NPG-2016-38. On the intrinsic time-scales of temporal variability in measurements of the surface solar radiation" by M. Bengulescu et al.

AUTHORS' RESPONSE

Dear Editor

The surface solar irradiance (SSI) is the main driver behind climate and weather. A literature survey reveals that its variability has not been investigated in much details. Our paper is a contribution to bridging this gap. It analyses the temporal variability of the SSI on a wide dynamic range, from two days to several years. Since the physical processes are non-linear and the data are non-stationary, we use the adaptive Hilbert-Huang Transform (HHT) to identify the characteristic time-scales of the signal. We found the presence of a high-frequency band (2-100 days) consisting of quasi-stochastic IMFs that have been shown to be amplitude-modulated by the yearly cycle, a low power spectral spectral band in the 100 days to 300 days region, and a well-defined spectral peak at the one year mark accounting for the yearly variability. Our findings are relevant to meteorology, climatology and the solar power industry, where forecasts of the SSI can be improved by e.g. modelling it as an output of dynamic systems with low-frequency oscillations and noise effects that interact non-linearly.

We have had two reviews and one comment. They were encouraging and we are proud of submitting a revision. We have written below a point-by-point response to each Reviewer. We have made a large number of slight editing changes in the text; there are not all reported in this point-by-point response. The only change in the structure follows a suggestion by Reviewer #2 for reverse the order of Sections 5.1 and 5.3, and 5.2 and 5.4. Editing changes have also been made at the suggestion of Mr Tommaso Alberti in his comment, to whom we answered separately.

This point-by-point answer is followed by a marked-up manuscript version. You will note a large number of deleted lines and new sentences in Section 5. It does not mean that we have rewritten entirely Section 5; it originates from the re-ordering of sub-sections 5.1 to 5.4. Please, take also note that due to a known technical limitation, track changes mode is disabled for bibliographic references in the text. Hence, they all appeared as new references.

We believe that we have answered to all comments by Reviewers in a satisfactory way.

The authors

ANSWER to Anonymous Referee #1

We thank the Anonymous Referee #1 for the review and encouraging comments.

MAJOR COMMENT by Referee #1: *"Only a major point needs to be explained and justified more clearly, or modified. This concerns a discrimination method between "signal" and "noise""*

ANSWER. We understand that we have created a confusion between "noise" and "stochastic component". This is spread throughout the text, including the term "weather noise" that we have borrowed from other authors (Chekroun, M. D., Kondrashov, D., and Ghil, M.:, Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation, Proceedings of the National Academy of Sciences, 108, 11 766–11 771, doi:10/bpt5kk, 2011). In their work on ENSO forecasting, these authors model the ENSO as a climate signal (slowly varying signal) influenced by fast (i.e. rapidly varying) processes that they called weather noise. This terming was certainly appropriate in their case but not in ours. Adopting this terming has created confusion somehow which is the basis of most of the following comments made by the Reviewer.

To remove this confusion, we have rewritten the text in some parts to avoid the use of "weather noise" and to make a clear distinction between the stochastic component created by physical processes and noise.

COMMENT. *"This needs to receive a much more precise definition of terms, because it seems that, for the authors, something stochastic is purely noisy and not relevant for the physics of the problem studied. If this is correctly understood by the reader, it is not correct, since stochastic variability possesses of course in general much more rich information than a pure noise."*

ANSWER: We are fully aware of this and we have clearly stated in Section 5.3: "It will be subsequently shown that, for the first five <u>IMFs</u> at least, this is indeed the case; although (quasi-)stochastic in nature, they are not completely devoid of information."

COMMENT on section 3.2. *"The procedure which is applied here to separate what is assumed to be "noisy" and deterministic information, is explained in section 3.2. The main idea is to state that "noisy" parts of the signal generate dyadic filtering in the EMD method, and a detection method based on this property is applied here. This is problematic because if white noise or fractional Brownian motion have been shown to generate EMD modes which are dyadic, the reciprocal is wrong, many studies have found the dyadic property for stochastic processes and also for observed data, that are not noises. The problem seems here the confusion by the authors between noise and random processes."*

ANSWER: We are well aware of this. This is why, in the original section 5.3, we have stated: "At this point, several precautionary notes are compulsory. First, the rule of inference used here is *modus tollens*, i.e. the results from figure 9 do not imply that the modes who experience down-shift in their SWMFs are made up of pure noise."

In addition to our answer, we may mention a comment made in the interactive discussion by Dr. Tomasso Alberti, who fully agrees and even appreciates our approach. He wrote "*I think that the null-hypothesis test proposed in Section 3.2 is a simple but powerful test to investigate the noise-*

*like existence of IMFs. If I do not misunderstood, this is particularly suitable when the EMD really acts as a dyadic filter.*"

COMMENT on section 5.3: *"The same confusion is visible in section 5.3, lines 17-19 and line 25. All this methodology and the discussion in section 5.3 must be changed or suppressed"*

ANSWER. The Section 5.3 "Discriminating signals from noise in the IMFs" has an interest because of the underlying question of the significance of each IMF. It is important to ascertain whether an IMF results from a physical phenomenon, possibly of stochastic nature, or from noise. This section has been reformulated to avoid the confusion mentioned by the Reviewer. We have also changed the title accordingly; it is now "Discriminating deterministic signals from stochastic components in the IMFs".
Please, take also note that following suggestions by Reviewer #2, the Section 5.3 is now Section 5.1.

COMMENT. *"the Hilbert-Huang marginal power spectrum of the data given by equation (8) should be displayed, for some locations and also globally"*

ANSWER. This comment is partly unclear to us as we have not used the term "Hilbert-Huang marginal power spectrum". Eq. 8 defines the Hilbert marginal spectrum which is displayed for one (BOU) of the four studied locations.
We read this comment as a suggestion to add three Figures similar to Figure 6. If we do this, we are facing a major increase in the length of the paper. Displaying the four Figures needs room and they should be accompanied by comments. To illustrate this, we have excerpted 4 pages (pages 57-60) from the PhD thesis of Marc Bengulescu (defended in July 2017) that comprise the four suggested graphs (Fig. 4.3) and the associated comments. These pages are reproduced here. Please, skip the first paragraph of the first page.

With the scrutiny of the these low frequency components, the discussion of the time-scale distribution of the IMFs from figure 4.1 can now be concluded. However, as previously mentioned, this particular illustration, although instructive, is incomplete. First, the box plot representation does not take the instantaneous variations of frequency into account, but renders global aggregates instead – much like the traditional Fourier methods, with the interquartile range spread in addition. This is done on purpose, with the intent of making it easier for the readership not accustomed to the HHT to create analogies with the more familiar methods (e.g. Fourier analysis, wavelets, etc.). Second and last, this particular representation is totally devoid of any information pertaining to the local amplitude, or power, or variance, of the data. With these consideration in mind, the Hilbert spectra of the data, a true time-frequency representation for non-linear and non-stationary data, will be discussed next.

Figure 4.3 depicts the Hilbert spectra of the four BSRN time-series: BOU, CAR, PAY and TAT. These spectra are similar to the one already introduced in figure 3.11, with the exception that the Hilbert marginal spectrum is no longer expressed in decibels.

