to some large-scale and long-term organization of magnetospheric fluctuations. The corresponding results were qualitatively confirmed in a recent study[34] utilizing a simple complexity measure [the linear variance decay (LVD) dimension density δ_{LVD}] based on the auto-correlation function of the data under study.

References 25, 26, and 33 applied the classical concept of Shannon entropy S in conjunction with Tsallis' non-extensive entropy S_q as well as some other entropy measures in order to characterize the degree of dynamical disorder. Their results consistently revealed a marked decrease of entropic characteristics during magnetic storms, pointing again towards an elevated degree of dynamical organization and associated determinism in the magnetosphere.

In two recent papers,[26,55] a suite of complementary entropy concepts have been utilized, which are not solely rooted in information theory and statistical mechanics (like Shannon or Tsallis entropy) but make use of distances among vectors in the system's reconstructed phase space. Specifically, the approximate entropy (AppEnt),[56] sample entropy (SampEnt),[57] and fuzzy entropy (FuzzyEnt)[58] used in the aforementioned works are based on a conceptual foundation closely related to that of the recurrence characteristics employed in this work (see Ref. 55 for more details on the respective methods and their relation to recurrence analysis). In this spirit, when using these measures we may expect results that are generally comparable with those of recurrence analysis as discussed above.

In the following, we provide an in-depth analysis on how well different previously considered dynamical characteristics discriminate between the state of the magnetosphere during storm and non-storm periods based on the Dst index data. For this purpose, we apply the measures listed below with the following parameters:

- The Hurst exponent H is estimated from the slopes of the wavelet power spectral densities in the frequency range between 2 and 128 h.[19]
- The LVD dimension density δ_{LVD}[34] is computed using $N = 100$ embedding components mutually shifted by $\tau = 1$ h and a threshold for the fraction f of explained variance at $f = 0.95$.
- The Shannon entropy S is calculated for a binary partition (with the mean value as threshold) of the underlying data. The block (H_n) and Tsallis entropies (S_q) are computed for the same partition based on a block (word) length of $n = 2$ (for details, see Ref. 26). The value of the Tsallis q index utilized for the calculation of non-extensive Tsallis entropy $S_q(q)$ is selected to be 1.8, as indicated in Ref. 25 and more recently in Ref. 59.
- The Kolmogorov entropy h is estimated from the linear scaling behavior of the block entropy per symbol, H_n/n for larger n.
- The T-complexity T is calculated using $n = 2$ and $m = 1$, respectively (for details, see Ref. 26).
- The approximate, sample and fuzzy entropies are computed using embedding dimension $m = 2$, $\tau = 1$ h and a distance threshold $\varepsilon = 0.65\sigma$ (with σ being the standard deviation of the data within the considered time window) to comply with the setting of previous work.[55] For FuzzyEnt, the fuzzy membership function described in Ref. 55 is utilized with $n = 2$.

For brevity, we present only the results for a window width of $w = 256$ h and a mutual offset of $\Delta w = 1$ h. A summary of the temporal variations of all measures is provided in Fig. 7. Note
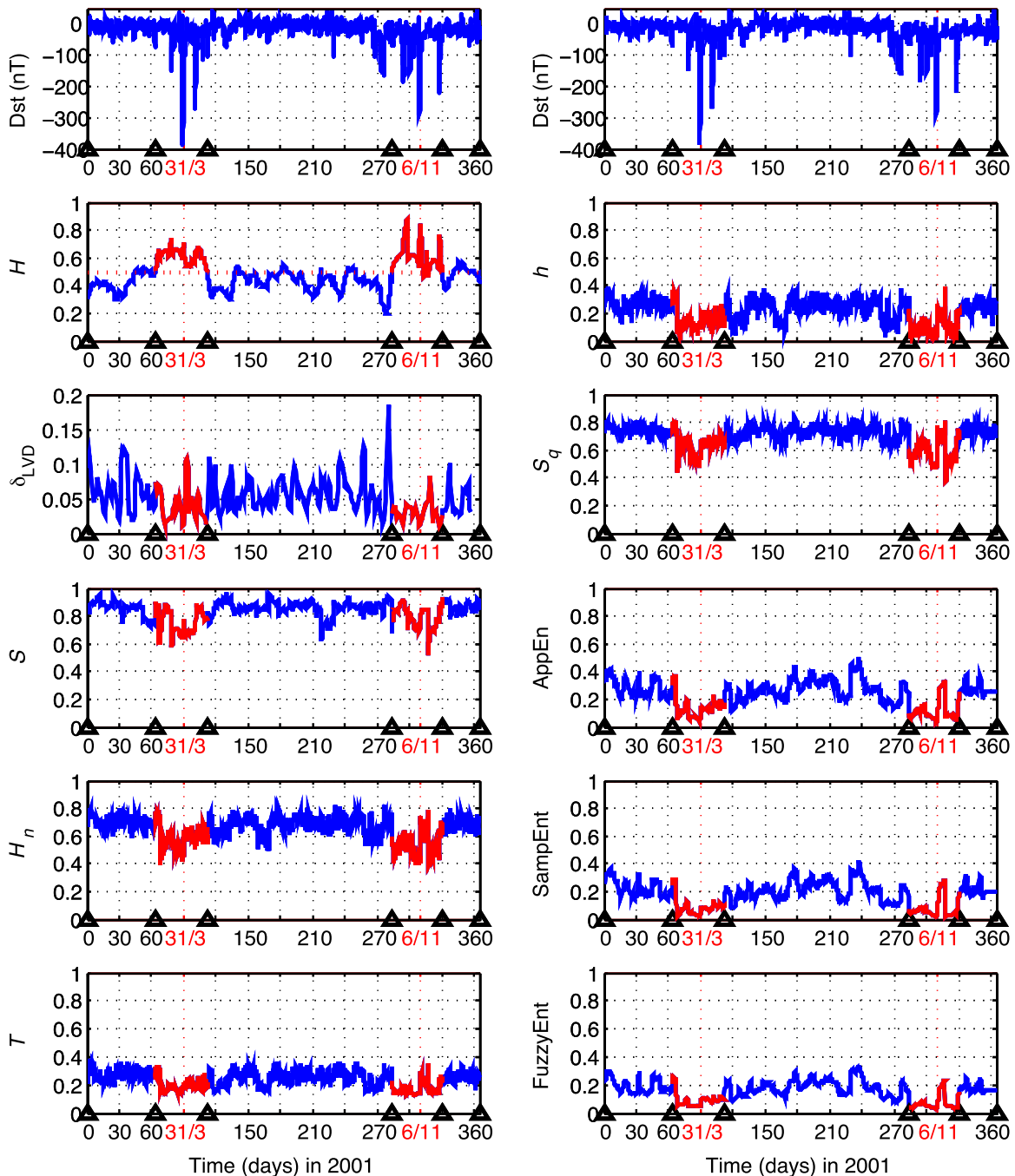
FIG. 7. Time evolution of Dst index (top panels) and its entropy and correlation-based characteristics: Hurst exponent H, LVD dimension density δ_{LVD}, Shannon entropy S, block entropy H_n, T-complexity T (left panels, from top to bottom), Kolmogorov entropy h, Tsallis entropy S_q, AppEnt, SampEnt, and FuzzyEnt (right panels, from top to bottom). Further details on the respective methodological settings are provided in the text. Red colors mark the storm periods II and IV from Fig. 1. Dotted red vertical lines indicate the timing of the two major storms during the considered interval of observations. For the Hurst exponent H, the dashed horizontal line gives the value of 0.5 discriminating between persistent and anti-persistent dynamics.

