Disentangling the solar activity–solar wind predictive causality at Space Climate scales

Raffaele Reda[1] · Mirko Stumpo[2] · Luca Giovannelli[1] · Tommaso Alberti[3,2] · Giuseppe Consolini[2]

Abstract

The variability in the magnetic activity of the Sun is the main source of the observed changes in the plasma and electro-magnetic environments within the heliosphere. The primary way in which solar activity affects the Earth's environment is via the solar wind and its transients. However, the relationship between solar activity and solar wind is not the same at the Space Weather and Space Climate time scales. In this work, we investigate this relationship exploiting five solar cycles data of Ca II K index and solar wind parameters, by taking advantage of the Hilbert–Huang Transform, which allows to separate the contribution at the different time scales. By filtering out the high-frequency components and looking at decennial time scales, we confirm the presence of a delayed response of solar wind to Ca II K index variations, with a time lag of ∼ 3.1-year for the speed and ∼ 3.4-year for the dynamic pressure. To assess the results in a stronger framework, we make use of a Transfer Entropy approach to investigate the information flow between the quantities and to test the causality of the relation. The time lag results from the latter are consistent with the cross-correlation ones, pointing out the presence of a statistical significant information flow from Ca II K index to solar wind dynamic pressure that peaks at time lag of 3.6-year. Such a result could be of relevance to build up a predictive model in a Space Climate context.

"This paper belongs to the Topical collection "Frontiers in Italian studies on Space Weather and Space Climate", that includes papers written on the occasion of the Second National Congress of SWICo, "Space Weather Italian Community", held on February 9-11 2022 in Rome at ASI, "Agenzia Spaziale Italiana"."

✉ Raffaele Reda
    raffaele.reda@roma2.infn.it

  Mirko Stumpo
    mirko.stumpo@inaf.it

  Luca Giovannelli
    luca.giovannelli@roma2.infn.it

  Tommaso Alberti
    tommaso.alberti@ingv.it

  Giuseppe Consolini
    giuseppe.consolini@inaf.it

[1] Department of Physics, University of Rome Tor Vergata, Via Della Ricerca Scientifica 1, Rome 00133, Italy

[2] INAF - Istituto di Astrofisica e Planetologia Spaziali, Via del Fosso del Cavaliere 100, Rome 00133, Italy

[3] Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, Rome 00143, Italy

1 Introduction

The presence of a magnetic field in the solar atmosphere is the most prominent manifestation of solar variability. Observations of such magnetic variability on the Sun can provide strong observational constraints on the solar dynamo theory, helping to understand the physical mechanisms underlying magnetic flux emergence and evolution. This is particularly interesting on long-term scales, as the solar cycle can offer insights into the complex dynamics of the global dynamo (e.g., Usoskin et al. 2007).

The goal of Space Climate is to describe long-term variations in solar activity and their impact on the heliosphere and the Earth's environment. The time scale usually used to distinguish the two regimes (i.e., Space Weather and Space Climate) is a few solar rotations (Mursula et al. 2007). Here, we focus on the multi-year time scale relationship between solar activity and near-Earth solar wind properties. To achieve this goal, we utilize the full extent of space–age solar wind observations via the OMNI database (King and Papitashvili 2005). Furthermore, we make use of a century-long dataset of the Ca II K index, a widely employed

physical solar activity indicator. This emission index has been demonstrated to be a reliable proxy for the magnetic flux density at the Sun (Schrijver et al. 1989; Ortiz and Rast 2005; Chatzistergos et al. 2019b) and it has also been proven to trace long-term variations in solar activity (e.g., Judge 2006; Bertello et al. 2016; Chatzistergos et al. 2019a).

The first evidence of a strong link between solar wind properties and the solar cycle comes from observations of a period close to 11 years since the very first satellite observations (Siscoe et al. 1978; King 1979; Neugebauer 1981). In later studies, long-term solar wind properties were mostly compared with sunspot numbers, revealing a non-perfect match among the periodicities of the solar cycle and solar wind speed, density, pressure, and magnetic field, further highlighting a delayed response of solar wind signals compared to sunspot numbers (Petrinec et al. 1991; Köhnlein 1996; El-Borie 2002; Katsavrias et al. 2012; Richardson and Cane 2012; Li et al. 2016, 2017; Venzmer and Bothmer 2018; Samsonov et al. 2019). A recent observational analysis of near-Earth solar wind measurements in relation to the Ca II K index was the first to study these properties over the last five solar cycles (Reda et al. 2023b). A 3.2-year lag of solar wind speed with respect to the Ca II K index is found using both cross-correlation and mutual information analysis, while a 3.6-year lag is found between the magnetic proxy and solar wind dynamic pressure. The analysis on the time lag behaviour between the Ca II K index and the same solar wind parameters was further extended in Reda et al. (2023a), studying how their pairwise relative lags vary over the last five solar cycles, with values ranging from 6 years to 1 year.

In Reda et al. (2023a) and Reda et al. (2023b), the Space Climate scales were studied by applying a 37-month moving average to the Ca II K index and solar wind parameters, following the approach used by Köhnlein (1996). However, the use of the moving average as a low-pass filter can remove some relevant features at the Space Climate scales, such as the double maxima present in some solar cycles. In this study, we propose to overcome these issues using the Hilbert–Huang Transform (Huang et al. 1998), and in particular the Empirical Mode Decomposition, to filter out the intrinsic modes with mean periods below 3 years. We, therefore, reproduce a similar analysis as in other studies, studying the delays between the signals using a cross-correlation analysis.

Furthermore, we aim to investigate the causal link between the proxy of solar activity and the characteristics of the solar wind. This causal connection can guide us in exploring the underlying physical mechanisms responsible for this relationship, opening significant prospects for understanding the mechanisms of the solar dynamo at Space Climate scales. For this reason, we intend to use the Transfer Entropy (Schreiber 2000a), a novel approach from information theory recently applied to the analogous complex

dynamics of the Earth's magnetosphere–ionosphere system (Stumpo et al. 2020; Balasis et al. 2023; Stumpo et al. 2023). Transfer entropy can track down the information flow between variables in different directions, thus showing the causality relation between the variables. Although a causal relationship between the driver of solar variability and solar wind properties is expected, such a measure is not assured and can shed light on the parameters of the solar wind that are more influenced. Additionally, this type of analysis provides an independent measure of the solar wind's response times to solar magnetic variability, which is not constrained by a linear analysis. In particular, in this study, we aim to reanalyse the unfiltered data, sampled monthly, to study the solar wind properties on climatological time scales.

These investigations are crucial in an era of significant expansion of human activity in space. The variations of the space radiation environment, for example, have been forecasted for the next 80 years in the perspective of long-term changes in the Space Climate (Barnard et al. 2011). We are at the beginning of a new era of human exploration of deep space, which also aims to colonize and inhabit environments unprotected by the Earth's magnetosphere. Solar wind constitutes the main low-energy and high-flux component of charged particles in the space environment and has a significant impact on the possibility of permanently inhabiting remote locations in the solar system, such as the lunar surface in the near future and Mars in the medium term. For this reason, it is essential to understand the physical mechanisms governing the variability of the solar wind on multi-year scales.

The present article is structured as follows: Sect. 2 gives a description of the data used; Sect. 3 explains the techniques adopted for the analysis here performed; Sect. 4 presents the results of the analysis; finally, in Sect. 5, we list and discuss the results obtained.

2 Data

The data we use in the present study, to investigate the relationship between the solar magnetic activity and the near-Earth solar wind, are measurements of the Ca II K index and of the solar wind speed and dynamic pressure. In particular, we use here the monthly averages of the parameters listed above over the period 1965–2021.