The BOU Hilbert spectrum from the top left panel of figure 4.3 exhibits a high-frequency feature between 2 days and $\backsim$ 100 days, which corresponds to the first five IMFs of the time-series. The instantaneous time-scales of these modes overlap (figure 4.1), hence the appearance on the Hilbert spectrum of a continuum instead of distinct bands. This spectral feature has relatively low power, that decreases with increasing period, as can be inferred from the sloped dent in the marginal Hilbert spectrum corresponding to this region. In the 2 days to 32 days band, amplitude modulation by the yearly cycle can be inferred from the periodic change in color, with yellow-green tones, occurring mostly during the high irradiance regime of summer, that turn blue during the winterly minima. Next, in the band between 100 and 300 days, a gap in the spectrum is apparent, as can also be inferred from the lack of support in this region for any of the BOU IMFs in figure 4.1. The yellow trace, corresponding to IMF6, exhibits frequency modulation around the one year period, seen as oscillations in the range of 300 to 450 days, which is also the support of this mode in the box plot of IMF time-scales. The colour of this IMF indicates that it has the highest power of all the components, as can also be inferred from the large peak on the marginal spectrum. The corresponding time-scale fluctuations are centred in 365 days, and frequency modulation is greatest during 2003 through 2005. From 2006 onwards, however, frequency modulation is less pronounced – perhaps capturing the low solar activity around the 2008 minimum in the eleven year cycle solar cycle [Hathaway, 2015]. The final two low-frequency, blue-green traces on the spectrum correspond to IMF7 and IMF8. For IMF7, mode mixing is apparent through the occasional sharing of the yearly time-scale band with IMF6, between mid-2003 and 2005. IMF7 has such low power that it fails to leave an imprint on the marginal spectrum and it seems to suddenly spring

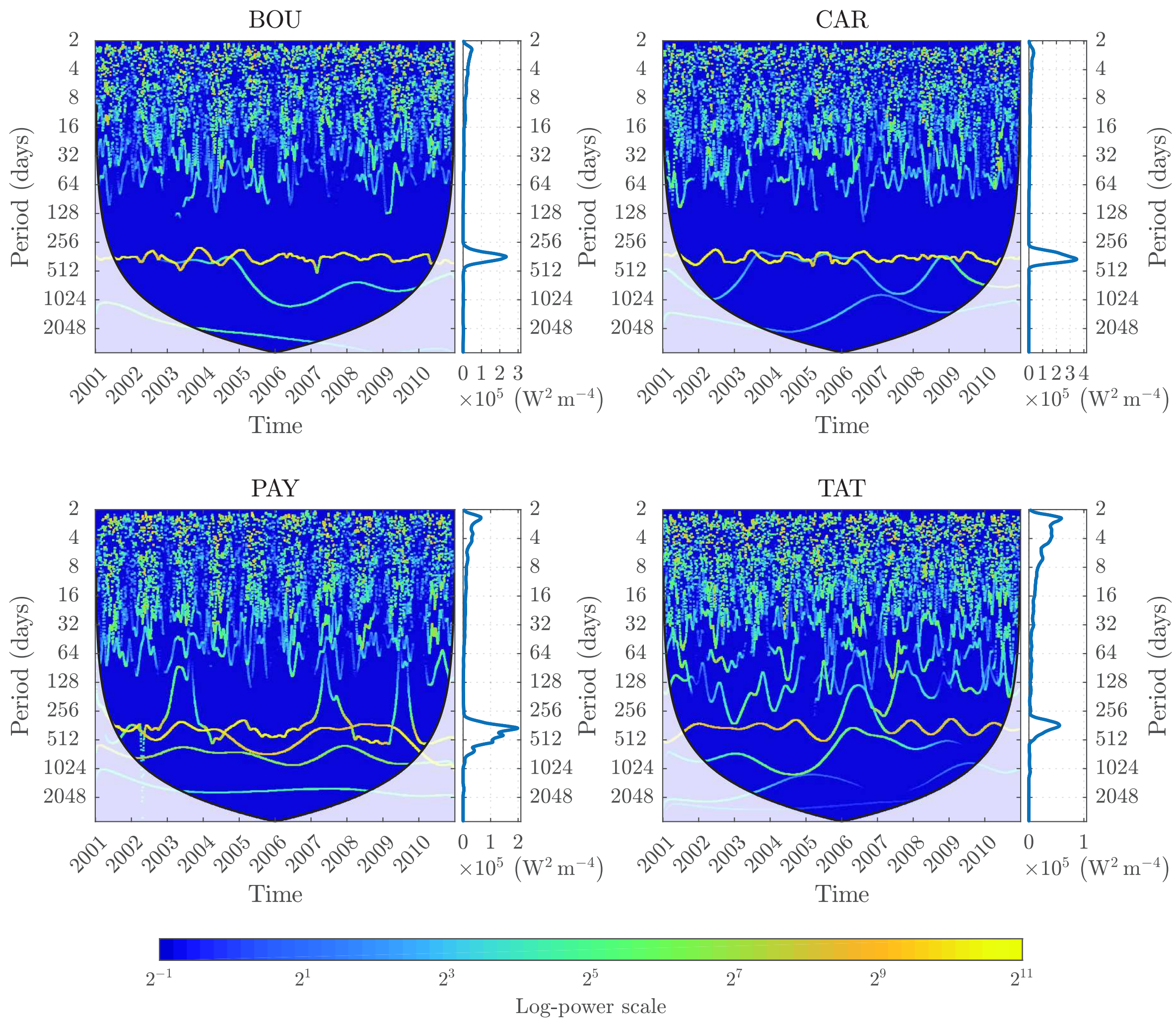


**Figure 4.3** The Hilbert energy spectra of the 10-year BSRN time-series, spanning 2001 through 2010; clockwise from top left: BOU, CAR, TAT, PAY. Pixel colour encodes power (logarithmic scale colour bar at the bottom) at each time (abscissa) and each scale (ordinate). Time markers on the abscissa denote the start of the corresponding year. The white-out area indicates the regions where edge effects become significant. The Hilbert marginal spectra in the panels on the right indicate the amount of power at each scale.

into existence during summer 2003, i.e. it has negligible amplitude during the first two and a half years. IMF8 starts out in light-green hues and slowly vanishes around 2007. This last mode for BOU has a median period of 1500 days (see figure 4.1), with most of its power lying within edge effect territory; interpretation of this feature is thus ambiguous at best.