that as for the recurrence measures (Sec. III) and the LVD dimension density,[34] the performance of all characteristics in discriminating between storm and non-storm periods depends on the window width as well as the definition of the storm period. However, we will not discuss this aspect further in the present work.

Our results shown in Fig. 8 reveal that the dynamic entropies (block, Tsallis, and Kolmogorov entropy, T-complexity as well as AppEnt, SampEnt, and FuzzyEnt) and the LVD dimension density provide a reasonable

discrimination with AUC values of 0.84 and higher, whereas the purely statistical Shannon entropy and the Hurst exponent exhibit significantly poorer performance. The three phase space-based entropies yield the highest AUC values among all measures, which are, however, still markedly lower than those for the four best-suited recurrence characteristics. One possible reason for this is that the parameters of the recurrence analysis (especially the embedding parameters) have been adjusted to the specific data set, whereas the "embedding" dimension and delay have been chosen at the lowest
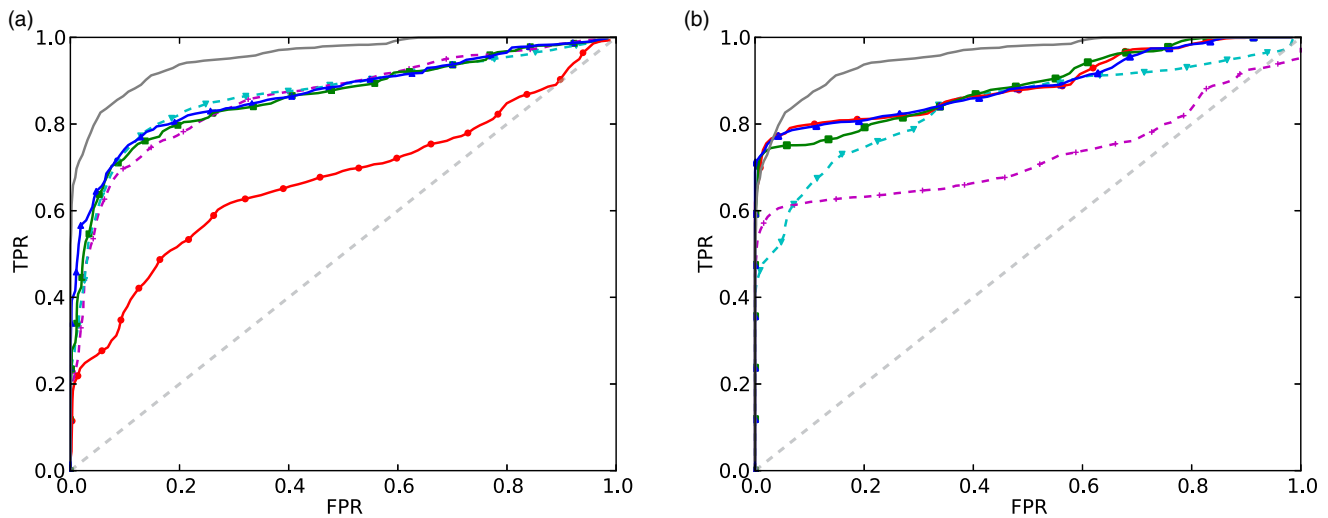
FIG. 8. ROC curves for other measures recently considered in the literature ($w = 256$ h, $\Delta w = 1$ h). (a) Symbolic dynamics based characteristics Shannon entropy S (red, ○, solid, AUC = 0.670), block entropy H_n (green, □, solid, AUC = 0.862), Tsallis entropy S_q (blue, △, solid, AUC = 0.871), T-complexity T (magenta, +, dashed, AUC = 0.860), and Kolmogorov entropy h (cyan, ▽, dashed, AUC = 0.867). (b) Phase space-based entropic quantities AppEnt (red, ○, solid, AUC = 0.893), SampEnt (green, □, solid, AUC = 0.890), and FuzzyEnt (blue, △, solid, AUC = 0.892) together with the correlation-based measures Hurst exponent H (magenta, +, dashed, AUC = 0.730) and LVD dimension density δ_{LVD} (cyan, ▽, dashed, AUC = 0.845). For reference, the corresponding ROC curve for the recurrence network transitivity $\mathcal{T}$ is displayed as well (solid gray line).

possible values for the entropy estimations. Notably, the latter setting might not allow resolving all essential time scales of variations associated with magnetospheric dynamics (unlike for the choice of embedding parameters used for RQA and RNA analysis). By further tuning the parameters for AppEnt, SampEnt, and FuzzyEnt, a systematic improvement of the skills of these methods could be achieved. As a somewhat surprising result, we emphasize that the correlation-based LVD dimension density performs not much worse than the dynamic entropies (even though it is based on some specific statistical model of the correlations that does not necessarily provide a reasonable approximation of the actual data properties). Reversing the aforementioned argument regarding the phase space-based entropies, this finding could be related to the relatively large number of embedding components used in estimating this characteristic.

V. DISCRIMINATION USING A DATA-ADAPTIVE CLASSIFICATION

Instead of the previous fixed and thus rather inflexible classification of storm and non-storm periods, in the following we turn to some more data adaptive strategy by defining for each running window a storm as being present if the mean Dst value taken over the window is below some threshold value $\langle Dst \rangle^*$. Unless stated otherwise, we will commonly consider $\langle Dst \rangle^* = -30$ nT. The choice of this threshold value will be justified in the course of the analysis described in this section.