The Ca II K index is a physical proxy of solar magnetic activity that accounts for the emission in the K line of Ca II at 393.4 nm. Such a line is originated in the middle solar atmosphere (i.e., the chromosphere) and it is related to the mean chromospheric emission of the Sun. The Ca II K index is one of the more commonly used activity indices and it has been proven to be a great proxy for the line-of-sight

(LoS) unsigned magnetic flux density along all the phases of the cycle, and not only when sunspots are present (see, e.g., Schrijver et al. 1989; Ortiz and Rast 2005; Chatzistergos et al. 2019b). Specifically, in this work, we make use of the Ca II K index composite presented and described in Bertello et al. (2016), which is freely accessible from the National Solar Observatory (NSO) website at https://solis.nso.edu/0/iss/. The Ca II K 0.1 nm emission index contains inter-calibrated measures from three different observatories (Kodaikanal Solar Observatory, Sacramento Peak and ISS-SOLIS at NSO), and overall, it covers the time period between February 1907 and October 2017. After that date, the SOLIS facility has been offline and no more data from this instrument are available. However, this dataset has been already extended to April 2021 in Reda et al. (2023b), by making use of the Mg II index (University of Bremen composite), which has been proven to strongly correlate with the Ca II K index (see, e.g., Donnelly et al. 1994; Reda et al. 2021, 2023b).

The near-Earth solar wind data are taken from the OMNI database, which can be accessed at https://omniweb.gsfc.nasa.gov/hw.html. The OMNI dataset is a collection of various near-Earth solar wind parameters, both magnetic and plasma ones, provided with different time resolutions (King and Papitashvili 2005). It is compiled using validated data from several spacecrafts, such as IMP, ISEE, ACE, Wind, and Geotail. Among the set of parameters provided by the OMNI database, the analysis we perform in this study regards two dynamic parameters of the solar wind: speed (V_{sw}) and dynamic pressure ($P_{d,sw}$). The latter has been computed starting from the speed (V_{sw}) and the ion density ($n_{i,sw}$), as $P_{d,sw} = 1/2\, m_p n_{i,sw} V_{sw}^2$,

where the proton mass m_p is assumed as the mean ion mass. Data concerning these parameters are available starting from July 1965, thus constituting the main limit for the temporal extension of the analysis we carry out here.

The missing monthly data of Ca II K index, solar wind speed, and dynamic pressure were filled using a simple linear interpolation between the previous and the following monthly data. This procedure allows us to continuously investigate, in the present study, the time interval that goes from July 1965 to April 2021, ensuring to almost fully cover the solar cycles from 20 to 24, together with the beginning of solar cycle 25 (Fig. 1).

3 Methods

3.1 The Hilbert–Huang transform: empirical mode decomposition and hilbert spectral analysis

The Empirical Mode Decomposition (EMD), e.g., the first step of the Hilbert–Huang Transform (HHT), has been first introduced by Huang et al. (1998) as an adaptive and a posteriori decomposition method whose decomposition basis is derived via an iterative process, known as sifting process, based on the local properties of signals (Huang et al. 1998).

Let $y(t)$ be a time-dependent signal, the EMD allows us to write

$$y(t) = \sum_{k=1}^{N} c_k(t) + r(t), \tag{1}$$

Fig. 1 Monthly averages of the time series used for this work: Ca II K index (green, top), solar wind speed (red, middle), and solar wind dynamic pressure (blue, bottom) (color figure online)

52

Rendiconti Lincei. Scienze Fisiche e Naturali (2024) 35:49–61
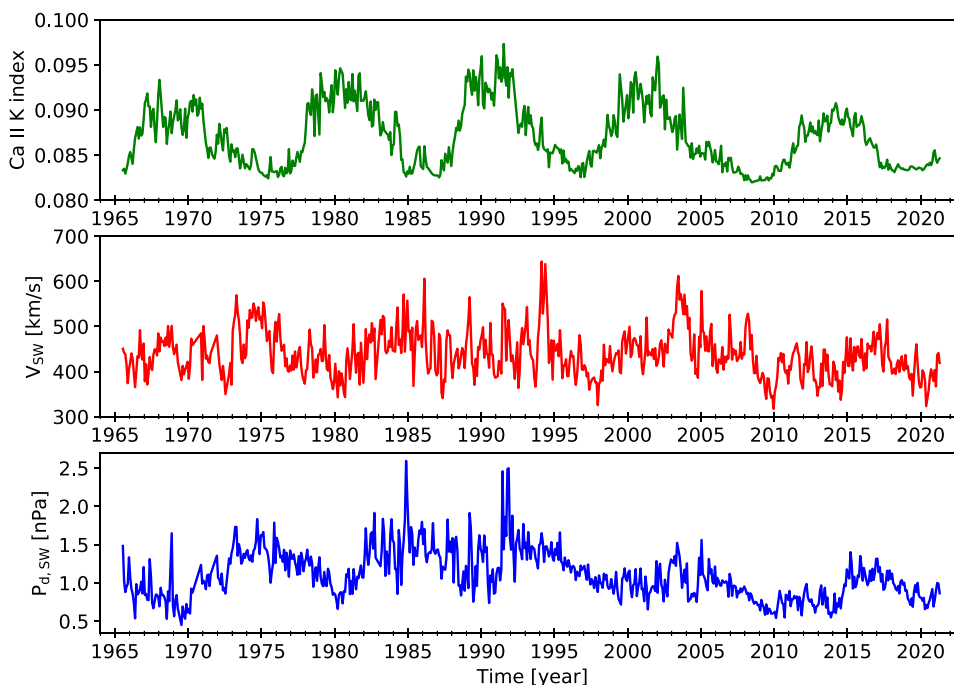
where the set $\{c_k(t)\}$, named as Intrinsic Mode Functions (IMFs) or empirical modes, forms the decomposition basis, while $r(t)$ is the residue of the decomposition. The latter is a non-oscillating function of time, while an IMF is defined as a function having the same (or differing at most by one) number of extrema and zero crossings and a zero-average mean envelope derived from local maxima and minima envelopes, obtained by interpolating them using a cubic spline (e.g., Huang et al. 1998; Huang and Wu 2008). Table 1 summarizes the main steps of the sifting process.

The authors proposed in Huang et al. (1998) the following constraints as exit condition to stop the sifting process

$$\sigma_n = \sum_j \frac{\left[\delta_n(t_j) - \delta_{n+1}(t_j)\right]^2}{\delta_n(t_j)^2} < \epsilon, \tag{2}$$

being fixed $\epsilon \in [0.2, 0.3]$. This criterion has been refined by Rilling et al. (2003) by the so-called threshold method based on two thresholds, θ_1 and θ_2, to guarantee globally small fluctuations (as in Huang et al. 1998) and to avoid locally large excursions (Flandrin et al. 2004).

In this way, a completely adaptive procedure is built, allowing us in deriving embedded oscillations without assuming linearity and/or stationarity. The derived set of empirical modes $\{c_k(t)\}$ satisfies mathematical requirements of completeness, convergence, and local orthogonality by construction (Huang et al. 1998), while global orthogonality is a posteriori guaranteed, since $\langle c_k, c_{k'} \rangle = \delta_{kk'}$, being $\langle \dots \rangle$ the scalar product between functions, and $\delta_{kk'}$ the Kronecker tensor (e.g., Huang and Wu 2008).