Referring to the CAR Hilbert spectrum, depicted in the upper right panel of figure 4.3, some features can be identified. A high-frequency plateau between 2 days and ∽ 100 days is notable, corresponding to first five IMFs of the time-series. As can be seen in figure 4.1, the supports of these modes overlap, thus the appearance on the Hilbert spectrum of a contiguous plateau, as is the case for BOU. Also similar to BOU, the power of this feature is low, manifested by a slight indentation on the marginal Hilbert spectrum for this band. A cyclic shift in color can also be observed in this band, which points to an amplitude modulation mechanism linked to the seasonal cycle – darker blue tones during winter turn yellow-green during summer. Similar to BOU, between 100 and 300 days a gap in the spectrum is again manifest, as hinted at by the time-scales of the CAR IMFs in figure 4.1, which do not cover this band. IMF6, in the form of the yellow trace oscillating around the one year period, in the range of 300 to 450 days, resembles the same mode for BOU. This component has the greatest power, denoted by the large peak on the marginal spectrum. It exhibits frequency modulation around 365 days, with shifts towards greater frequencies taking place predominantly during 2002 through 2006. Between 2007 and 2009, however, the frequency modulations are less pronounced – corresponding to a period when solar activity is at a minimum. The last two large time-scale, blue-green traces denote IMF7 and IMF8. The seventh exhibits some mode mixing with IMF6, between the end of 2003 and the first half of 2006, and shortly again before 2009. These two last components have such low power that they fail to leave a mark on the marginal spectrum.

In the Hilbert spectrum for the PAY data, shown in the lower left panel of figure 4.3, the contiguous high-frequency plateau between 2 days and $\backsim$ 100 days is also present, unsurprisingly, since the first five IMFs of the time-series closely resemble those for the CAR data. The power in this band, however, is slightly greater than for BOU or CAR, especially considering the predominance of yellow hues in the 2 – 4 days region that corresponds to the first and second IMFs. This is also apparent when looking at the marginal spectrum, where a distinct peak at this time-scale can be clearly made out. The amplitude modulation phenomenon in phase with the seasonal variations, previously identified in the BOU and CAR data, is even more pronounced in the PAY spectrum; once again dark blue tones that occur during low insolation in winter turn green and even yellow during the high irradiance regime of summer. Here too, between 100 and 300 days a gap in the spectrum is also apparent, with the notable exception of the mode mixing phenomena associated with IMF6 that occur during 2003, 2007 and 2009. As previously shown in figure 4.1, the lower range of the frequency distribution of the sixth mode overlaps the high frequency plateau, which is portrayed by the three jutting spikes from the yearly band into the sub-100 days plateau on the PAY spectrum. Since these protruding filaments have such low power that they leave no imprint on the marginal spectrum, their most probable origin can be attributed to some sort of numerical artefacts. The yearly variability of the data can be, unsurprisingly, identified with the sharp peak at roughly 365 days on the marginal spectrum. In terms of the Hilbert time-frequency-energy representation, however, the seasonal cycle cannot be attributed to one sole component. This is due to mode mixing as seen in figure 4.1 where the range of IMF7 completely overlaps that of IMF6. It could be argued that IMF6 should represent the "true" seasonal cycle, as indicated by the median of its frequency distribution, however for 2003 and 2007 the Hilbert spectrum reveals that whenever IMF6 extends its tendrils into the high-frequency plateau and drastically reduces its power, IMF7 seems to “pick up the slack” by reaching into the vacated one year band. During 2008, and briefly during spring 2004, it is found that the two components trade places altogether, with IMF6 assuming lower values that IMF7 on the frequency scale. The last two low-frequency components, IMF8 and IMF9, can be seen here too to have relatively low power – both only just manage to make a very slight indentation on the marginal spectrum. For IMF9, the frequency spread in the Hilbert representation is in good agreement with its narrow support from figure 4.1.

The Hilbert spectrum for the TAT data, shown in the lower right panel of figure 4.3, also shows the contiguous high-frequency plateau between 2 days and $\backsim$ 100 days, owing to the first five IMFs of the data closely resembling those for BOU, CAR and PAY data. However, the power found in this band is much greater than for either BOU, CAR, or PAY, especially in the 2 – 4 days region denoting the first two IMFs. For these time-scales of TAT, the marginal spectrum shows a peak that is even greater than the one associated with the yearly cycle. This feature is also evident in figure 4.2, where the upper whisker of the amplitude of IMF1 is seen to extend beyond 150 $\mathrm{W\,m^{-2}}$, whereas it is less than 80 $\mathrm{W\,m^{-2}}$ for annual cycle depicted by IMF7. As a result, the slanting of the high-frequency plateau is clearly visible on the marginal spectrum, with a somewhat lesser slope that for the other stations. The amplitude modulation phenomenon in phase with the seasonal variations, also identified for the other datasets, is present here too, although to a lesser extent than for PAY. Unlike the previous datasets, no gap can is evident in the spectrum, owing to a sixth IMF that covers the region between 50 and 300 days, as shown in figure 4.1. The yearly variability of the time-series is denoted by the spectral peak at roughly 365 days on the marginal spectrum. In the Hilbert time-frequency representation the seasonal cycle is attributed to the dark yellow trace of IMF7, which can be seen in figure 4.1 to have a median time-scale of 366.6 days. Mode mixing in the yearly band is nevertheless apparent between the second half of 2005 through 2007, when IMF8 approaches the 365 days mark while IMF7 protrudes below 256 days. The last two low-frequency components, IMF9 and IMF10, have reduced power and fail to leave a mark on the marginal spectrum.

So far, all time-series have been shown to share a high-frequency constituent between 2 days and 100 days composed of five IMFs with mean periods following a dyadic sequence, and an IMF around 365 days that captures the yearly variability. For BOU, CAR and PAY, a low power region can be found in the 100 days to 300 days band. Beyond the one year time-scale, the low-frequency variability in the 1.5 years to 6 years band is captured by another two (BOU and CAR) or three (PAY and TAT) components. The TAT data is the only time-series that has an IMF in the low power band between the high-frequency feature and the yearly cycle (median period 143.2 days).

ANSWER to Anonymous Referee #2

We thank the Anonymous Referee #2 for the review and encouraging comments.

MAJOR COMMENT *"Primarily I believe it is important to establish the signal/noise status of the components before discussing their physical origin i.e. sections 5.3 and 5.4 should be placed before sections 5.1 and 5.2. These sections then question the validity of linking the various components to features observed in solar data e.g. the discussion of the high frequency components with solar rotation, which appear to be due to noise and the dyadic properties of EMD."*

ANSWER. We thank the Reviewer for this suggestion. We agree that it facilitates the reading of this Section. This was done. Note that we have renamed Section 5.1 in "Discriminating deterministic signals from stochastic components in the IMFs".

COMMENT. *"Along the same lines in Section 5.3 it is stated that 'unambiguous interpretations of QBO-like components seems to be out of reach' and yet the authors still discuss the possibility that it could be related to the solar QBO. If the authors insist on including this discussion I believe the terrestrial QBO should also be mentioned as this also has a well know impact on weather on Earth, such as the severity of winters, which would also affect cloud cover. However, it is my opinion that the authors should either not try and make any conclusions concerning the QBO or at least stress that with the current analysis they cannot be sure that this is a real signal.*
*Finally with regards to the QBO I believe that the link between galactic cosmic rays and cloud coverage is still highly debated and so I would either remove the comment concerning this or refer to papers concerning the debate."*

ANSWER. This paragraph has been rewritten and is now:
Lastly, the components indicative of low-frequency variability on time-scales greater than one year are discussed. The intrinsic time-scales found in these IMFs seem to match once more those pertaining to the so-called quasi-biennial oscillations that have been observed in solar activities and proxies with periodicities between 0.6 and4 years (Bazilevskaya et al., 2015; Kolotkov et al., 2015; Vecchio et al., 2012), as well in meteorological data like Harrison (2008) who identifies a 1.68 year peak in cloud cover or high-latitude stratospheric temperatures and geopotential heights (Labitzke and Loon, 1988). Nevertheless, within the scope of the current analysis, the interpretation of these low frequency variability components as as a real, possibly QBO-like, signal is uncertain.