A. Recurrence based characteristics

Figures 9(a) and 9(b) show the ROC curves of all six recurrence measures for two different choices of the window width w. In addition, we present the dependence of AUC on the discrimination threshold $\langle Dst \rangle^*$ and running window width w, respectively, in Figs. 9(c) and 9(d).

When using the same window length of $w = 256$ h as in Secs. III and IV, the comparative performance of the different characteristics follows the expectations based on the results of the correlation analysis in Sec. III. That is, $\mathcal{T}$, TT, and DET generally provide the best discrimination between storm and non-storm conditions, followed by LAM, $\mathcal{C}$, and $\mathcal{L}$ [Fig. 9(b)]. More specifically, with the variable threshold, TT competes with $\mathcal{T}$ and DET again, i.e., the performance characteristics are closer to the findings from correlation analysis than when using the heuristic coarse-grained classification into storm and non-storm periods from Sec. IV. This difference is most likely caused by the dynamical properties of time windows with Dst values close to zero during the previously considered storm periods. Given the relatively high fraction of time windows with such conditions (see Fig. 2), it is not unexpected that the six recurrence measures actually behave somewhat differently especially during such time intervals where the heuristic and data-adaptive classifications do not agree with each other. The potential relevance of this result—particularly regarding the problem of anticipating or early detecting approaching magnetospheric disturbances from Dst index data—could offer an interesting avenue for further research. However, for the latter purpose, additional information on solar wind parameters needs to be considered as well,[32] bearing in mind that the magnetosphere is a driven system heavily affected by extraterrestrial forcing.

If we turn towards shorter windows [e.g., Fig. 9(a)], the general picture does not change much but exhibits some interesting details: For shorter windows, the discrimination between storm and non-storm conditions becomes gradually worse for all measures, which is to be expected, since a lower amount of data is available for computing the measures the classification is based on. In turn, for longer windows, the classification skills of all recurrence parameters saturate at a window size between about 250 and 350 h, depending on the
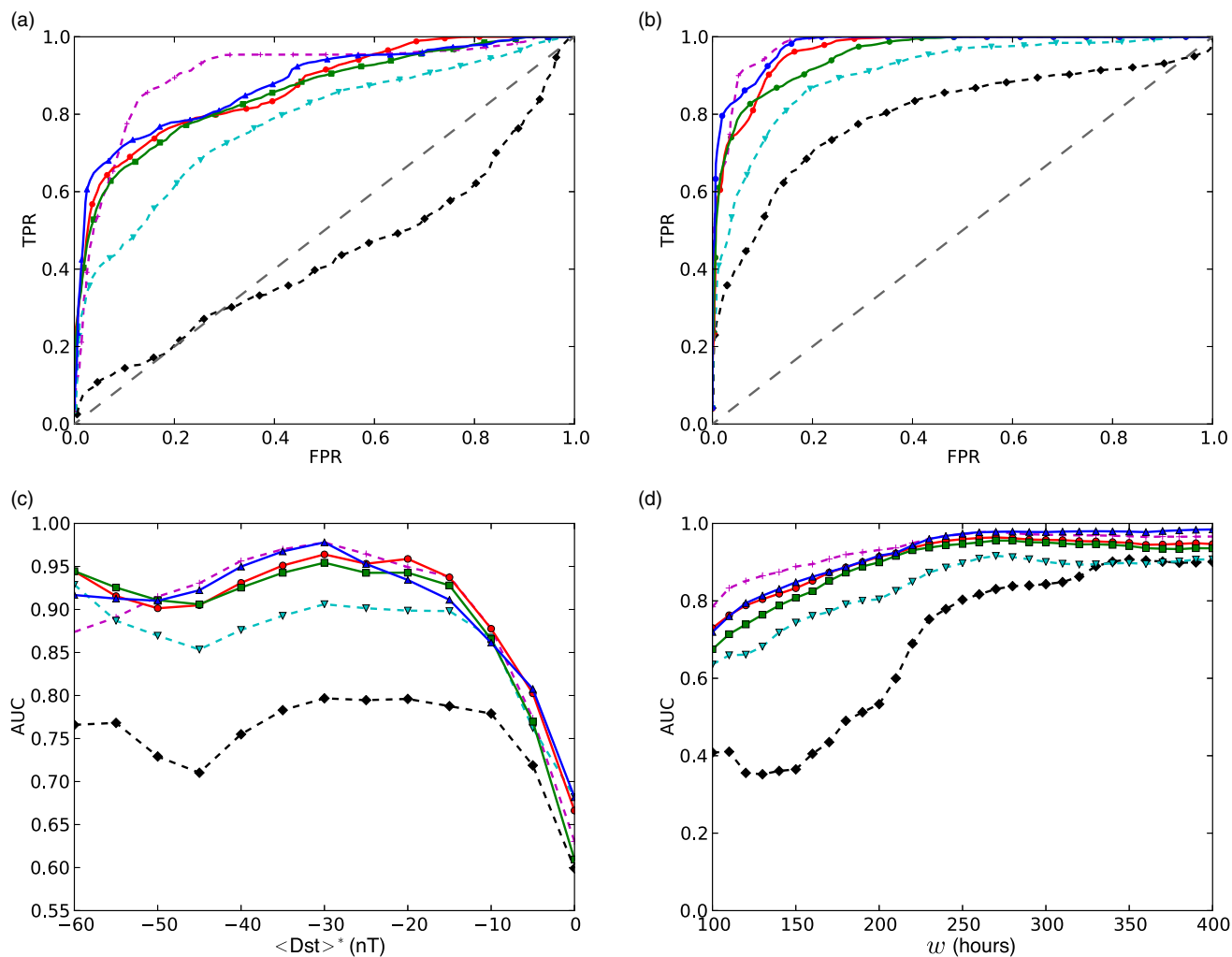
FIG. 9. Results of ROC analysis for the RQA and RNA measures (line colors and symbols as in Fig. 6) when using a data-adaptive discrimination between storm and non-storm conditions based on a threshold value $\langle Dst \rangle^*$. (a) ROC curves using windows of $w = 168$ h, $\langle Dst \rangle^* = -30$ nT. (b) As in (a) for $w = 256$ h. (c) Dependence of AUC on $\langle Dst \rangle^*$ ($w = 256$ h), (d) AUC as a function of the window width w ($\langle Dst \rangle^* = -30$ nT).

specific measure [Fig. 9(d)]. The latter scale of about 10 to 15 days appears related to the typical time scales at which magnetospheric dynamics exhibits transitions between intense storm events and quiescence conditions.