Table 1 The main steps of the sifting process

$y(t) \to y_m(t) = y(t) - \langle y(t) \rangle$

$\delta(t) = y_m(t)$

1. find local extrema of $\delta(t)$

2. find upper and lower envelopes by using cubic spline
   $\to \mathcal{U}(t), \mathcal{L}(t)$

3. find the mean envelope $\to \mathcal{M}(t) = \frac{\mathcal{U}(t) + \mathcal{L}(t)}{2}$

4. update $\delta(t) \to \delta(t) - \mathcal{M}(t)$

if $\delta(t)$ is an IMF

store $c_k(t) = \delta(t)$

$\delta(t) \to \delta(t) = y_m(t) - \delta(t)$

repeat steps 1.-4.

else

iterate steps 1.-4. until $\delta(t)$ is an IMF

store $c_k(t) = \delta(t)$

$\delta(t) \to \delta(t) = y_m(t) - \delta(t)$

repeat steps 1.-4.

stop the process when $r(t) = \delta(t)$ is a non-oscillating function or has only two extrema

Being derived the set of empirical modes, by means of the so-called Hilbert Transform (HT), i.e., the second step of the HHT, we can write each of them as modulated both in amplitude and in frequency (e.g., Huang et al. 1998). Indeed, given an empirical mode $c_k(t)$, we can define its Hilbert Transform $\hat{c}_k(t)$ as

$$\hat{c}_k(t) = \frac{1}{\pi} \mathcal{P} \int_0^\infty \frac{c_k(t')}{t - t'} dt', \tag{3}$$

where $\mathcal{P}$ is the Cauchy principal value. By introducing the complex signal

$$\zeta_k(t) = c_k(t) + i\,\hat{c}_k(t) = \alpha_k(t)e^{i\,\varphi_k(t)}, \tag{4}$$

it follows

$$\alpha_k(t) = \sqrt{c_k(t)^2 + \hat{c}_k^2} \tag{5}$$

$$\varphi_k(t) = \tan^{-1}\left[\frac{\hat{c}_k(t)}{c_k(t)}\right], \tag{6}$$

where $\alpha_k(t)$ and $\varphi_k(t)$ are the instantaneous amplitude and phase of the $k-$th empirical mode, respectively. The definition of instantaneous frequency derives from the instantaneous phase as $\omega_k(t) = \frac{1}{2\pi}\frac{d\varphi_k(t)}{dt}$. Similarly, the mean time scale is $\tau_k = \langle \omega_k^{-1}(t) \rangle_t$, with $\langle \dots \rangle_t$ identifying the time average.

3.2 Transfer entropy

The notion of cause–effect is a delicate question when data from controlled experiments are not available. This is the case of complex systems in general: when dealing with a system whose complete set of dynamical variables is not known a priori and the state of the system is monitored by some indices (which work as proxies) derived empirically, correlation may be confused with causation.

Data-driven methods for studying the degree of causation have been developed in the recent years. Generally, these methods are based on the notion of predictability, i.e., it is said that X drives Y if the knowledge of X's past gives us information about Y's future, but not vice versa. This type of causality is known as predictive causality, and it is restricted to only two variables, X and Y respectively. Mathematically, the concept of predictive causality is expressed through conditional independence, i.e., it is reasonable to assume that X does not drive Y if

$$p(Y_t|\mathbf{Y}_{t-1}^{(k)};\mathbf{X}_{t-\tau}^{(l)}) = p(Y_t|\mathbf{Y}_{t-1}^{(k)}), \tag{7}$$

where $\mathbf{X}_{t-\tau}^{(l)} = (X_{t-\tau},...,X_{t-\tau-l})$, $\mathbf{Y}_{t-1}^{(k)} = (Y_{t-1},...,Y_{t-1-k})$, $p(\cdot)$ denotes the probability and τ is a time lag. Therefore, to measure predictive causality the idea is to test Eq. (7). One

Rendiconti Lincei. Scienze Fisiche e Naturali (2024) 35:49–61

53

way to quantify the distance between the right-hand side (r.h.s.) and the left-hand side (l.h.s.) of Eq. (7) is using the Kullback–Leibler Divergence. In this case, testing Eq. (7) becomes equivalent to test whether the expression

$$T_{X \to Y}^{(k,l)}(\tau) = \sum_{Y_t, \mathbf{Y}_{t-1}^{(k)}, \mathbf{X}_{t-\tau}^{(l)}} p(Y_t, \mathbf{Y}_{t-1}^{(k)}, \mathbf{X}_{t-\tau}^{(l)}) \log \frac{p(Y_t | \mathbf{Y}_{t-1}^{(k)}, \mathbf{X}_{t-\tau}^{(l)})}{p(Y_t | \mathbf{Y}_{t-1}^{(k)})}$$

(8)

is different from zero (Schreiber 2000b). Equation (8) is known as *transfer entropy* (TE). Note that $T_{X \to Y}^{(k,l)}(\tau = 0) \neq T_{Y \to X}^{(k,l)}(\tau = 0)$, i.e., the transfer entropy is asymmetric as expected from a measure of predictive causality (Schreiber 2000b).

In principle to test causality. one needs to include in the l.h.s. of Eq. (7) all the information available in the universe at time $t - 1$ and all the information available in the universe with the exception of $\mathbf{X}_{t-\tau}^{(l)}$ in the r.h.s (Pearl 2009). For controlled systems, all the variables influencing the measurements of Y's state are assumed to be known and the transfer entropy becomes measurable. This naturally set a limit to the application of such causal inference technique. For example, in the cases in which all the relevant variables are not known *a priori*, we can still compute Eq. (8), but it can be different from zero even though the interaction between X and Y is mediated by, e.g., a third variable Z (indirect causation; see, e.g., Bossomaier et al. 2016). Thus, the predictive causality does not generally imply the true cause–effect relationship if the information of the whole set of relevant variables is not available. However, it can be still used to measure lags between variables and directional coupling (Wibral et al. 2013).

From a numerical point of view, a key question is whether or not the values found for the transfer entropy are statistically significant. In our case, the critical value of the transfer entropy $T^{(*)}$ above which we can reject the null hypothesis is computed by generating surrogate time series satisfying Eq. (7) and with the same statistical properties of X and Y. To achieve this, we create surrogate trials by randomly shuffling the time series $X_{t-\tau}$. This allows us to estimate the distribution of the null hypothesis, to fix a confidence bound, and to find the critical value $T^{(*)}$ which adapts to our dataset. The values such that $T_{X \to Y}^{(k,l)}(\tau) > T^*$ are considered statistically significant.

4 Results

The results of Empirical Mode Decomposition are shown in Fig. 2 for the Ca II K index, in Fig. 3 for the solar wind speed and in Fig. 4 for the solar wind dynamic pressure. The mode decomposition generates 6 IMFs for the Ca II K index,

7 IMFs for the solar wind speed, and 7 IMFs for the solar wind dynamic pressure.

The characteristic time scales (or the average period) of the IMFs obtained with the EMD are shown, for each signal, in Table 2.

The Ca II K index, as expected, displays an intrinsic component (IMF 5) related to the 11-year solar activity cycle, in particular with a mean period of 12.4 years. Also, the solar wind speed and dynamic pressure display a component at solar cycle time scales. For both parameters, it is the IMF 6, corresponding to a mean period of 11.2 years for the speed and 13.9 years for the dynamic pressure. It is interesting to notice that for solar wind speed and dynamic pressure, we obtain a mode with a quasi-biennial periodicity (IMF 4 in both cases), while this is not true for Ca II K index.

To understand which IMF has the highest contribute to the overall variability of the signal, it is possible to compute for each IMF a weighted variance as it follows. The value of the variance of each IMF (σ_{IMF}^2) is normalized to that of the total signal ($\sigma_{IMF_{tot}}^2$), obtained by summing the contribution of all the IMFs but excluding the residual term, and plotted as a function of the mean period of the IMF itself. Such values are shown for the Ca II K index the solar wind speed and the solar wind dynamic pressure in Fig. 5. We can use this information to quantify the contribution of each IMF to the overall observed behaviour. We focus for our analysis on the IMFs with a mean period over 1 year. In the case of Ca II K index (left panel of Fig. 5), the greatest weighted variance is from IMF 5, the one related to the 11-year cycle. In the case of the solar dynamic pressure (right panel of Fig. 5), the major contribution is from IMF 6, once again the one corresponding to the solar cycle time scales. The same is true for the solar wind speed (central panel of Fig. 5), for which over the yearly time scale, the highest contribution is from IMF 6.

To further investigate the results of the decomposition, it is possible to look at how the power is distributed among the IMFs, searching for the presence of possible power laws. This can be done by plotting, for each IMF, the value of the variance (σ_{IMF}^2) multiplied by the corresponding mean period (P_{IMF}), as a function of the mean period of the IMF itself (log-log scale). This is the analogous of a spectral density. The results are shown for Ca II K index, solar wind speed, and dynamic pressure in Fig. 6. It is possible to notice the presence of a power law, characterized by increasing intensity from yearly to solar cycle scales, thus extending over at least one decade, in all the signals. Although, in this time range, the power law is clearly evident, the behaviour at very high and very low frequencies remains uncertain due to the limited data points available in the plots.