MINOR REMARKS.
Thank you for spotting these points. All of them have been taken into account and the text was rewritten accordingly.

# On the intrinsic time-scales of temporal variability in measurements of the surface solar radiation

Marc Bengulescu, Philippe Blanc, and Lucien Wald

MINES ParisTech, PSL Research University, Centre for Observation, Impacts, Energy
CS 10207 - 06904 Sophia Antipolis Cedex, France

*Correspondence to:* M. Bengulescu (marc.bengulescu@mines-paristech.fr)

**Abstract.** This study is concerned with the intrinsic temporal scales of the variability of the surface solar irradiance (SSI). The data consist of decennial time-series of daily means of the SSI ~~spanning ten years,~~ obtained from high quality measurements of the broadband solar radiation impinging on a horizontal plane at ground level, issued from different Baseline Surface Radiation Network (BSRN) ground stations around the world. First, embedded oscillations ~~roughly sorted by ranges~~ sorted in terms of increasing time-scales of the data are extracted by empirical mode decomposition. Next, Hilbert spectral analysis is applied to obtain an amplitude modulation – frequency-modulation (AM–FM) representation of the data. The time-varying nature of the characteristic time-scales of variability, along with the variations of the signal intensity, are thus revealed. A novel, adaptive null-hypothesis based on the general statistical characteristics of noise is employed, in order to discriminate between the different features of the data, those that have a deterministic origin and those being realisations of various stochastic processes. The data has a significant spectral peak corresponding to the yearly variability cycle and features quasi-stochastic high-frequency ~~"weather noise"~~variability components, irrespective of the geographical location or of the local climate. Moreover, the amplitude of this latter feature is shown to be modulated by variations of the yearly cycle, indicative of non-linear multiplicative cross-scale couplings. The study has possible implications on the modelling and the forecast of the surface solar radiation, by clearly discriminating the deterministic from the quasi-stochastic character of the data, at different local time-scales.

**Keywords.** solar radiation; temporal variability; Hilbert-Huang transform; empirical mode decomposition; Baseline Surface Radiation Network (BSRN); fractional re-sampling; stochastic components.

## 1 Introduction

The power of the electromagnetic radiation from the Sun that reaches the surface of the Earth is estimated at around $10^{17}$ W. Thus, solar irradiance is the main driver behind the weather and climate systems on the planet. As such, the Global Climate Observing System (GCOS) program has identified the surface solar irradiance (SSI) as an Essential Climate Variable that helps understand climate evolution and guides adaptation and mitigation efforts (Bojinski et al., 2014). Long term time-series of the SSI are instrumental in engineering and finance by enabling, e.g. the optimal determination of geographical sites for solar power plants and guiding investment decisions, respectively (Schroedter-Homscheidt et al., 2006). Thus, better knowledge of the SSI and of its temporal variability, as recorded in long term time-series, is one of the intents of this work.

Temporally, the SSI exhibits a very wide dynamic range. Its short-term time-scales of variability, such as clouds briefly obscuring the Sun, are ~~measured in~~ observed on seconds. At the opposite scale thousands or even millions of years are to be used, as related to the change of the orbital parameters of the Earth-Sun system or to stellar evolution (Beer et al., 2006). In spite of this large span of characteristic scales of temporal variability, most of the studies dealing with this physical quantity have focused primarily on a few selected time-scales of interest. As such, reports have either dealt with global averages and long-term trends (Trenberth et al., 2009; Wahab et al., 2010; Pachauri et al., 2014; Blanc et al., 2015), have only scrutinized the short-term, high-frequency variability (Yordanov et al., 2013; Lauret et al., 2016), or have focused exclusively on a few intermediate scales (Coskun et al., 2011; Medvigy and Beaulieu, 2012). Although considerably differing in methods, taken together the previously cited studies add valuable contributions to our knowledge of the SSI. But is it possible to analyse the variability of the SSI across multiple time-scales in a unitary way?

To do so, first a decomposition of the time-series into uncorrelated sub-constituents with distinct characteristic time-scales should be preferred. Analysis would then ensue in a like manner for each scale. The time-scales, or characteristic periods of a time-series can be identified with the inverse of the frequency at which the processes that generate the data occur. It then follows that methods portraying the changes of the spectral content of a time-series with respect to time are potentially good candidates. This would enable both the identification of the periodicities and of the dynamic evolution of the processes generating the data. A general class of useful signal processing techniques can thus be identified in the so-called time-frequency distributions, that depict the intensity (or energy) of a signal in the time and the frequency domains simultaneously (Cohen, 1989). Such methods are commonly employed for geophysical signal processing (Tary et al., 2014).

Another factor to be taken into account are the non-linear and non-stationary characteristics of the measured solar radiation data (Zeng et al., 2013). Handling such data issued from the non-linear interaction of physical processes, often also found under the influence of non-stationary external forcings calls for an adaptive data analysis approach (Wu et al., 2011).

The study at hand will make use of the Hilbert-Huang Transform (HHT), an adaptive, data-driven analysis technique designed specifically for investigating non-linear and non-stationary data (Huang et al., 1998). The HHT adaptively decomposes any dataset into basis functions that are derived solely from the local properties of the time-series. A time-frequency-energy representation of the data is then constructed from these basis functions. The HHT has seen extensive use in geophysical signal analysis and spectral estimation (Solé et al., 2007; Huang and Wu, 2008; Vecchio et al., 2010; Lee and Ouarda, 2011; Alberti et al., 2014; Huang and Shen, 2014; Tary et al., 2014). The HHT has also been previously employed on SSI datasets (Duffy, 2004; Calif et al., 2013; Bengulescu et al., 2016a, b). A similar technique has been independently proposed by Nagovitsyn (1997) for the analysis of the non-linear, non-stationary, long range solar activity. In this light, the use of the HHT for the study of the temporal variability of the SSI appears to be appropriate. The inner workings of this data processing method will be detailed in a dedicated subsection.

Regardless of the methods used, when analysing data ~~a question always needs to be addressed, namely: what is "signal" and what is "noise"? More precisely,~~ there is always the need to discriminate between deterministic signals and what are assumed to be ~~stochastic realizations of a noisy background~~ background stochastic realizations (Rios et al., 2015). The classical way to solve this when employing the HHT on geophysical signals, such as the SSI, is to presume some model for the background power spectrum, against which the identified features are then compared (Huang and Wu, 2008; Franzke, 2009, 2012). In contrast, the present study parts with the traditional approach, by adopting a novel, adaptive null-hypothesis that requires no *a priori* knowledge of the ~~background noise~~nature of the background processes, introduced by Chen et al. (2013); further discussion thereof will be presented in due course. A somewhat similar objective can be found in the work of Rios and de Mello (2016), though their method of discrimination between stochastic and deterministic components is fundamentally different. Kolotkov et al. (2016) also propose a method for discriminating frequency–dependent ~~noise~~ stochastic components by empirically estimating their power law spectral energy distribution and respective confidence bounds. Approaches for discriminating high-frequency fluctuations from large time-scale modulations are also described by Flandrin et al. (2004b) and Alberti et al. (2016).