In general, we find that for shorter windows the skills of $\mathcal{T}$ as expressed in terms of AUC decay somewhat slower than those of *TT* and *DET*, making this measure (among the six characteristics studied in this work) most suitable for a temporally localized tracing of magnetospheric complexity variations. In turn, the behavior of $\mathcal{L}$ even changes qualitatively, providing AUC values below 0.5 [i.e., worse than a random classification, cf. Figs. 9(a) and 9(d)]. This observation is related to a transition from negative towards positive correlations between $\mathcal{L}$ and $\langle Dst \rangle$ as w decreases. This qualitative change in the behavior of $\mathcal{L}$ might be understood as follows: Larger windows typically cover a succession of different individual storm/quiescence intervals, indicating persistent changes in magnetospheric complexity with a two-state pattern within the same time window. In such a situation, we can expect high values of $\mathcal{L}$. In turn, shorter windows possibly only capture either one storm or non-storm phase so

that the recurrence characteristics relate to the dynamics of individual storms, which are more homogeneous and could therefore give rise to lower values of $\mathcal{L}$ than during (intermittent) short quiescence periods. In this spirit, both very high and very low values of $\mathcal{L}$ can be considered as indicators of regime changes,[31] a feature that has not yet been explored in full detail. Our present results provide indications towards a possible general explanation of this observational fact in terms of heterogeneity in the system's reconstructed phase space.

When varying the threshold $\langle Dst \rangle^*$ for distinguishing between storm and non-storm conditions [Fig. 9(c)], we observe that all recurrence measures lose their skills when the classification becomes less informative (i.e., the threshold reaches values close to the normal background level of geomagnetic variations). In turn, for very high negative thresholds, there are only a few storm periods remaining so that the data available for classification become too sparse. As a reasonable trade-off, we recommend (for $w = 256$ h) an operational window of $\langle Dst \rangle^*$ between about -40 and -20 nT, for which the AUC values are relatively stable, justifying our initial choice of a threshold at -30 nT.

B. Comparison with other methods

We repeat our ROC analysis for the non-recurrence characteristics previously discussed in Sec. IV. Figures 10(a) and 10(b) display the resulting ROC curves, and Figs. 10(c) and 10(d) the dependence of the resulting AUC values on the thresholds applied to the window-wise mean Dst values. Most of the studied characteristics again discriminate rather well between conditions with strongly negative versus close to zero Dst values. The observation that this feature is not unique to recurrence measures was to be expected given the previous results using entropic and correlation-based characteristics.

In general, we again find the same three measures which perform worse than the others in the ROC analysis. In case of the Shannon entropy S, the relatively bad discriminatory skills are to be expected, since this measure does not characterize dynamical properties but is calculated based on the (heavily coarse-grained) PDF of the Dst data within each window. For the Hurst exponent, we find that even for very high FRP, we do not reach TPR values sufficiently close to one, which indicates that there are certain storms where this measure takes values

distinctively different from other storm periods. One possible reason for this could be the general problem of properly estimating H from rather short time series segments.[60] Finally, the LVD dimension density δ_{LVD} displays better discriminatory skills than H and S but distinguishes not as good as all considered dynamical entropies between storm and non-storm conditions. Note again that this measure is based on some linear approach, whereas entropic characteristics potentially account for nonlinearities. The imperfect discrimination between storm and non-storm periods by means of δ_{LVD} has already been reported previously[34] and could potentially be improved by systematically optimizing the parameters of this method. However, the latter measure has its particular advantage when considering very short windows, a setting where "fully nonlinear" characteristics commonly experience problems in their estimation.

Under ideal conditions, the three phase space based entropies AppEnt, SampEnt, and FuzzyEnt provide the best skills in distinguishing storm from non-storm conditions, with AUC values of a similar order as for the best recurrence measures. This means that the observed differences between these entropies and some of the recurrence measures in case of
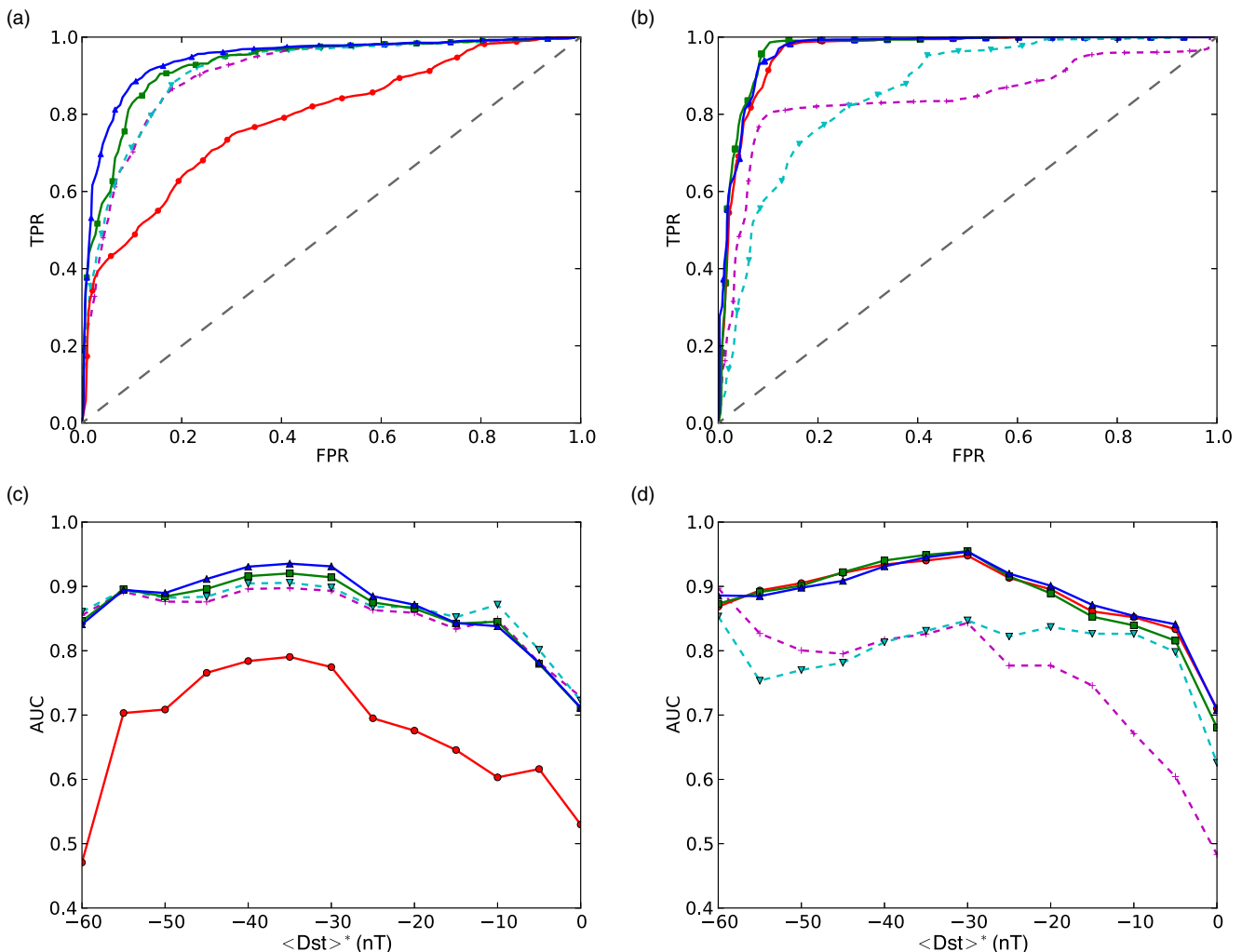