54

Rendiconti Lincei. Scienze Fisiche e Naturali (2024) 35:49–61

Fig. 2 Empirical mode decomposition of Ca II K index. The top row shows the starting monthly means. The subsequent rows show the successive order IMFs, while the last row shows the residual signal (color figure online)
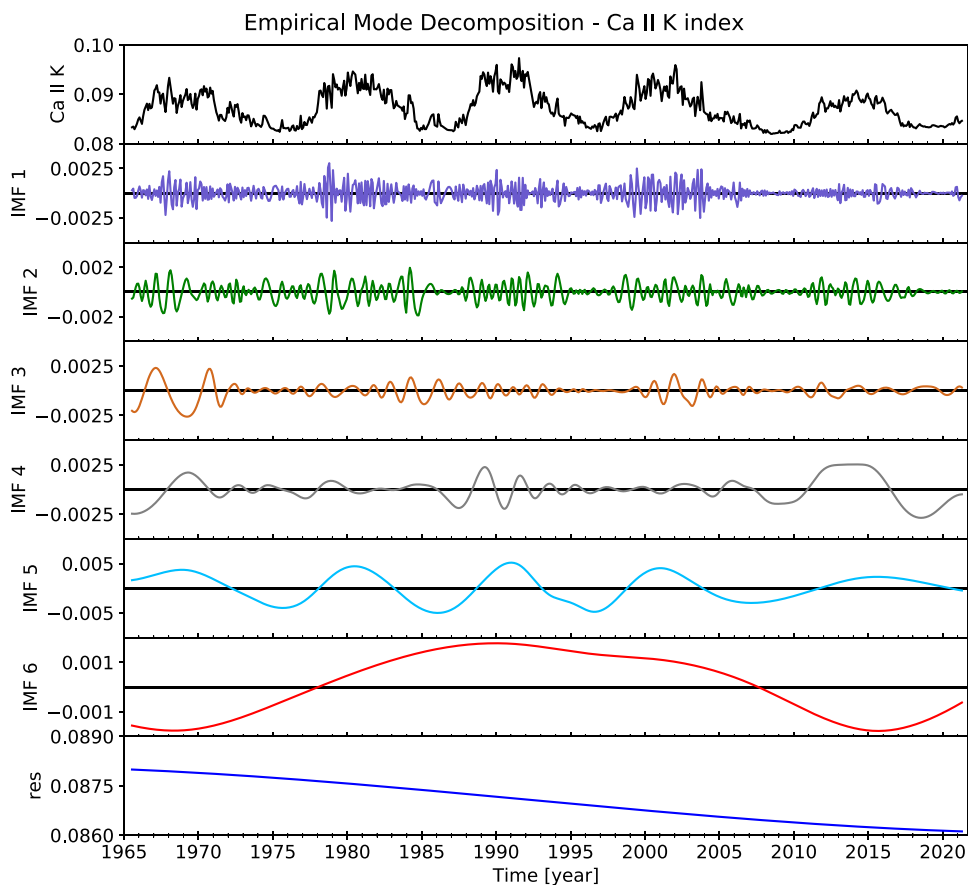

Once the signals have been decomposed via EMD as shown above, it becomes possible to filter the time series of the three parameters by means of the obtained IMFs. To this scope, we subtracted from the monthly time series of Ca II K index, solar wind speed, and dynamic pressure the contribution of the IMFs with mean periods smaller than 3 years. This criterion is chosen to be consistent and to compare the results with the 37-month filtering previously applied in Reda et al. (2023b). In particular, for the Ca II K index, the contribution from IMFs 1–3 have been subtracted, while for both solar wind parameters, we subtract the IMFs 1–4. The comparison between monthly means, 37-month averages, and IMFs filtered data for the three signals is shown in Fig. 7. As it can be seen, the signals obtained by filtering the high-frequency IMFs are consistent with the 37-month moving averages, but they seem to better follow the behaviour of the monthly means with respect to the latter. Indeed, the signals filtered by means of the IMFs retain more information, such as the double solar cycle peak visible in Ca II K index (top panel of Fig. 7).

According to the analysis performed in Reda et al. (2023b), once the high-frequency components of the signals have been filtered out, it is possible to investigate the time lag between them. To this scope, we use here a cross-correlation analysis considering only positive delay of the solar

wind parameters with respect to the activity of the Sun (Ca II K index here). The results of the cross-correlation analysis are shown in Fig. 8. The maximum correlation between Ca II K index and solar wind speed occurs at a time lag of 3.1 ± 0.1 yr, in agreement with the result of 3.2 ± 0.1 yr found in Reda et al. (2023b) using 37-month averaged data. The maximum correlation of Ca II K index with solar wind dynamic pressure, instead, is found for a time lag of 3.4 ± 0.1 yr. This value is in agreement, within the confidence intervals, with the values found in Reda et al. (2023b) with cross-correlation (3.6 ± 0.1 yr) and mutual information (3.4 ± 0.1 yr). These findings strengthen the analysis, as the results do not depend on the technique adopted to filter out the high-frequency components.

The scatter plots of Fig. 9 show the relation of Ca II K index with solar wind speed and solar wind dynamic pressure, respectively, once the time lags from the cross-correlation analysis are taken into account. In both figures, the black lines show the best linear fits to the data points. The Pearson's correlation coefficient is $r = 0.57$ in the case of the speed and $r = 0.56$ in the case of the dynamic pressure, indicating in both cases a positive moderate correlation. For the case of Ca II K index with solar wind dynamic pressure, the correlation coefficient is almost equal to that found in Reda et al. (2023b) using 37-month averaged data ($r = 0.57$),

Fig. 3 Empirical mode decomposition of solar wind speed. The top row shows the starting monthly means. The subsequent rows show the successive order IMFs, while the last row shows the residual signal (color figure online)
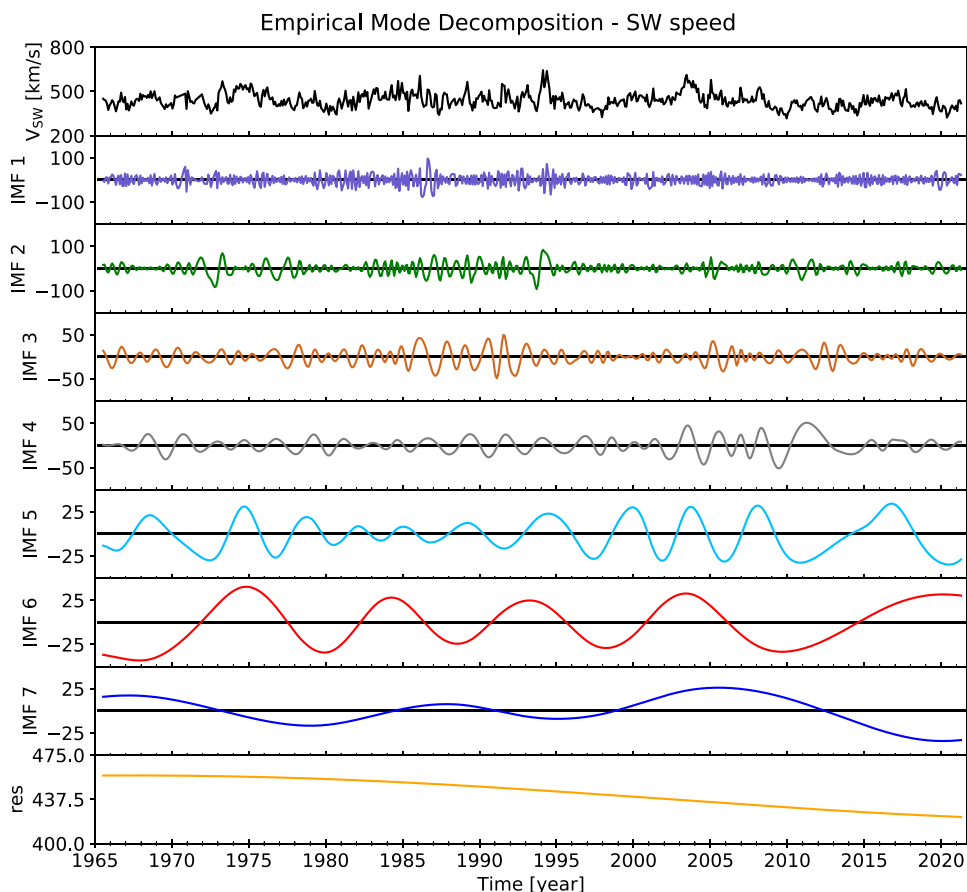

while in the case of solar wind speed, it is smaller compared to the value they found ($r = 0.65$).