At this point, the general outline of our study can be summarized as follows. We analyse measurements of daily means of SSI at different geographical locations. We focus on identifying and analysing the intrinsic modes of the temporal variability of the SSI, as revealed by the HHT. We also investigate the physical and statistical significance of these modes. We show that the HHT is able to discriminate between a deterministic yearly cycle and ~~a~~ multiple high-frequency (quasi-)stochastic ~~part, that we termed "weather noise" following . Although the weather noise appears to be random in nature, we nevertheless~~ components. We also find a non-null, statistically significant rank correlation between ~~its amplitude envelopes~~ the amplitude envelopes of the high-frequency scales and the yearly cycle. We then discuss the possible implications of our findings on the modelling and forecast of the SSI.

The study is organised as follows. Section 2 discusses the data sources and the preprocessing. In section 3 the adaptive data analysis approach will be described. Section 4 will present the results obtained, with the discussion thereof being deferred to section 5. Conclusions and outlook are presented in section 6. Code and data availabilities are indicated in sections 7 and 8, respectively. Lastly, acknowledgements and a bibliographical list conclude the study.

## 2 Data sources and pre-processing

The data under scrutiny in this study consist of ten-year time-series of daily means of SSI obtained from high-quality measurements performed at four different locations (table 1 and figure 1). The measurement stations are part of the Baseline Surface Radiation Network (BSRN), a worldwide radiometric network providing accurate readings of the SSI at 1 min temporal resolution and with an uncertainty requirement at 5 $\mathrm{W\,m^{-2}}$ (Ohmura et al., 1998).

The four time-series for the period 2001–2010 have been quality checked according to Roesch et al. (2011). Next, daily means of SSI were then calculated from these raw time-series only if more than 80% of the data during daylight were valid. Lastly, any isolated missing daily means were completed by linear interpolation applied to the daily clearness index, $K_T$, which is the ratio between the daily mean of SSI and the daily mean of the total solar irradiance received on a horizontal surface at the top of atmosphere for the same geographical coordinates.

Two measuring stations are located in Europe, one in Japan, and one in North America in order to capture various climatic conditions. Boulder (herefater abbreviated in BOU) experiences a mid-latitude steppe, cool type of climate (Köppen-Geiger:

BSk), while at Carpentras (abbreviated in CAR) the climate is a humid subtropical, Mediterranean one (Köppen-Geiger: Csa). Both sites experience many sunny days during the year. As a rule of thumb, $K_T$ equal to 0.2–0.3 denotes cloudy, overcast conditions, while $K_T$ around 0.7 indicates sunny conditions. Figure 2 exhibits the histograms of $K_T$ for the four stations. One may observe the high frequencies of the greatest values of $K_T$ for BOU and CAR. The median $\widetilde{K}_T$ is equal to 0.63 for both BOU and CAR, which means that half of the days exhibit $K_T$ greater than 0.63. The climate in Payerne (PAY) is classified as marine west coast, mild (Köppen-Geiger: Cfb), and Tateno (TAT) has a humid subtropical, east coast climate (Köppen-Geiger: Cfa). Compared to BOU and CAR, PAY and TAT exhibit more uniform histograms, with less days with cloud-free conditions, and experience more overcast and broken clouds conditions. The median $\widetilde{K}_T$ is equal to 0.47 for PAY and 0.51 for TAT. Except for TAT, which is embedded in an urban setting, the stations are located in rural environments; the local topography for BOU and TAT is flat with grassy surfaces, while for CAR and PAY the area is hilly with cultivated surfaces (BSRN, 2015).

Any further reference to seasons and seasonal phenomena shall be understood as occurring in the northern hemisphere since the stations are situated at boreal latitudes.

## 3 Adaptive data analysis

Ideally, data analysis methods should require no assumptions to be made about the nature of the scrutinised time-series, i.e. neither linearity, nor stationarity should be presumed. This is because the true character of the underlying processes that have generated the data is usually not known beforehand. Adaptivity to the analysed data would also be a sought after feature, in the sense of not imposing a set of patterns against which data would be decomposed, but rather letting the data itself drive the decomposition. This latter criterion ensures both that the extracted components carry physical meaning, and that the influence of method-inherent mathematical artefacts on the rendered picture of temporal variability is kept to a minimum (Wu et al., 2011). Since such a decomposition is only determined by the local characteristic time scales of the data, its appropriateness to non-linear and non-stationary time-series analysis is immediate (Huang et al., 1998).

### 3.1 The Hilbert-Huang Transform

The Hilbert-Huang Transform (HHT) is an adaptive data analysis technique built with the previous considerings in mind. It involves two distinct steps, the empirical mode decomposition followed by Hilbert spectral analysis. In-depth discussion of each step is carried out within the dedicated subsections that follow.

#### 3.1.1 The Empirical Mode Decomposition

The first step of the HHT is the empirical mode decomposition (EMD), an algorithmic procedure in essence, by which oscillations that present a common local time-scale are iteratively extracted from the data. These oscillatory components of the data are called Intrinsic Mode Functions (IMFs). An IMF is any function that satisfies two criteria: (1) its number of extrema and zero crossings differ at most by one; and (2) at any data point, the mean value of its upper and lower envelopes is zero. These two properties ensure that IMFs have a well behaved Hilbert transform (Huang et al., 1998). Owing to the adaptive nature of the EMD, the IMFs represent the basis functions onto which the data is projected during decomposition. This is in contrast with the Fourier or wavelet transforms where the basis functions are fixed in advance (Huang and Wu, 2008). Once all the IMFs have been extracted, all that is left of the time-series is a residue, or trend, which cannot be mathematically thought of as an oscillation at the span of the data. A sketch of the EMD algorithm is provided in algorithm 1.

**Algorithm 1** EMD

**Require:** $x(t) \in \mathbb{R}$

1: Initialize IMF counter: $k \leftarrow 0$
2: Initialize residual: $r(t) \leftarrow x(t)$
3: **while** $r(t)$ is not monotonic **do**
4:     Increment IMF counter: $k \leftarrow k+1$
5:     (Re)process residual: $h(t) \leftarrow r(t)$
6:     **while** $h(t)$ is not an IMF[a] **do**
7:         Find minima and maxima of $h(t)$
8:         Interpolate minima to find lower envelope: $L(t)$
9:         Interpolate maxima to find upper envelope: $U(t)$
10:         Find mean of envelopes: $m(t) \leftarrow (L(t)+U(t))/2$
11:         Remove mean of envelopes: $h(t) \leftarrow h(t)-m(t)$
12:     **end while**
13:     Store IMF: $c_k(t) \leftarrow h(t)$
14:     Update residual: $r(t) \leftarrow r(t)-c_k(t)$
15: **end while**
16: **return** $c_{1\ldots N}(t), r(t)$

[a]See text for the definition of IMF. This "sifting" loop should be run approximately 10 times (Wu and Huang, 2009, 2010).