FIG. 10. [(a) and (b)] As in Figs. 8(a) and 8(b) for a data-adaptive classification with a fixed threshold $\langle Dst \rangle^* = -30$ nT ($w = 256$ h, $\Delta w = 1$ h). [(c) and (d)] AUC values calculated from the ROC curves obtained for different values of $\langle Dst \rangle^*$. Line colors and symbols as in (a) and (b).

the heuristic classification of storm and non-storm periods are largely relieved when considering a fully data-adaptive discrimination. In comparison with the phase space based entropies, the information-theoretic and statistical mechanics-based entropy characteristics exhibit slightly lower maximum AUC values, while the three aforementioned measures Shannon entropy, Hurst exponent, and LVD dimension density generally display significantly lower values. For most measures, an optimum discrimination is again found for Dst threshold values of about -20 to -40 nT consistent with the corresponding results for the RQA/RNA measures.

VI. DISCUSSION

The results described in Sec. V indicate that recurrence-based characteristics generally have great potentials in tracing temporal variations in the dynamical complexity of geomagnetic variations, but also other non-stationary geophysical systems. Regarding the inferred dynamical complexity of magnetospheric fluctuations during storm and non-storm conditions, the obtained results are in good agreement with the existing body of literature on this subject.[26]

From the methodological perspective, the consideration of multiple measures from RQA and RNA in this study—based on the same underlying structure, but characterizing different types of statistical properties—allowed us to discriminate between statistics that are better versus such that are less well suited for the purpose of distinguishing magnetic storms and quiescence phases based on the dynamical complexity of fluctuations of the Dst index. The obtained results provide interesting and (beyond the limits of this work) relevant information on the general potentials and applicability of these different characteristics. In turn, the fact that different recurrence measures appear to exhibit (at least quantitatively) different behaviors for transition periods between "physiological" and "pathological" states of the magnetosphere (in our case most notably short periods of relative quiescence of the magnetic field in between marked magnetic storms) makes such a multi-measure perspective a prospective approach for further studying magnetic field variations beyond general storm/quiescence variability, for example, regarding storm/substorm sequences based on other geomagnetic indices. This aspect shall be further addressed in future work, thereby extending this study to both, additional time periods and other geomagnetic activity indices that are better suited for tracing other magnetospheric phenomena like substorms.

More specifically, the findings of this work demonstrate that some of the recurrence-based measures perform (depending on the specific setting) at least comparatively well as the best of the previously studied dynamical entropy characteristics in discriminating between storm and non-storm periods according to the dynamical complexity of magnetospheric variations. However, this tentative result should be further challenged from a methodological perspective: The so far best performing entropies have been developed based upon similar phase space based considerations as recurrence analysis and thus include the same intrinsic methodological parameters (i.e., embedding dimension m and delay τ). However,

unlike for the recurrence measures used in this work, the optimal choice of these parameters has not yet been systematically addressed for these measures. Therefore, to this end, we cannot provide a fair and finally conclusive comparison between entropies and recurrence characteristics, since a detailed study of different methodological settings for all entropy measures has been beyond the scope of this work. In general, we emphasize the potential of a further performance gain by systematically tuning embedding dimension and delay such as to achieve maximum AUC values.

From a conceptual viewpoint, the latter aspect raises the additional question of how to choose the embedding parameters in some optimum way in situations where the correlation properties of the data change crucially with time. For example, correlations in Dst values are much stronger expressed during magnetic storms than within periods of quiescence, as also indicated by the corresponding Hurst exponent and LVD dimension density.[19,34] In this work, we have chosen to fix the embedding delay at some globally justified value (selected according to commonly accepted standard criteria) disregarding the extremely strong temporal changes in the correlation structure. We emphasize that this approach ensures mutual comparability of the values of all dynamical characteristics obtained for different time windows and is thus advantageous for tracing variations in dynamical complexity, since recurrence measures (as well as other nonlinear dynamical properties) can exhibit a marked dependence on the embedding parameters. On the other hand, it could also be justified to take the opposite perspective and choose the embedding delay adaptively for each window. In fact, we have repeated the same recurrence analysis as shown in the previous sections using a time-dependent embedding delay, but did not obtain any conclusive results regarding the discrimination between dynamical complexity during storm and non-storm periods in that case (not shown).

Going even further into the methodological details of this study, another interesting aspect to be addressed in future work is to examine the different performance of the considered recurrence characteristics in even more detail. For example, *TT* and *LAM* are based on the same rationale (i.e., quantifying statistical properties of the length distributions of vertical line structures in recurrence plots), but *TT* clearly outperforms *LAM*. Following this observation, a promising approach could be considering the temporal changes associated with the full PDF of these line lengths to systematically address the question which statistical property (or combination of statistics) associated with this PDF provides the best discrimination between storm and non-storm conditions (and why). In the same way, one could proceed for the PDF of diagonal line lengths (e.g., comparing *DET* with other characteristics like the mean diagonal line length *LMEAN* not considered in this work) or different recurrence network properties based on transitive relationships (beyond $\mathcal{T}$ and $\mathcal{C}$). Yet another possible extension of the present work would be additionally considering recurrence time statistics (i.e., statistics based on the length of vertical non-recurrence structures or "white lines" in the recurrence plots[61]), which would provide a systematic extension to the study of return periods of storms and other magnetospheric

disturbances towards finer scales and more dynamical aspects of variability.