To assess in a stronger framework the results obtained with the present analysis, we compute the transfer entropy by directly employing the monthly averaged data, without applying any filter. This approach allows us to investigate higher order correlation between data, i.e., predictive causality as explained in the previous section. In Fig. 10, we show the information flow, as measured by the transfer entropy, from Ca II K index to $P_{d,SW}$ (top-left panel) and *vice versa* (top-right panel). In both cases, the purple line shows the empirical threshold of 99% arising from the analysis of 500 surrogate time series data (see Sect. 3.2). The information flow from the Ca II K index to $P_{d,SW}$ exhibits a statistically significant structure/enhancement between ~25 and ~50 months. The maximum of the transfer entropy is at 43 months ($\simeq 3.6$-year), while the mean of the interval (assuming a symmetric peak) is at 37.5 months ($\simeq 3.1$-year). Both lags are comparable with the results found with the cross-correlation analysis in this work, but also in agreement with the findings by Reda et al. (2023b). The latter result highlights that the correlation found is not simply due to the synchronization of the time series peaks, but it means that there is a predictive link between Ca II K index and the solar wind dynamic pressure, suggesting a certain degree of causation.

On the other hand, the information flow in the reversed case (top-right panel of Fig. 10) is always below the 99% threshold with the exception of a peak at 71 months ($\simeq 5.9$-year; comparable with the distance between solar maximum and solar minimum). We interpret this finding as due to redundancies/periodicities induced by the solar cycle. To test quantitatively this hypothesis, in principle, Eq. (8) should be computed using $k > 1$. However, this is not reliable with only 670 data points.

The bottom panels of Fig. 10 show the results of the transfer entropy analysis for the solar wind speed. In both directions, i.e., from Ca II K index to V_{SW} (bottom-left panel) and *vice versa* (bottom-right panel), there are no strong evidences about the information flow. The exceedances of the threshold at lags of 37 months ($\simeq 3.1$-year) and 60 months ($\simeq 5.0$-year) from Ca II K to V_{SW}, and at 78 months ($\simeq 6.5$-year) from V_{SW} to Ca II K, are here interpreted as a fluctuation due to sampling effects. However, the peak at 3.1-year is consistent with the structures found in the analysis of solar wind dynamic pressure and with the results recently found in Reda et al. (2023b).

Note that, in general, the results of the transfer entropy analysis are noisy. This is due to the fact that we have only 670 data points, so that the estimation of high-dimensional transition probabilities is prone to fluctuations. To mitigate

Fig. 4 Empirical mode decomposition of solar wind dynamic pressure. The top row shows the starting monthly means. The subsequent rows show the successive order IMFs, while the last row shows the residual signal (color figure online)
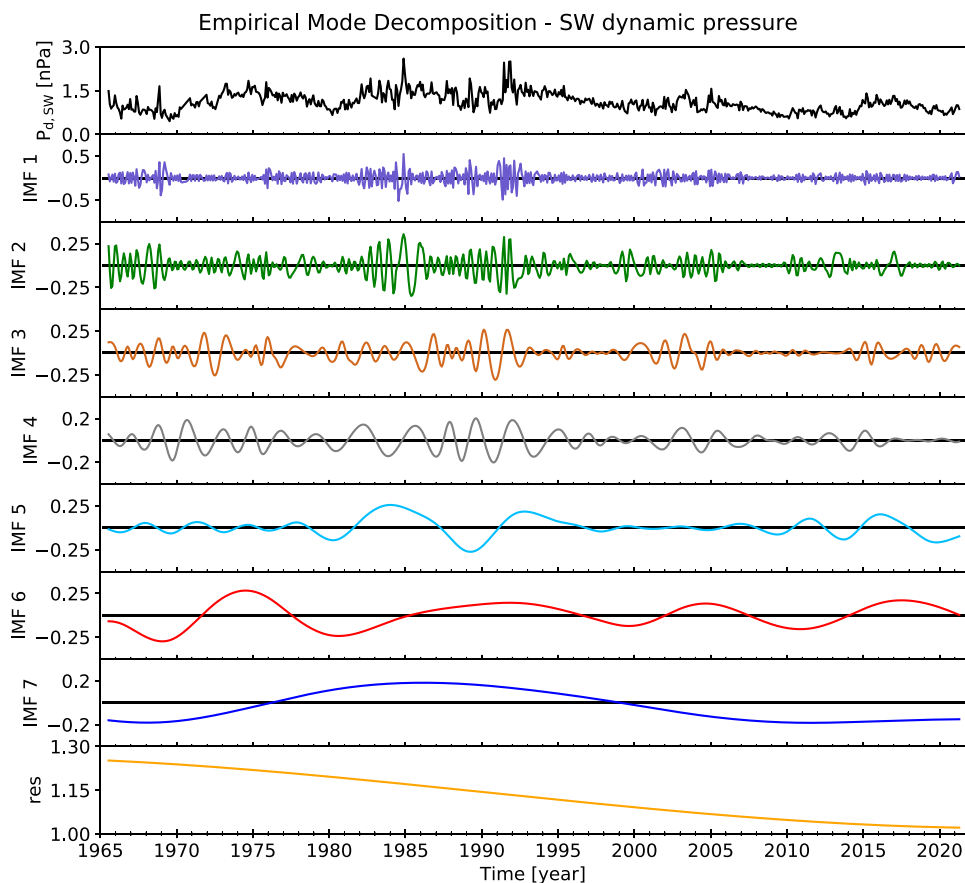
Table 2 Characteristic time scales of the extracted IMFs for each signal

IMF #	Mean period [years]		
	Ca II K index	SW speed	SW dynamic pressure
1	0.36	0.31	0.32
2	0.65	0.59	0.54
3	1.36	1.06	1.18
4	3.58	2.13	2.26
5	12.38	4.84	5.17
6	51.43	11.17	13.88
7	–	19.84	51.45

this effect, the estimation of transition probabilities must be performed by reducing the number of bins. In our case, the best trade-off between correct sampling and resolution (i.e., to have filled bins) is to choose less than 10 bins per dimension. Aware of this technical problem, we interpret threshold exceedances of single-point structures as fluctuations that are not statistically significant.

5 Discussion and conclusions

Starting from the monthly averages of a physical proxy of the solar activity (the Ca II K index) and solar wind parameters, we investigate in this work their relationship on Space Climate scales. To this scope, we take advantage of the Hilbert–Huang Transform. This method allows to decompose the starting signals into several modes and to obtain for each of them the instantaneous frequency and hence the mean characteristic time scale. Looking at how the energy is distributed among the time scales, we find a quite similar behaviour for all the signals between annual and solar cycle scales, characterized by an increasing power. Concerning the behaviour at lower and higher time scales, instead, it is not possible to draw conclusions here.