Lines 6–12 of the EMD algorithm represent the so-called "sifting loop" which has a two-fold purpose — to discard any riding waves and to render the IMFs more symmetric. The stoppage criterion for the sifting loop is closely related to how the latter controls the filter character of the EMD. On the one hand, an infinite number of sifting iterations would asymptotically approach the result of the Fourier decomposition (i.e. constant amplitude envelopes) (Wang et al., 2010). On the other hand, several studies performed on time-series of pure noise (Flandrin and Gonçalves, 2004; Flandrin et al., 2004a; Wu and Huang, 2004) have shown the decomposition behaves like an adaptive "wavelet-like" dyadic filter if the number of sifting iterations is kept small, around 10, which also assures maximum component separation and minimum leakage (Wu and Huang, 2010). This stoppage criterion of 10 sifting iterations is currently the recommended one for practical applications (Wu and Huang, 2009, 2010) and is also the one employed in the study.

It also worth noting that the preferred interpolation method in the EMD, i.e. lines 8 and 9 of algorithm 1, are cubic splines (Rilling et al., 2003). Because of oscillations of the these interpolating splines edge effects may appear in the EMD, but are usually contained within a half-period of a component at data boundaries (Wu et al., 2011).

One of the drawbacks of the original EMD is that it may introduce a phenomenon known as "mode mixing". This is the manifestation of oscillations with dissimilar time-scales in the same IMF, or the presence of oscillations with similar time-scales in different IMFs. A workaround was proposed by Wu and Huang (2009) with ensemble empirical mode decomposition (EEMD). The idea was to run the decomposition over an ensemble of copies of the original signal to which white Gaussian noise has been added, with the final result obtained by averaging. Although the EEMD improved the mode mixing problem, the different sums of signal and noise produced different numbers of modes, making the final averaging somewhat difficult. Added to this, the reconstructed signal still contained some residual noise, and thus was not identical to the original. To overcome this situation, Torres et al. (2011) have proposed another iteration of the EMD, the complete EEMD with adaptive noise (CEEMDAN). This method also decomposes the white noise into modes, along with the signal, such that at each stage of the decomposition a particular noise is added and a unique residue is computed to obtain each mode. However, the modes of CEEMDAN still contain some residual noise and sometimes spurious modes appear in the early stages of the decomposition. The next iteration of the method, the improved complete ensemble EMD (ICEEMD or ICEEMDAN), overcomes these issues by fixing the signal to noise ratio for all stages of the decomposition process (Colominas et al., 2014). The ICEEMD method is that used in this study. In addition, a fast EMD routine provided by Wang et al. (2014) has been used to decrease the computation time.

To illustrate the workings of the EMD, the eight IMFs of the BOU time-series are presented in figure 3 in the order they were obtained, from top to bottom. As EMD operates in time-domain, the IMFs have the same temporal support as the original data and, by construction, ~~an average of zero and symmetrical~~ upper and lower amplitude envelopes that are symmetrical with respect to zero. It can be observed in figure 3 that as the decomposition progresses, the time-scale of the IMFs increases, i.e. the intrinsic oscillations are getting further spaced apart with increasing IMF number. Another view of this is brought by figure 4, where the power spectral density (PSD) and a Fourier estimate of the mean period of each IMF are plotted. To aid the reader, the colours used to portray the individual IMF spectra are the same as for the time-domain representation from figure 3. The spectral shapes of the IMF1…IMF5 are similar in form, i.e. bell curves, and their median periods roughly follow a dyadic scale, i.e. doubling with increasing IMF number as : 3.1 days $\rightarrow$ 7.3 days $\rightarrow$ 13.9 days $\rightarrow$ 30.5 days $\rightarrow$ 54.0 days. This doubling of the time-scale for these first five IMFs is the hallmark output of an efficient dyadic filter. Subsequently, it will be shown that this dyadic repartition is pertinent to identifying deterministic signals from random realizations of quasi-stochastic background processes. This finding is even more interesting, since the median periods have been estimated with a Fourier-based method, which measures period *globally* over the whole time range of the IMFs. By opposition, a measure of *local* period in the Hilbert sense is a much better estimate, since it has an accuracy as low as a quarter wavelength of temporal resolution with respect to the average time-scale of the IMF (Huang et al., 2009).

### 3.1.2 Hilbert Spectral Analysis

Once the empirical mode decomposition is completed, the second and last step of the HHT consists in the Hilbert spectral analysis of the previously obtained IMFs. Each IMF and its Hilbert transform are used to construct a complex analytic signal, described by an amplitude modulation - frequency modulation (AM–FM) model. This decomposition into two time-varying parts corresponding respectively to instantaneous amplitude and instantaneous frequency is very useful for the purpose of this study. It enables the identification, in a time-varying sense, of how much power (i.e. the square of amplitude) occurs at which time-scale (i.e. the inverse of frequency).

The Hilbert transform of each real-valued IMF $c_k(t)$ can be written as:

$$\sigma_k(t) = \mathcal{H}(c_k(t)) = \frac{1}{\pi} P \int\limits_{-\infty}^{\infty} \frac{c_k(\tau)}{t-\tau} \, \mathrm{d}\tau \tag{1}$$

where subscript $k$ designates the $k$-th IMF, and $P$ indicates the Cauchy principal value. From each IMF and its Hilbert-transformed version, a unique complex-valued analytic signal can be obtained (Gabor, 1946):

$$z_k(t) = c_k(t) + i \cdot \sigma_k(t) = a_k(t) \cdot e^{i \cdot \theta_k(t)} \tag{2}$$

in which

$$a_k(t) = \sqrt{c_k^2(t) + \sigma_k^2(t)} \tag{3}$$

is the instantaneous amplitude and

$$\theta_k(t) = \tan^{-1}\left(\frac{\sigma_k(t)}{c_k(t)}\right) \tag{4}$$

is the instantaneous phase. The instantaneous frequency is the first time derivative of the instantaneous phase:

$$\omega_k(t) = \frac{1}{2\pi} \frac{\mathrm{d}\theta_k(t)}{\mathrm{d}t} \tag{5}$$

Figure 5 provides a visual guide to this concept, by illustrating the AM–FM decomposition of IMF5 for the BOU time-series. The top panel (IMF5) of the figure reproduces the mode function, which is also the real part of the analytic signal from equation (2). The amplitude of the latter (AM), given in equation (3), which is the envelope of the original signal, is then extracted and plotted in the ~~middle panel. Unlike in the Fourier decomposition, the~~ second panel. This amplitude is not a constant, but rather a time-dependent function. Next, by removing the AM component from the signal through simple division, the frequency modulation component is obtained, i.e. the complex exponential in equation (2); the real part of this component (FM) is plotted in the ~~bottom~~ third panel. The FM is ~~similar to~~ a trigonometric function ~~, but~~ with a phase argument that ~~unlike the Fourier transform is not a constant but~~ is a time-dependent function, as seen from equation (4). The local frequency ~~(and~~