Furthermore, it should be noted that the AUC values reported in this work should be considered as estimates of the "true" values, which differ to a certain extent depending on the specific calculation strategy (in particular, the level of detail of the underlying ROC curves). In this spirit, given the finite length of the used data and the intrinsically non-stationary character of the Dst index, it might be useful to apply a further optimized estimation strategy. Such a strategy could be based on computing the TPR and FPR values for each attained value of the respective measure in the full sample instead of some coarse-graining of the associated range or the PDF as used in this work. In addition, providing confidence bounds for the obtained AUC values (e.g., via cross-validation or bootstrapping techniques) would help statistically evaluating the differences between the skills of different methods in more detail. While such a treatment would have significantly enhanced the computational efforts of this study, given the multiple approaches used in this work, we are confident that the obtained results are at least qualitatively robust even without such further refinements.

Finally, we have demonstrated the robustness of our results for a given embedding dimension m chosen according to the restrictions originating from our sliding window analysis and the associated window widths. In turn, it would be worth accounting for the possibly larger dimensionality of Dst index variations (as suggested by the corresponding results in the supplementary material), thereby contributing to the understanding of the dynamical behavior of the magnetosphere (respectively, the subsystem represented by Dst) as some low-dimensional dynamical system. Moreover, a systematic study of the sensitivity of our quantitative results with respect to the choice of the recurrence rate RR would provide further information on characteristic ranges of Dst values during activity and quiescence periods of the magnetosphere (captured in terms of the associated recurrence threshold ε) which lead to qualitatively stable results of various flavours of recurrence analysis. Such further parameter studies have been, however, beyond the scope of the present work.

As emphasized above, all results obtained in this work are restricted to the properties of magnetospheric variability during one year of observations. From this analysis, we therefore cannot make any detailed statements about the generality of our findings for other years. Notably, not only the statistical distribution of storm and non-storm periods as well as storm magnitudes (as expressed by Dst variations or other geomagnetic indices) can be expected to vary from year to year (even beyond the solar Schwabe cycle). Even more, the specific characteristics of individual storm events may exhibit a certain range regarding the events' magnitudes, durations, and dynamical characteristics, and differences in the latter aspect might result in modifications of our results. We expect such differences to be only quantitative rather than qualitative, but it would require a detailed investigation extending the data to the entire available period of observations to verify this expectation. Complying with other recent studies focusing on individual years of activity as well, we leave this aspect to be addressed in future work.

VII. CONCLUSIONS

These days, space weather research is beginning to systematically use results from complex systems science (see, for instance, the recent Lorentz Center Workshop "Space Weather: A Multi-Disciplinary Approach," https://lorentzcenter.nl/lc/web/2017/921/info.php3?wsid=921&venue=Oort). However, complex system methods and analysis techniques have remained underutilized so far and are not yet widely featured in Space Weather forecasting. In this study, we have demonstrated that recurrence analysis and quantitative measures based upon it are able to successfully discriminate the different levels of dynamical complexity between quiet-time and storm-time conditions in the Earth's magnetosphere. The next step would be to apply the derived best-performing recurrence measures to solar wind parameters and geomagnetic activity indices with higher resolution than Dst, aiming to reveal signatures of the transition from the normal state to the magnetic storm, which could be considered as early warning signals. Such an achievement would clearly enrich our capacity of predicting dynamically the Space Weather. A first step into this direction has been recently made in a companion paper complementing this work.[32]

In the present work, we have employed a suite of selected characteristics from the modern toolbox of recurrence quantification analysis and recurrence network analysis to investigate the time-dependence of different aspects of dynamical complexity exhibited by the Dst index during one year with two marked periods of strong geomagnetic activity peaking in sequences of magnetic storms. While all considered measures have shown their ability to trace complexity variations associated with the succession of storm/non-storm periods very well (i.e., provide possibly relevant space weather diagnostics), different characteristics exhibit different degrees of sensitivity with respect to changes in the magnetospheric variability patterns. Specifically, the recurrence network transitivity $\mathcal{T}$—together with two RQA measures—has been identified as the most sensitive tracer of such variations. A detailed ROC analysis has shown that this property performs comparably well as (or even better than) the best dynamical entropy characteristics considered so far for the purpose of discriminating storm and non-storm conditions based on their nonlinear dynamical characteristics.

Our results provide new information that helps assessing and understanding the potentials of different measures emerging from this still quite a novel approach. In this spirit, the fact that $\mathcal{T}$ performs particularly well could be due to this measure being directly related to a generalized notion of fractal dimension[50] associated with the geometric structure of the data in the phase space reconstructed by means of time-delay embedding. Besides other recent applications of the same characteristic,[30,31,47] this study is among the few cases where fractal dimension concepts have been successfully used for tracing temporal variations of the dynamical complexity of geophysical systems based on single time series. This observation opens promising new research avenues by revisiting the classical concept of fractal dimensions in a geoscientific context.

Regarding the temporal organization of fluctuations inside the Earth's magnetosphere, our findings confirm previous results on the distinctive difference between magnetic storms and quiescence periods. Specifically, storm periods exhibit an elevated degree of dynamical regularity related to the gradual trends of the Dst index during the emergence of magnetic storms and the subsequent recovery phase. The multiplicity of recurrence measures studied in this work allows capturing different facets of the dynamical complexity. Specifically, RQA measures are explicitly linked with the dynamical organization of the system's fluctuations in the sense that they highlight persistent proximity relationships between the embedded values of the Dst index in the associated reconstructed phase space. In turn, RNA measures capture geometric properties of the multi-dimensional distributions of these state vectors and, thus, rather take a structural perspective. The fact that measures from both classes (particularly $\mathcal{T}$ versus TT and DET) exhibit very similar temporal variations underlines that both aspects are closely entangled in the magnetosphere. Further addressing this fact by considering other variability indicators based on different rationales than Dst could therefore allow more detailed insights into the dynamical complexity of the magnetosphere associated with different spatial and temporal scales and processes. In turn, this could potentially advance operational models used for space weather forecasting purposes.

SUPPLEMENTARY MATERIAL

This paper is accompanied by supplementary material including five additional figures and explanatory text providing further details on the employed methods (selection of embedding parameters, sliding window analysis, recurrence quantification, and recurrence network analysis together with a discussion of their respective bootstrapping approaches), discussing the choice of the second component of the embedding vectors as our reference time scale, and further justifying the choices of methodological settings made in our analysis.

