Causal inference in space weather by an information theory approach

M Stumpo\textsuperscript{1,2}, G Consolini\textsuperscript{2}, T Alberti\textsuperscript{2}
\textsuperscript{1} Dipartimento di Fisica, Università degli Studi di Roma Tor Vergata, via della Ricerca Scientifica 1, I-00133 Roma, Italy
\textsuperscript{2} INAF-Istituto di Astrofisica e Planetologia Spaziali, Via del Fosso del Cavaliere 100, 00133, Roma, Italy
E-mail: mirko.stumpo@gmail.com

Abstract. The variability of the interplanetary medium and solar wind conditions strongly affects the near-Earth environment causing several phenomena, such as the magnetic storms and substorms. Here, we present some preliminary results on the causal inference in the solar wind-magnetosphere-ionosphere coupling by means of information theory which could be helpful for Space Weather modeling.

1. Introduction
The near-Earth environment is strongly affected by the changes of the physical conditions of the solar wind and interplanetary medium due to the solar activity, being the responsible of several phenomena characterizing the solar wind-magnetosphere-ionosphere (SMI) system, such as for instance the geomagnetic storms and substorms, i.e., large magnetic field perturbations and disturbances [1]. Such phenomena can affect technological systems and anthropic activities; indeed, technologies, such as the global positioning system (GPS), which play an important role in our society, are vulnerable to electromagnetic disturbances. Therefore, it is important to be able to mitigate the damages of these phenomena. During the last decades, a new discipline, named Space Weather, was born with the aim to get a deeper knowledge of the physical processes responsible for the generation of such disturbances in the near-Earth environment and to attempt to model and forecast their complex behavior.

Although there is a well-known correlation between some changes of the solar wind conditions and the occurrence of geomagnetic storms and substorms [1], a clear link between the different physical quantities as well as the appropriate variables to describe the SMI dynamics have not been identified yet in terms of some causality relation. In the past, several attempts to forecast geomagnetic disturbances (in terms of both geomagnetic indices variations and/or ground-based measurements) during magnetic storms and substorms have been developed by means of artificial Neural Networks and linear or nonlinear regression models [2, 3]. Most of these models are based on the use of a large set of physical quantities describing the dynamical state of the interplanetary medium and solar wind conditions (interplanetary magnetic field, \textbf{B}, solar wind velocity, \textbf{v}, plasma density \(\rho\), etc.) without attempting to identify the most informative physical quantities. The reason is that all the past studies focussing on the relation between interplanetary conditions...
changes and the near-Earth environment (magnetospheric) response are essentially based on the identification of correlations [4].

According to Granger causality principles [5], forecasting is related to the identification of causal variables responsible for state transitions. Therefore, it is important to combine forecasting and causal inference techniques to get the best results. However, the methods based on the improvement of predictability proposed by Granger are formally applicable only to those cases in which the system response is governed by a linear dynamics. Unfortunately, Space Weather phenomena, such as the occurrence of geomagnetic disturbances, do not belong to this category. Indeed, the response of the Earth’s magnetosphere-ionosphere system to the solar wind changes is mainly nonlinear and complex (see, e.g., [6]). Thus, a more appropriate approach to the problem of the causal inference and the identification of the best physical quantities to model the magnetospheric response would be a point of view based on Information Theory, which provides model-free generalizations of the correlation through the Mutual Information (MI), and of the measure of predictability and causality via the Transfer Entropy (TE) [7, 8].

In this work we discuss some specific measures of correlation and causality in the framework of information theory, with an application to the solar wind-magnetosphere coupling during a disturbed geomagnetic period.

2. Methods: Mutual Information and Transfer Entropy

In the framework of information theory some theoretical quantities can be introduced to measure the distance and the flow of information between two processes. These are the $MI$, which is a way to measure the total linear and nonlinear correlation and the correlation delay between two signals, and the $TE$, which is a quantity capable of identifying the main direction of the information flow, useful to unveil causal relations.

The definition of the $MI$ is based on the Kullback-Leibler divergence,

$$D_{KL}(p||q) = \sum_{i=1}^{n} p_i(x) \log \frac{p_i(x)}{q_i(x)},$$

which, given an observable $x$, is a measure of the inefficiency of assuming that the probability distribution is $q(x)$ when the true distribution is $p(x)$ [8]. From Eq. (1) the $MI$ can be defined as:

$$MI(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij}(x, y) \log \left( \frac{p_{ij}(x, y)}{p_i(x)p_j(y)} \right) = H(X) + H(Y) - H(X,Y),$$

where $p_{ij}(x, y)$ is the joint probability distribution function of observing the pair of values $(x_i, y_j)$, $p_i(x)$ and $p_j(y)$ are the probability distributions of observing $x_i$ and $y_j$ independently, while $H(\cdot)$ denotes the Shannon Entropy [9]. Since this quantity estimates the information shared between the two datasets $X$ and $Y$, if $MI = 0$ one can effectively state that they are independent.