~~its inverse, the local time-scale of the signal)~~ is then just the first temporal derivative of this phase, as defined in equation (5). The inverse of the local frequency, i.e. the local time-scale of the signal, is depicted in the bottom panel (Time-scale), where its temporal variability can be clearly distinguished. Owing to their time-varying character, the amplitude and frequency are usually encountered in the literature under the terms instantaneous amplitude, and instantaneous frequency, respectively. The original time-series $x(t)$ can then be expressed as a sum of AM–FM signals riding onto the EMD trend, $r(t)$, as follows:

$$x(t) = \mathrm{Re}\left[\sum_{k=1}^{N} a_k(t) \cdot e^{i\int \omega_k(\tau)\,\mathrm{d}\tau}\right] + r(t) \tag{6}$$

The square of the instantaneous amplitude and the instantaneous frequency of the IMFs can then be used to represent the data as an energy density distribution overlaid on the time-frequency space, as in equation (7). This representation, called the Hilbert energy spectrum is defined by Huang et al. (2011) as "the energy density distribution in a time-frequency space divided into equal-sized bins of $\Delta t \times \Delta\omega$ with the value in each bin summed and designated as $a^2(t)$ at the proper time, $t$, and proper instantaneous frequency, $\omega$."

$$S(\omega,t) = \sum_{k=1}^{N} a_k^2(t) \cdot e^{\cancel{i\int \omega_k(\tau)\,\mathrm{d}\tau}\; i\sum \omega_k(t)} \tag{7}$$

The time-integrated version of equation (7), the Hilbert marginal spectrum $S_M(\omega)$, is similar, but not identical to, the traditional Fourier spectrum:

$$S_M(\omega) = \int_0^T S(\omega,t)\,\mathrm{d}t \tag{8}$$

An example of Hilbert spectral representation is given in figure 6 (left panel) where the BOU time-series is shown as an energy density distribution over-imposed on a time-frequency space as in equation (7). Each pixel in the Hilbert spectrum is identified by three attributes – color, abscissa, and ordinate – through which it denotes the local power (color, log-scale) of the corresponding time-series, at a certain time (abscissa), and at a certain time-scale (ordinate, log-scale). For the sake of the readability, the spectrum is binned in time, scale, and colour space and has been smoothed. Hence, some aliasing may occur. Some features may be represented as continuous lines while others are rendered as point-like, especially where rapid frequency modulation takes place, such as in the high-frequency bands.

Interpretation of Hilbert spectral features at data boundaries must be done with care due to possible oscillations of the spline interpolants used in the EMD (see algorithm 1). This effect is similar to the "cone of influence" in the popular wavelet transform (Torrence and Compo, 1998). With the EMD edge effects are usually contained within a half-period of a component at data boundaries (Wu et al., 2011). In figure 6 this region has been whitened out.

The plot in the right panel of figure 6 is the Hilbert marginal spectrum, or the time-integrated variant of of the image at its left, indicating the amount of power at each time-scale. This time agnostic representation is comparable, but not identical to, the

Fourier spectrum of the same time-series. It should be once again emphasized that the Hilbert marginal spectrum is obtained from *local* features of the data, its components having instantaneous amplitude and instantaneous frequency, as opposed to the *global* outlook of the Fourier spectrum whose constituents have constant amplitude and constant frequency throughout the whole domain.

### 3.2 Adaptive background null hypothesis

Which confidence can be attributed to the information extracted by the EMD? More specifically, ~~knowing that measurements may be contaminated by noise,~~ how can one ascertain that a certain IMF is the result of a real physical process as opposed to it possibly being a stochastic manifestation of background ~~noise~~processes?

In the past, several investigations have been carried out in order to identify the effects of the EMD when applied to time-series issued from various ~~noise~~ models, such as white, red, or fractional Gaussian noise, (Huang et al., 2003; Flandrin and Gonçalves, 2004; Flandrin et al., 2004a; Wu and Huang, 2004; Flandrin et al., 2005; Rilling et al., 2005; Schlotthauer et al., 2009; Colominas et al., 2012). As a result, it has been consistently shown that, irrespective of the assumed noise model, the EMD acts as an efficient "wavelet-like" dyadic filter, decomposing the ~~noise input~~ stochastic inputs into IMFs having the same spectral shape, but that are shifted in the frequency domain.

Nevertheless, the rejection of a null hypothesis based on an *a priori* assumed ~~class of noise~~ model of the background does not preclude the probability that the now statistically significant deemed signals originate from a stochastic process of a different kind. Furthermore, as the EMD is an adaptive, data driven decomposition, it would be desirable to also employ a null hypothesis that shares the same characteristics, making no beforehand assumptions about the character of the background ~~noise~~processes.

Following Flandrin (2015) and Chen et al. (2013) this study will make use of the robust statistical properties of the EMD with respect to a wide class of ~~noise~~ background models, in order to adaptively contrast potential signals against ~~the presumably stochastic background~~presumed stochastic realizations, as detailed hereafter. Owing to its dyadic filter character, the EMD decomposes noise ~~inputs~~ time-series into IMFs having similar spectral shape, but that are translated to roughly the next lower octave in the spectral domain. When the sampling step is increased, i.e. the sampling frequency is reduced by fractionally re-sampling the input, these components cannot preserve their original locations in the spectral domain and will instead be shifted towards lower frequencies. Hence, significance testing of IMFs is done by verifying if the IMF remains unchanged in the time-frequency representation of the signal, during fractional re-sampling of the latter.

A Hilbert marginal spectrum $S_{M_k}(\omega)$ is first constructed for each IMF from its instantaneous amplitude $a_k(t)$ and instantaneous frequency $\omega_k(t)$. Next the spectrum-weighted mean frequency (SWMF) $\overline{\omega}_k$ of each IMF is computed (Chen et al., 2013):

$$\overline{\omega}_k = \frac{\displaystyle\int S_{M_k}(\omega)\,\omega\,\mathrm{d}\omega}{\displaystyle\int S_{M_k}(\omega)\,\mathrm{d}\omega} \tag{9}$$

Then, the time-series is fractionally re-sampled by making the original sampling rate $\Delta t$ progressively larger, i.e. the time-spacing of the data points becomes:

$$\Delta t_l = \Delta t \cdot l, \quad l \in \{1.1, 1.2, \ldots, 1.9\} \tag{10}$$

For each sampling rate $l$, and for each IMF $k$, the SWMFs are then recomputed, obtaining a set $\overline{\omega}_{k,l}$. To enhance the visibility of the evolution of frequency as a function of the re-sampling rate, normalization is performed as in:

$$\widehat{\omega}_{k,l} = \frac{\overline{\omega}_{k,l}}{\overline{\omega}_{k,1}} \tag{11}$$

with $\overline{\omega}_{k,1}$ being the SWMFs of the modes of the data having the original sampling rate. Therefore, the normalized SWMFs for the IMFs of the original data will be unity, i.e. $\widehat{\omega}_{k,1} = 1, \forall k$.