ACKNOWLEDGMENTS

[1]P. Song, H. J. Singer, and G. L. Siscoe, *Space Weather*, AGU Geophysical Monograph Series Vol. 125 (American Geophysical Union, Washington, 2011).

[2]H. J. Singer, P. T. Loto'Aniu, J. C. Green, J. V. Rodriguez, B. J. Anderson, J. J. Love, V. Angelopoulos, D. N. Baker, M. G. Connors, W. F. Denig, E. F. Donovan, O. Lecontel, T. G. Onsager, T. Nagatsuma, A. Runov, and E. L. Spanswick, "Multipoint observations of the large substorm associated with the galaxy 15 anomaly," in *American Geophysical Union, Fall Meeting 2010* (American Geophysical Union, WA, 2010), pp. SM22B–05.

[3]D. H. Boteler, R. J. Pirjola, and H. Nevanlinna, "The effects of geomagnetic disturbances on electrical systems at the Earth's surface," Adv. Space Res. **22**(1), 17–27 (1998).

[4]L. Bolduc, "GIC observations and studies in the Hydro-Québec power system," J. Atmos. Solar-Terrestrial Phys. **64**(16), 1793–1802 (2002).

[5]M. Wik, R. Pirjola, H. Lundstedt, A. Viljanen, P. Wintoft, and A. Pulkkinen, "Space weather events in July 1982 and October 2003 and the effects of geomagnetically induced currents on Swedish technical systems," Ann. Geophys. **27**(4), 1775–1787 (2009).

[6]V. Bothmer and A. Zhukov, *The Sun as the Prime Source of Space Weather* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007), pp. 31–102.

[7]I. G. Richardson and H. V. Cane, "Solar wind drivers of geomagnetic storms over more than four solar cycles," AIP. Conf. Proc. **1539**(1), 422–425 (2013).

[8]D. N. Baker, A. J. Klimas, and D. Vassiliadis, *Nonlinear Dynamics in the Earth's Magnetosphere* (Springer New York, New York, NY, 2007), pp. 53–67.

[9]D. N. Baker, *Introduction to Space Weather* (Springer, Berlin, Heidelberg, 2005), pp. 3–20.

[10]I. A. Daglis, G. Balasis, N. Ganushkina, F.-A. Metallinou, M. Palmroth, R. Pirjola, and I. Tsagouri, "Investigating dynamic coupling in geospace through the combined use of modeling, simulations and data analysis," Acta Geophys. **57**(1), 141–157 (2009).

[11]G. Consolini, P. De Michelis, and R. Tozzi, "On the Earth's magnetospheric dynamics: Nonequilibrium evolution and the fluctuation theorem," J. Geophys. Res. Space Phys. **113**(A8), A08222 (2008).

[12]N. W. Watkins, M. P. Freeman, S. C. Chapman, and R. O. Dendy, "Testing the SOC hypothesis for the magnetosphere," J. Atmos. Solar-Terrestrial Phys. **63**(13), 1435–1445 (2001).

[13]D. V. Vassiliadis, A. S. Sharma, T. E. Eastman, and K. Papadopoulos, "Low-dimensional chaos in magnetospheric activity from AE time series," Geophys. Res. Lett. **17**(11), 1841–1844 (1990).

[14]J. Takalo, J. Timonen, and H. Koskinen, "Properties of AE data and bicolored noise," J. Geophys. Res. Space Phys. **99**(A7), 13239–13249 (1994).

[15]H. L. Wei, S. A. Billings, and M. Balikhin, "Analysis of the geomagnetic activity of the D_{st} index and self-affine fractals using wavelet transforms," Nonlinear Process. Geophys. **11**(3), 303–312 (2004).

[16]X. Shao, M. I. Sitnov, S. A. Sharma, K. Papadopoulos, C. C. Goodrich, P. N. Guzdar, G. M. Milikh, M. J. Wiltberger, and J. G. Lyon, "Phase transition-like behavior of magnetospheric substorms: Global MHD simulation results," J. Geophys. Res. Space Phys. **108**(A1), 1037 (2003).

[17]A. S. Sharma, "Global and multiscale features of the magnetosphere: data-derived and global MHD modeling," in *Proceedings of the ILWS Workshop 2006, Goa, India, February 19–24, 2006*, edited by N. Gopalswamy and A. Bhattacharyya (2006), p. 340.

[18]M. I. Sitnov, A. S. Sharma, K. Papadopoulos, and D. Vassiliadis, "Modeling substorm dynamics of the magnetosphere: From self-organization and self-organized criticality to nonequilibrium phase transitions," Phys. Rev. E **65**, 016116 (2001).

[19]G. Balasis, I. A. Daglis, P. Kapiris, M. Mandea, D. Vassiliadis, and K. Eftaxias, "From pre-storm activity to magnetic storms: A transition described in terms of fractal dynamics," Ann. Geophys. **24**(12), 3557–3567 (2006).

[20]A. T. Monk and A. H. Compton, "Recurrence phenomena in cosmic-ray intensity," Rev. Mod. Phys. **11**, 173–179 (1939).

[21]N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," Phys. Rep. **438**(5), 237–329 (2007).

[22]*Recurrence Quantification Analysis: Theory and Best Practices*, edited by C. L. Webber, Jr. and N. Marwan (Springer International Publishing, Cham, 2015).

[23]G. Robinson and M. Thiel, "Recurrences determine the dynamics," Chaos **19**(2), 023104 (2009).

[24]R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths, "Recurrence-based time series analysis by means of complex network methods," Int. J. Bifurcat. Chaos **21**(4), 1019–1046 (2011).

[25]G. Balasis, I. A. Daglis, C. Papadimitriou, M. Kalimeri, A. Anastasiadis, and K. Eftaxias, "Dynamical complexity in $D_s t$ time series using nonextensive Tsallis entropy," Geophys. Res. Lett. **35**(14), L14102 (2008).

[26]G. Balasis, I. A. Daglis, C. Papadimitriou, M. Kalimeri, A. Anastasiadis, and K. Eftaxias, "Investigating dynamical complexity in the magnetosphere using various entropy measures," J. Geophys. Res. Space Phys. **114**(A9), A00D06 (2009).

[27]J. A. Bastos and J. Caiado, "Recurrence quantification analysis of global stock markets," Physica A **390**(7), 1315–1325 (2011).