The advantage of the HHT is that the EMD makes possible to filter out the noisy high-frequency components, which are not of interest for the purpose of this work. The time lags we find between Ca II K index and both solar wind speed (3.1-year) and dynamic pressure (3.4-year), after subtracting the contribution at scales smaller than 3 years, are consistent with the results previously obtained by Reda et al. (2023b) using a 37-month smoothing on the same dataset.
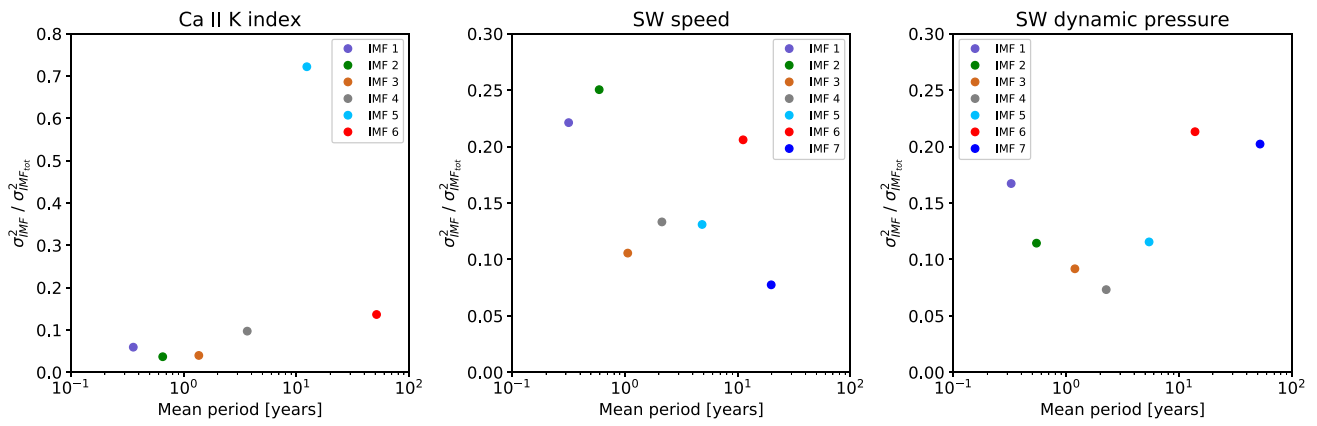
Fig. 5 IMF variance weighted on the IMFs total variance as a function of the IMF mean period. The subplots are for Ca II K index (left), solar wind speed (center), and solar wind dynamic pressure (right) (color figure online)
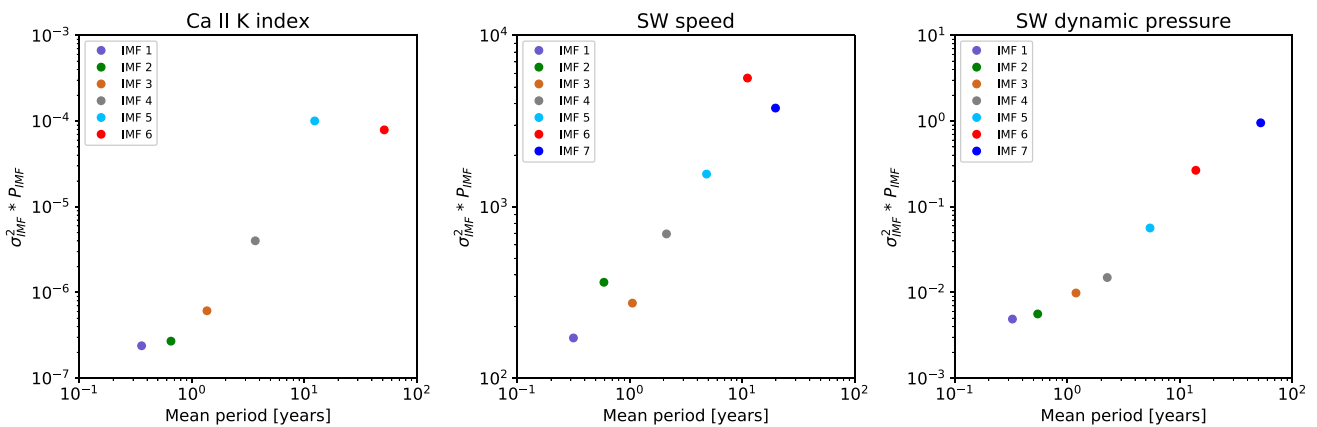

Fig. 6 IMF variance multiplied by the mean period as a function of the IMF mean period (log–log scale). The subplots are for Ca II K index (left), solar wind speed (center), and solar wind dynamic pressure (right) (color figure online)

Fig. 7 Monthly averages with superimposed 37-month moving averages and IMFs filtered signals for Ca II K index (top), solar wind speed (middle), and solar wind dynamic pressure (bottom) (color figure online)
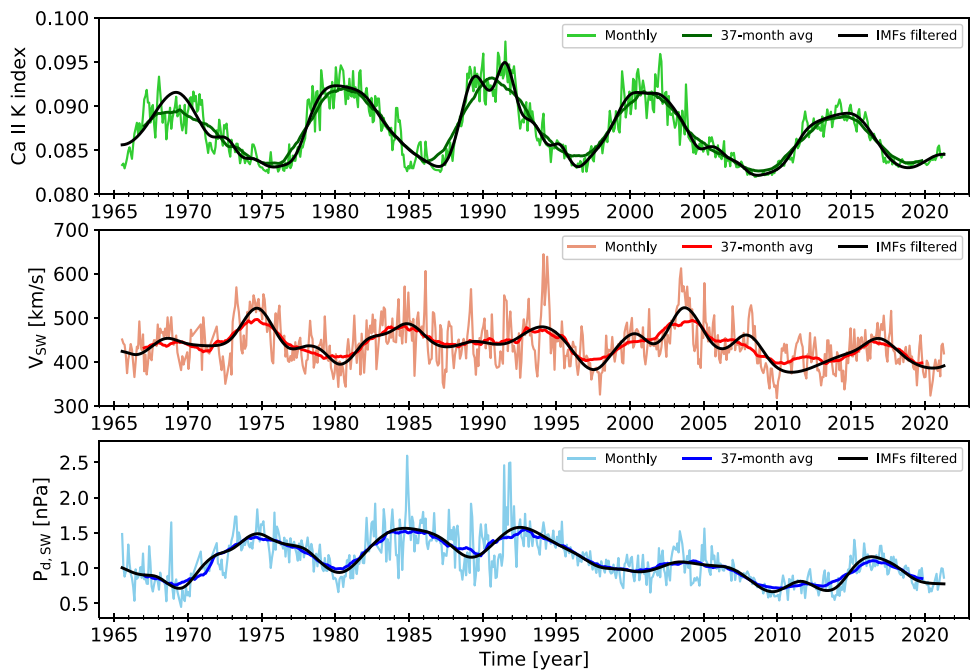
Fig. 8 Cross-correlation of the IMFs filtered signals. **a** Cross-correlation between Ca II K index and solar wind speed; **b** comparison of Ca II K index (green) with solar wind speed (red) shifted backward by 3.1 years; **c** cross-correlation between Ca II K index and solar wind dynamic pressure; **d** comparison of Ca II K index with solar wind dynamic pressure (blue) shifted backward by 3.4 years (color figure online)
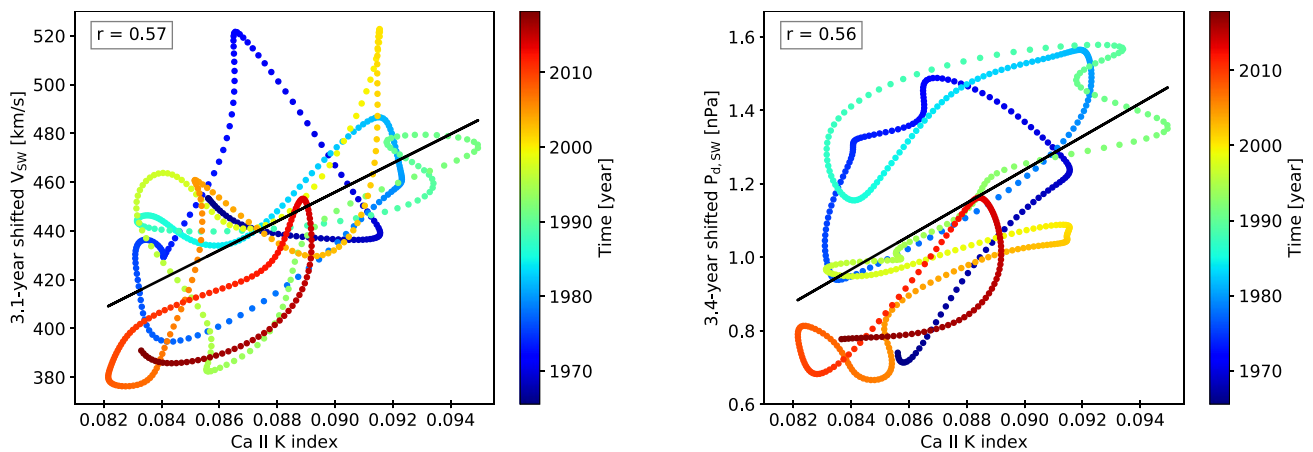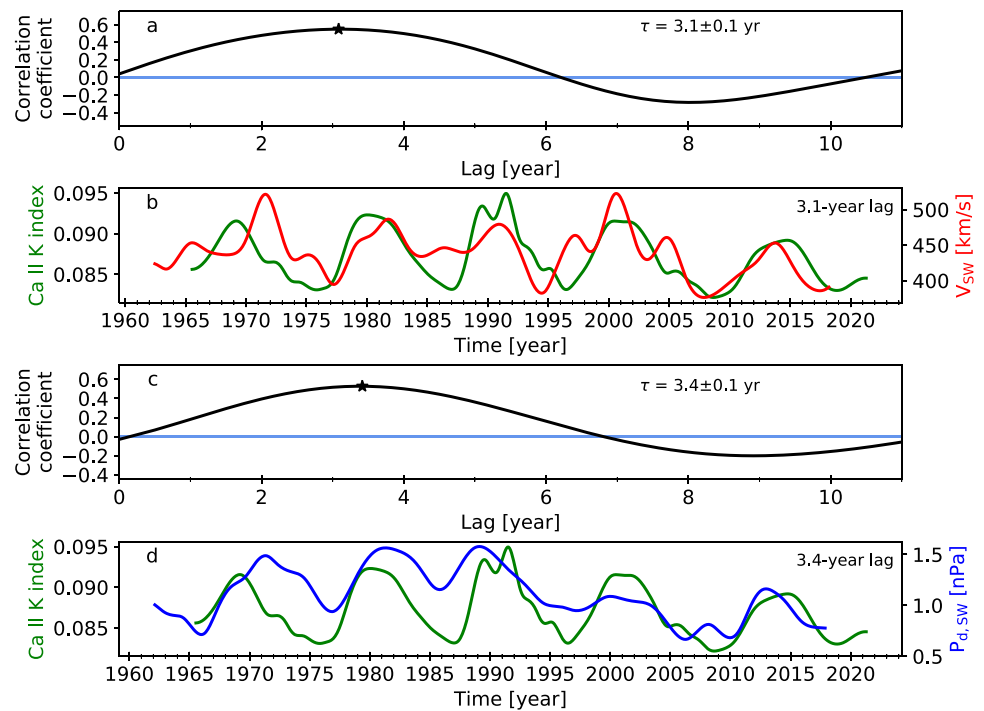
Fig. 9 Scatter plot showing the relationship of Ca II K index with solar wind speed (left panel) and solar wind dynamic pressure (right panel) once shifted by the time lags found with the cross-correlation analysis. In both panels, the color map shows how the relation changes with time, while the black line shows the best linear fit to the data points. The Pearson correlation coefficients are reported on the upper left (color figure online)