In time series analysis, the mutual information can be used in a time-delayed form (DMI) (see, e.g., [10, 11]), accounting for time-dependent interactions between the systems. Assuming that the variables $X$ and $Y$ are represented by measured signals $x(t)$ and $y(t)$, the $DMI$ can be written as

$$DMI_{x,y}(\tau) = MI(x(t), y(t+\tau)),$$

where $\tau$ is the time delay. The value of the time delay $\tau$ corresponding to the maximum of the $DMI$ provides a measure of the correlation time delay. Nevertheless, the evidence that the process $X$ precedes another process $Y$ cannot be read as an indication of a causal relation between them. Indeed, it is not possible to exclude that there could be a third process $Z$ which drives both the aforementioned processes with different time delays. Precisely, in order to unveil
the occurrence of an information transfer (i.e. a causality relation) from the process $X$ to $Y$, the information (about $X$ and $Y$) that is contained in a third variable $Z$ has to be removed. This can be done introducing the conditional mutual information ($CMI$)

$$CMI(X, Y|Z) = H(X, Z) + H(Y, Z) - H(Z) - H(X, Y, Z).$$

(4)

To deal with this problem Schreiber proposed the Transfer Entropy ($\mathcal{T}E$) as the direct generalization of Granger causality principles to nonlinearly related signals [7].

Assuming that two time series can be approximated by a Markov process, Schreiber proposed a measure of directionality to compute the deviation from the generalized Markov condition

$$p(y_{t+\tau}^n|y_t^n, x_t^m) = p(y_{t+\tau}^n|y_t^n),$$

(5)

where $x_t^m = (x_t, ..., x_{t-m+1})$ and $y_t^n = (y_t, ..., y_{t-n+1})$, while $n$ and $m$ are the orders of the Markov processes $X$ and $Y$. Eq. (5) is satisfied if and only if the transition probability, or dynamics, of $Y$ is independent of the past of $X$, i.e., in the absence of causality from $X$ to $Y$. To measure the “inefficiency” of assuming independence, Schreiber uses the Kullback-Leibler divergence between the two probability distributions to define the transfer entropy from $X$ to $Y$ as

$$\mathcal{T}E_{X\rightarrow Y}(\tau) = \sum_{y_{t+\tau}^n, y_t^n, x_t^m} p(y_{t+\tau}^n|y_t^n, x_t^m) \log \left( \frac{p(y_{t+\tau}^n|y_t^n, x_t^m)}{p(y_{t+\tau}^n|y_t^n)} \right),$$

(6)

where $\tau$ is the time delay.

Palaš et al. [12] showed that this definition of transfer entropy is equivalent to that of the Conditional Mutual Information ($CMI_{X\rightarrow Y}$) when the variable $Z$ is the past history of $Y$. Thus we can rewrite Eq. (6) in terms of the mutual information between the past values of $X$ and the present values of $Y$ conditioned to the past history of $Y$, i.e.,

$$CMI_{X\rightarrow Y} = I(x(t), y(t+\tau)|\hat{P}y(t)),$$

(7)

beating $\hat{P}y = (y(t), y(t-\eta), ..., y(t-(d-1)\eta))$, and $\eta$ and $d$ the embedding delay and dimension, respectively.

3. Dataset and Analysis

To reveal causality relations in the framework of the SMI system we evaluated the $\mathcal{T}E$ between a solar wind and a geomagnetic proxy during the 2013 St. Patrick’s storm by considering the time interval between March, 10 to March, 30, 2013. Specifically, we use the Perrault-Akasofu coupling function, $\varepsilon$, as a proxy of the solar wind [13]. This quantity is defined as:

$$\varepsilon = \frac{4\pi}{\mu_0} I_0^2 v B^2 \sin^4(\theta_c/2),$$

(8)

where $\mu_0$ is the vacuum permeability, $I_0$ is the stand-off distance of the nose of the magnetosphere (typically of the order of $7 - 10$ R$_E$, being $R_E$ the Earth’s radius), $v$ is the flow speed of the solar wind, $B$ is the magnitude of the interplanetary magnetic field and $\theta_c$ is the magnetic field clockwise angle in the plane $yz$. On the other hand, the proxy we used for the magnetosphere-ionosphere system is the Auroral Electrojet (AE) index which provides an estimate of the intensity of the currents flowing in the high-latitude polar ionosphere. This current enhances during geomagnetic storms and substorms, also providing an estimation of the energy deposition rate in the auroral regions [14]. Fig. 1 reports the time series of the two aforementioned proxies with a time resolution of 5 minutes.
To compute transfer entropy $\mathcal{T}E$ by using $CMI_{X\to Y}$, we developed a Python package which implements the estimation of Information Theoretic tools via the $k^{th}$-nearest neighbours distance approach [15]. Our estimator for $\mathcal{T}E$ is:

$$\hat{\mathcal{T}E}_{X\to Y}(\tau) = \psi(k) - \langle N_{x_t,y_{t+\tau}} + N_{y_t,y_{t+\tau}} - N_{y_{t+\tau}} \rangle,$$