[28]J. P. Zbilut, N. Thomasson, and C. L. Webber, "Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals," Med. Eng. Phys. **24**(1), 53–60 (2002).

[29]S. Schinkel, N. Marwan, and J. Kurths, "Brain signal analysis based on recurrences," J. Physiol. **103**(6), 315–323 (2009).

[30]J. F. Donges, R. V. Donner, K. Rehfeld, N. Marwan, M. H. Trauth, and J. Kurths, "Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis," Nonlinear Process. Geophys. **18**(5), 545–562 (2011).

[31]J. F. Donges, R. V. Donner, M. H. Trauth, N. Marwan, H. J. Schellnhuber, and J. Kurths, "Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution," Proc. Natl. Acad. Sci. U.S.A. **108**, 20422–20427 (2011).

[32]R. V. Donner, G. Balasis, V. Stolbova, M. Georgiou, M. Wiedermann, and J. Kurths, "Recurrence based quantification of dynamical complexity in the Earth's magnetosphere at geospace storm timescales," e-print arXiv:1801.09412, 2018.

[33]G. Balasis, I. A. Daglis, C. Papadimitriou, A. Anastasiadis, I. Sandberg, and K. Eftaxias, "Quantifying dynamical complexity of magnetic storms and solar flares via nonextensive tsallis entropy," Entropy **13**(10), 1865–1881 (2011).

[34]R. V. Donner and G. Balasis, "Correlation-based characterisation of time-varying dynamical complexity in the earth's magnetosphere," Nonlinear Process. Geophys. **20**(6), 965–975 (2013).

[35]N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," Phys. Rev. Lett. **45**, 712–716 (1980).

[36]F. Takens, *Detecting Strange Attractors in Turbulence* (Springer, New York, 1981), pp. 366–381.

[37]J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," Europhys. Lett. **4**(9), 973 (1987).

[38]Since the Dst data are given as integer values, we add some small Gaussian white noise with a standard deviation of 0.01 before computing the distances to avoid problems by properly fixing the recurrence rate. It has been checked that different realizations or somewhat different standard deviations of the noise do not affect the results presented in this paper.

[39]K. H. Krämer, R. V. Donner, J. Heitzig, and N. Marwan, Dimension-scalable recurrence threshold estimation, e-print arXiv:1802.01605, 2018.

[40]N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths, "Complex network approach for recurrence analysis of time series," Phys. Lett. A **373**(46), 4246–4254 (2009).

[41]R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Recurrence networks—A novel paradigm for nonlinear time series analysis," New J. Phys. **12**(3), 033025 (2010).

[42]J. P. Zbilut and C. L. Webber, Jr., "Embeddings and delays as derived from quantification of recurrence plots," Phys. Lett. A **171**(3–4), 199–203 (1992).

[43]C. L. Webber and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," J. Appl. Physiol. **76**(2), 965–973 (1994).

[44]N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, "Recurrence plot based measures of complexity and its application to heart rate variability data," Phys. Rev. E **66**(2), 026702 (2002).

[45]S. Schinkel, N. Marwan, O. Dimigen, and J. Kurths, "Confidence bounds of recurrence-based complexity measures," Phys. Lett. A **373**(26), 2245–2250 (2009).

[46]Y. Zou, R. V. Donner, J. F. Donges, N. Marwan, and J. Kurths, "Identifying shrimps in continuous dynamical systems using recurrence-based methods," Chaos **20**(4), 043130 (2010).

[47]J. F. Donges, R. V. Donner, N. Marwan, S. F. M. Breitenbach, K. Rehfeld, and J. Kurths, "Non-linear regime shifts in holocene asian monsoon variability: Potential impacts on cultural change and migratory patterns," Clim. Past **11**(5), 709–741 (2015).

[48]Note that we have omitted here the ε-dependence of A_{ij} for brevity.

[49]J. R. Johnson and S. Wing, "A solar cycle dependence of nonlinearity in magnetospheric activity," J. Geophys. Res. Space Phys. **110**(A4), A04211 (2005).

[50]R. V. Donner, J. Heitzig, J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "The geometry of chaotic dynamics—A complex network perspective," Eur. Phys. J. B **84**(4), 653–672 (2011).

[51]R. J. Barlow, *Statistics—A Guide to the Use of Statistical Methods in the Physical Sciences* (Wiley, Chichester, 1989).

[52]F. Wilcoxon, "Individual comparisons by ranking methods," Biom. Bull. **1**(6), 80–83 (1945).

[53]T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett. **27**(8), 861–874 (2006).

[54]G. Balasis, I. A. Daglis, A. Anastasiadis, and K. Eftaxias, *Detection of Dynamical Complexity Changes in Dst Time Series Using Entropy Concepts and Rescaled Range Analysis* (Springer Netherlands, Dordrecht, 2011), pp. 211–220.

[55]G. Balasis, R. V. Donner, S. M. Potirakis, J. Runge, C. Papadimitriou, I. A. Daglis, K. Eftaxias, and J. Kurths, "Statistical mechanics and information-theoretic perspectives on complexity in the earth system," Entropy **15**(11), 4844–4888 (2013).

[56]S. M. Pincus, "Approximate entropy as a measure of system complexity," Proc. Natl. Acad. Sci. **88**(6), 2297–2301 (1991).

[57]J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," Am. J. Physiol. Heart Circ. Physiol. **278**(6), H2039–H2049 (2000).

[58]W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," IEEE Trans. Neural. Syst. Rehabil. Eng. **15**(2), 266–272 (2007).

[59]G. Balasis, I. A. Daglis, A. Anastasiadis, C. Papadimitriou, M. Mandea, and K. Eftaxias, "Universality in solar flare, magnetic storm and earthquake dynamics using Tsallis statistical mechanics," Physica A **390**(2), 341–346 (2011).

[60]A. Witt and B. D. Malamud, "Quantification of long-range persistence in geophysical time series: Conventional and benchmark-based improvement techniques," Surv. Geophys. **34**(5), 541–651 (2013).

[61]E. J. Ngamga, D. V. Senthilkumar, A. Prasad, P. Parmananda, N. Marwan, and J. Kurths, "Distinguishing dynamics using recurrence-time statistics," Phys. Rev. E **85**, 026217 (2012).

[62]J. F. Donges, J. Heitzig, B. Beronov, M. Wiedermann, J. Runge, Q. Y. Feng, L. Tupikina, V. Stolbova, R. V. Donner, N. Marwan, H. A. Dijkstra, and J. Kurths, "Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package," Chaos **25**(11), 113101 (2015).