However, the presence of a correlation with a time delay does not ensure a cause–effect relation, as well as the presence of mutual information peak does not guarantee the nature of the directional coupling. For this reason, to further investigate these results, we use the transfer entropy as predictive causality test for higher order correlation between data. The results from transfer entropy analysis suggest the presence of statistically significant structures from Ca II K index to solar wind dynamic pressure, with a peak at time lag of 3.6-year, once again in agreement with the time lag found by Reda et al. (2023b). This finding confirms that

the knowledge of past values of the Ca II K index gives information about the future state of the solar wind dynamic pressure. Indeed, the cross-correlation analysis is based on co-variation of data and it is naturally stronger when peaks are synchronized, while the transfer entropy is based on (temporal) transition probabilities between the states and it is a dynamical and time-asymmetric concept. As reported by Wibral et al. (2013), the time-delays found from cross-correlation analysis and transfer entropy analysis are not necessarily the same.

Fig. 10 Transfer entropy from Ca II K index to solar wind dynamic pressure and speed (top- and bottom-left panels, respectively) and *vice versa* (top- and bottom-right panels, respectively), computed using monthly averaged data. The purple line marks the 99% significance threshold obtained from the analysis of surrogate time series data (color figure online)
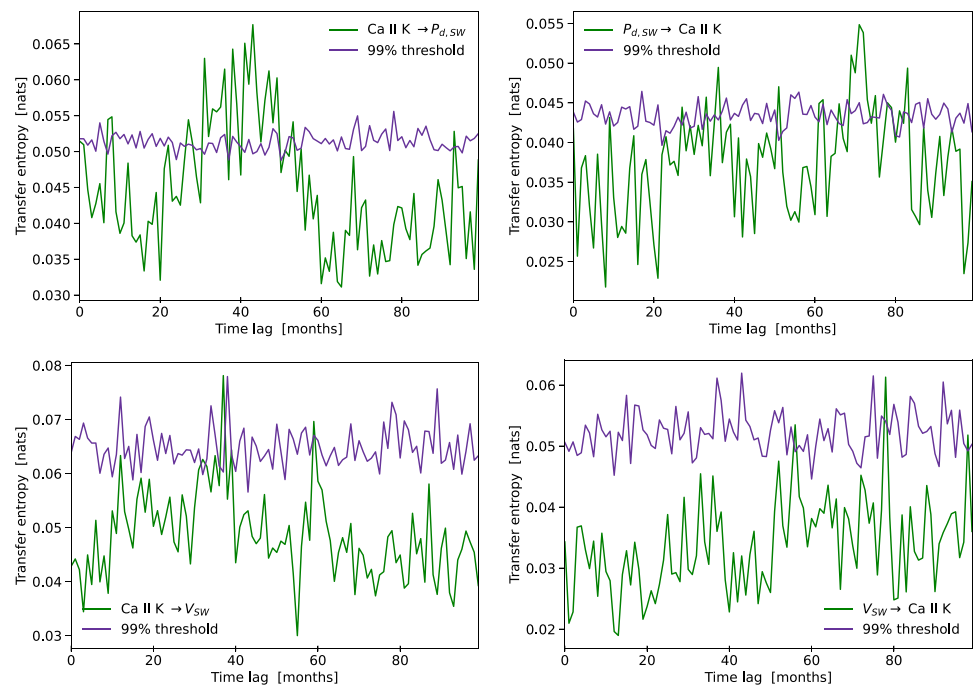
Considering the information flow from Ca II K index to solar wind speed, the transfer entropy shows a single peak that exceeds the 99% threshold, at time lag of 3.1-year. As in the case of the dynamic pressure, this results is consistent with the result recently found by Reda et al. (2023b). However, because it is a single peak, it may also be interpreted as a fluctuation.

Our results suggest that over the last five solar cycles, there is a better information flow from Ca II K index to solar wind dynamic pressure than from Ca II K index to solar wind speed. Since the dynamic pressure depends both on the speed and the density of the solar wind, thus making it an energy related parameter, we interpret this result as a phenomenon connected to energy transfer processes from the Sun to the heliosphere. The former result could be of interest to build up a predictive model in a Space Climate context.

We remark that, although the results found with the transfer entropy analysis are in agreement with the previous findings, further work is needed to assess the causal relationship. Indeed, due to technical limitations (e.g., few data points), we are not able to investigate thoroughly the statistical significance of our results. However, we emphasize that, at the moment, the application of the transfer entropy is promising and may be extremely helpful in the future to disentangle the (non-linear) causal relations between the solar activity and the solar wind at Space Climate scales. Dataset with a higher time resolution, thus with more data points, will be considered for a future analysis to verify this hypothesis and to confirm the result obtained via the transfer entropy presented in this work.