(9)

where $N(\cdot)$ is the number of neighbors in marginal spaces and $\psi(\cdot)$ is the digamma function.

Fig. 2 shows the results obtained for the transfer entropy analysis applied to the two time series of solar wind and geomagnetic proxies.

As expected, we note a clear transfer of information from the $\varepsilon$ parameter to the AE-index, being the reverse flow of information very small. The characteristic time delay (i.e. the response time) $\tau_{\text{max}}$ is of the order of 120 min, which is in agreement with the propagation time of a solar wind structure from the Advance Composition Explorer (ACE) spacecraft, located at the L1 Lagrangian point in the interplanetary medium ($\sim 1.5 \times 10^6$ km), to the Earth’s magnetopause (typically of the order of 30 - 45 min), and then from the Earth’s magnetopause to the polar ionosphere (typically 60 - 80 min) (see, e.g., [1]).

4. Conclusions

In this work we have presented a very preliminary application of an Information Theoretic measure, e.g., the Transfer Entropy, to unveil the capability of these methods to reveal causality relations in the SMI system, which is governed by strong nonlinearities. Indeed, in this framework the traditional Granger’s approach to causal inference is unfeasible.

The results have shown that the transfer entropy is capable of measuring the directionality of the information flow from the source (e.g., the solar wind) to the responding system (e.g., the Earth’s magnetosphere). Indeed, the $\mathcal{T}E$ indicates that the high latitude response of the Earth’s magnetosphere-ionosphere system (as monitored by AE) to solar wind changes is driven by the Poynting vector flux ($\varepsilon$), which is thus a primary quantity to be considered for predictive
models. Therefore, we conclude that TE is a powerful instrument to unveil the causal relations among the different quantities involved in the SMI coupling.

This can be viewed as a first step in characterizing and modeling the geomagnetic response to solar activity, which is one of the main targets of Space Weather studies. Clearly, the application of such methods requires an optimization of these techniques on ensemble-based estimation, which will be the subject of future works.

Acknowledgments
The authors acknowledge J.H. King and N. Papatashvilli at NASA and CDA Web for solar wind data. Acknowledgments are also due to N. Ness of Bartol Research Institute, PI of ACE magnetic field instrument, and to D.J. McComas of Southwest Research Institute, PI of ACE/SWEPAM Solar Wind Experiment. The OMNI data were obtained from the GSFC/SPDF OMNIWeb interface at http://omniweb.gsfc.nasa.gov. The authors also acknowledge World Data Center for Geomagnetism (Kyoto) for the use of the geomagnetic indices data.

References
[1] Alberti T et al 2017 J. Geophys. Res. 122 4266 doi:10.1002/2016JA023175
[2] Palocchia G et al 2008 J. Atmos. Sol.-Terr. Phys. 70 663
[3] Camporeale E, Wing S, and Johnson JR (Eds.) 2018 Machine learning techniques for space weather (Netherlands: Elsevier)
[4] Tsurutani B et al 1990 Geophys. Res. Lett. 17 279 doi:10.1029/GL017i003p00279
[5] Granger CWJ 2004 American Economic Rev. 94 421
[6] Consolini G, Alberti T, and De Michelis P 2018 J. Geophys. Res. 123 9065 doi:10.1029/2018JA025952
[7] Schreiber T 2000 Phys. Rev. Lett. 85 461
[8] Širca S 2016 Probability for Physicists (Springer)
[9] Shannon CE 1948 The Bell System Technical Journal 27
[10] De Michelis et al 2011 J. Geophys. Res. 116 A08225 doi:10.1029/2011JA016535
[11] Materassi M et al 2011 Adv. Space Res. 47 877
[12] Paluš M et al 2018 Chaos 28 075307 doi: 10.1063/1.5019944
[13] Perrault P, and Akasofu S-I 1978 Geophys. J. R. Astron. Soc. 54 547
[14] Ahn BH, Akasofu S-I, and Kamide Y 1983 J. Geophys. Res. 88 6275
[15] Kraskov A, Stögbauer H, and Grassberger P 2004 Phys. Rev. E 69 066138