Declarations

References

Balasis G, Balikhin MA, Chapman SC et al (2023) Complex systems methods characterizing nonlinear processes in the near-earth electromagnetic environment: Recent advances and open challenges. Space Sci Rev 219(5):38

Barnard L, Lockwood M, Hapgood MA et al (2011) Predicting Space Climate change. Geophys Res Lett 38(16):L16103. https://doi.org/10.1029/2011GL048489

Bertello L, Pevtsov A, Tlatov A, et al (2016) Correlation Between Sunspot Number and Ca II K Emission Index. Solar Phys. 291(9-10):2967–2979. https://doi.org/10.1007/s11207-016-0927-9, arXiv:1606.01092 [astro-ph.SR]

Bossomaier T, Barnett L, Harré M et al (2016) Transfer entropy. Springer. https://doi.org/10.1007/978-3-319-43222-9_4

Chatzistergos T, Ermolli I, Krivova NA, et al (2019a) Analysis of full disc Ca II K spectroheliograms. II. Towards an accurate assessment of long-term variations in plage areas. *Astron. Astrophys.* 625:A69. https://doi.org/10.1051/0004-6361/201834402, arXiv:1902.07122 [astro-ph.SR]

Chatzistergos T, Ermolli I, Solanki SK, et al (2019b) Recovering the unsigned photospheric magnetic field from Ca II K observations. Astron. Astrophys. 626:A114. https://doi.org/10.1051/0004-6361/201935131, arXiv:1905.03453 [astro-ph.SR]

Donnelly RF, White OR, Livingston WC (1994) The solar Ca II K index and the Mg II core-to-wing ratio. Solar Phys. 152(1):69–76. https://doi.org/10.1007/BF01473185

El-Borie MA (2002) On Long-Term Periodicities In The Solar-Wind Ion Density and Speed Measurements During The Period 1973–2000. Solar Phys. 208(2):345–358. https://doi.org/10.1023/A:1020585822820

Flandrin P, Rilling G, Goncalves P (2004) Empirical Mode Decomposition as a Filter Bank. IEEE Signal Process Lett 11(2):112–114. https://doi.org/10.1109/LSP.2003.821662

Huang NE, Wu Z (2008) A review on Hilbert-Huang transform: Method and its applications to geophysical studies. Reviews of Geophysics 46(2):RG2006. https://doi.org/10.1029/2007RG000228

Huang NE, Shen Z, Long SR et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond Ser A 454(1971):903–998. https://doi.org/10.1098/rspa.1998.0193

Judge P (2006) Observations of the Solar Chromosphere. In: Leibacher J, Stein RF, Uitenbroek H (eds) Solar MHD Theory and Observations: A High Spatial Resolution Perspective, p 259

Katsavrias C, Preka-Papadema P, Moussas X (2012) Wavelet Analysis on Solar Wind Parameters and Geomagnetic Indices. *Solar Phys.* 280(2):623–640. https://doi.org/10.1007/s11207-012-0078-6, arXiv:1205.2229 [astro-ph.SR]

King JH (1979) Solar cycle variations in IMF intensity. J. Geophys. Res'. 84(A10):5938–5940. https://doi.org/10.1029/JA084iA10p05938

King JH, Papitashvili NE (2005) Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. Journal of Geophysical Research (Space Physics) 110(A2):A02104. https://doi.org/10.1029/2004JA010649

Köhnlein W (1996) Cross-correlation of solar wind parameters with sunspots ('Long-term variations') at 1 AU during cycles 21 and 22. Astrophys Space Sci 245(1):81–88. https://doi.org/10.1007/BF00637804

Li KJ, Zhanng J, Feng W (2016) A Statistical Analysis of 50 Years of Daily Solar Wind Velocity Data. Astron J 151(5):128. https://doi.org/10.3847/0004-6256/151/5/128

Li KJ, Zhang J, Feng W (2017) Periodicity for 50 yr of daily solar wind velocity. Mon. Not. Roy. Astron. Soc. 472(1):289–294. https://doi.org/10.1093/mnras/stx1904

Mursula K, Usoskin IG, Maris G (2007) Introduction to Space Climate. Adv Space Res 40(7):885–887. https://doi.org/10.1016/j.asr.2007.07.046

Neugebauer M (1981) Observations of Solar-Wind Helium. Fund. Cosmic Phys. 7:131–199

Ortiz A, Rast M (2005) How good is the Ca II K as a proxy for the magnetic flux? Mem. Soc. Astron. Italiana 76:1018

Pearl J (2009) Causality, 2nd edn. Cambridge University Press, https://doi.org/10.1017/CBO9780511803161

Petrinec SP, Song P, Russell CT (1991) Solar cycle variations in the size and shape of the magnetopause. J. Geophys. Res'. 96(A5):7893–7896. https://doi.org/10.1029/90JA02566

Reda R, Giovannelli L, Alberti T, et al (2021) Correlation of solar activity proxy with solar wind dynamic pressure in the last five solar cycles. Il Nuovo Cimento della Societa Italiana di Fisica 44 C:120. https://doi.org/10.1393/ncc/i2021-21121-7

Reda R, Giovannelli L, Alberti T (2023a) On the time lag between solar wind dynamic parameters and solar activity UV proxies. Advances in Space Research 71(4):2038–2047. https://doi.org/10.1016/j.asr.2022.10.011, arXiv:2210.07855 [astro-ph.SR]

Reda R, Giovannelli L, Alberti T, et al (2023b) The exoplanetary magnetosphere extension in Sun-like stars based on the solar wind-solar UV relation. Mon. Not. Roy. Astron. Soc. 519(4):6088–6097. https://doi.org/10.1093/mnras/stac3825, arXiv:2203.01554 [astro-ph.SR]

Richardson IG, Cane HV (2012) Solar wind drivers of geomagnetic storms during more than four solar cycles. Journal of Space Weather and Space Climate 2:A01. https://doi.org/10.1051/swsc/2012001

Rilling G, Flandrin P, Gonçalves P (2003) On empirical mode decomposition and its algorithms. Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03 3

Samsonov AA, Bogdanova YV, Branduardi-Raymont G et al (2019) Long-Term Variations in Solar Wind Parameters, Magnetopause Location, and Geomagnetic Activity Over the Last Five Solar Cycles. Journal of Geophysical Research (Space Physics) 124(6):4049–4063. https://doi.org/10.1029/2018JA026355

Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85:461–464. https://doi.org/10.1103/PhysRevLett.85.461

Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85(2):461

Schrijver CJ, Cote J, Zwaan C, et al (1989) Relations between the Photospheric Magnetic Field and the Emission from the Outer Atmospheres of Cool Stars. I. The Solar CA II K Line Core Emission. Astrophys. J. 337:964. https://doi.org/10.1086/167168

Siscoe GL, Crooker NU, Christopher L (1978) A solar cycle variation of the interplanetary magnetic field. Solar Phys. 56(2):449–461. https://doi.org/10.1007/BF00152484

Stumpo M, Consolini G, Alberti T et al (2020) Measuring Information Coupling between the Solar Wind and the Magnetosphere-Ionosphere System. Entropy 22(3):276. https://doi.org/10.3390/e22030276

Stumpo M, Benella S, Consolini G et al (2023) Dynamical information flow within the magnetosphere-ionosphere system during magnetic storms. Rend Fis Acc Lincei 34(1):1–9

Usoskin IG, Berdyugina SV, Moss D et al (2007) Long-term persistence of solar active longitudes and its implications for the solar dynamo theory. Adv Space Res 40(7):951–958. https://doi.org/10.1016/j.asr.2006.12.050

Venzmer MS, Bothmer V (2018) Solar-wind predictions for the Parker Solar Probe orbit. Near-Sun extrapolations derived from an empirical solar-wind model based on Helios and OMNI observations.

Rendiconti Lincei. Scienze Fisiche e Naturali (2024) 35:49–61

61

Astron. Astrophys. 611:A36. https://doi.org/10.1051/0004-6361/201731831, arXiv:1711.07534 [astro-ph.SR]

Wibral M, Pampu N, Priesemann V, et al (2013) Measuring information-transfer delays. PloS one 8(2):e55,809