Since the EMD is an efficient dyadic filter, frequency deviation from the unity line will occur for ~~noise-like IMFs~~ IMFs generated by stochastic processes. It follows that when $\widehat{\omega}_{k,l} \simeq 1, \forall l$, the null hypothesis that mode $k$ is the realization of stochastic processes can be rejected.

## 4 Results

The IMFs obtained from the BOU time-series from figure 3 have already served as an illustrative example on the operation of the EMD. The IMFs for the other datasets (not shown) are very similar and will be discussed in due time. It must be noted that, like BOU, the CAR time-series is decomposed into 8 IMFs, while the PAY data has 9 IMFs and 10 IMFs are obtained for TAT. Besides the IMFs, for each time-series the decomposition also yields a residual, or trend (also not shown). With respect to the decennial time span of the analysis (10 years), the trend can be thought of as a low-pass approximation of the data (Moghtaderi et al., 2013)~~. Nevertheless, these~~, but not as an oscillation. Since this work focuses mostly on the characteristic scales of temporal variability, the EMD trends along with their statistical significance and physical meaning do not fall within the scope of ~~this~~ the study; for such discussion, see e.g. (Franzke, 2012).

From the Fourier spectra of the IMFs in figure 4 it can be seen that, owing to its median period of 364.8 days, IMF6 can be unambiguously associated with the yearly cycle, as dictated by the orbital parameters of the Earth-Sun system. IMF6 also accounts for the most prominent visual feature in the original data (figure 1, top panel: BOU), with its maxima and minima denoting summer and winter, respectively. Further evidence is brought by the spectral shape of IMF6, distinguished by a sharp peak that has the largest power in figure 4. Also noteworthy is that IMF6 seems to modulate the previous five IMFs, as these latter seem to exhibit amplitude excursions that are approximately in phase with the amplitude of IMF6, a phenomenon that is most visually distinguishable in figure 3 for the first three IMFs during the year 2005.

Finally, the last two components, IMF7 and IMF8, having median periods of 783.3 days and 1457.4 days (figure 4), respectively, are seen to exhibit only slight amplitude deviation from zero in their temporal representation. Moreover, these fluctuations in amplitude occur at the end of the signal for IMF7 and at the front edge for IMF8. Interpretation of these components should, thus, be done with care, since edge effects for the EMD are known to be usually contained within a half-period of a component at data boundaries (Wu et al., 2011), i.e. approximately one year for IMF7 and two years for IMF8.

With the FM components obtained, it becomes possible to illustrate the frequency contents of each time-series in terms of its individual IMFs, as shown in figure 7, where by means of box plots the distribution of the local time-scale of each mode is conveyed. This box plot representation is somehow incomplete, as it only accounts for the period distribution of the modes and does not take into account either the amplitude or the temporal localization of the events. For readability, the characteristic period of each IMF with its range of variability is also shown numerically in table 2. The box plots of the instantaneous amplitude of each IMF are given in figure 8.

For all time-series, IMF1…IMF5 have very similar median periods (figure 7), that approximate the dyadic sequence: 3.5 days → 7 days → 14 days → 28 days → 56 days. This dyadic repartition of their median time-scales is worthy of attention since, as it will become apparent in sections 3.2 and 5.1, it plays a major role in discriminating which IMFs can be attributed to deterministic phenomena as opposed to being the output of random realizations of background processes. Moreover, besides the notable similarity among the medians of these modes, for all the datasets both the interquartile ranges and the total ranges of these first five modes exhibit approximately the same variability. Added to this, IMF6 for BOU, CAR, and PAY, and IMF7 for TAT, whose median periods are respectively 368.2, 364.3, 356.6, and 366.6 days, can clearly be associated to the yearly cycle given by the revolution of the Earth around the Sun. This yearly component is very similar for BOU, CAR and to a lesser extent TAT, with an interquartile range that is concentrated around almost the same median value, the only minor difference being the slightly extended range for TAT of ~~200~~ 300 days as opposed to ~~300~~ 200 days for the other two. The PAY yearly mode differs from those of the other stations, its interquartile range and foremost its range being much larger, the latter even overlapping the interquartile ranges of IMF5 and IMF4. This is a result of the mode mixing phenomenon described in section 3.1 that may arise with the EMD, i.e. the coexistence or mixing of different time-scales in the same IMF, mainly related to the intermittence of signal and to contamination with noise (Huang et al., 2003). Nevertheless, the spectral part of IMF6 which overlaps IMF5 and IMF4 has very low power (Bengulescu et al., 2016b), thus this phenomenon does not influence the validity of the analysis. With this in mind, one notes that for BOU and CAR no spectral components are present in the 100 days to 300 days band. Furthermore, TAT is the only dataset that has a transitional mode of 143.2 days median period in between the first five IMFs common to all stations and the yearly cycle.

At this point, the Hilbert frequency distribution of the IMFs for BOU may be compared to the Fourier one from the PSD in figure 4. As previously mentioned, the Hilbert estimates are based on *local* features of the data, and thus are more accurate than the Fourier ones when applied to non-stationary signals. This can be seen especially when comparing the range of the first five high-frequency IMFs, which is upper bounded to about 100 days in figure 7, whereas in the PSD from figure 4 the spectra of the same components are seen to span the whole time-scale range. This also holds for IMF6, which has very narrow Hilbert period range, whose Fourier analogue is the sharp peak in the PSD of the same mode. Similar statements can be made for IMF7 and IMF8. To sum up, it is found that while the Hilbert period distributions of the modes have compact supports, the Fourier representations of the same components span the whole frequency range. Nevertheless, most of the power in the Fourier PSD is assigned to a frequency band that closely corresponds to the Hilbert range. Owing to the global nature of the Fourier transform, however, additional spectral coefficients are needed to provide a complete mathematical description of the data.

Resuming the discussion of the IMF time-scales from figure 7, it can be observed that the low-frequency, i.e. greater than one year, variability of the data, trend notwithstanding, is assigned into slightly overlapping (within the same time-series) IMFs that span the spectrum starting from the one year mark. For BOU and CAR time-series, there are only two modes extending beyond one year. First, IMF7 can be seen to span approximately the same range for both these stations, from about one year to slightly more than three years. For BOU however, the interquartile range and especially the median period is shifted towards higher periods, i.e. 724.7 days vs. 469.5 days for CAR. The last modes IMF8 of these stations are very different, with a very narrow range around the median of 1531.5 days for BOU, and a range of 900 days to over 2000 days and median of 1305.1 days for CAR. For the PAY data, the low-frequency components have narrower spectral support, with two IMFs (IMF7 and 8) that cover the band from 1 to 2.5 years and median periods of 413.6 and 707.5 days, and the IMF9 around 4.5 years ($\backsim$ 1668 days) with a very narrow range. It must also be noted that IMF7 for BOU, CAR and PAY have the same lower end support, and that the couple (IMF7, IMF8) of PAY taken together somehow emulates IMF7 for BOU and CAR. Lastly, TAT is the only dataset whose the low-frequency variability is expressed by three components, IMF8...IMF10, with mean periods of 609 days, 1440.3 days and 2402.6 days. While the first quartile of IMF9 coincides with the upper range of IMF8, the upper range of IMF9 is slightly below the lower range of IMF10, hence the last two modes do not overlap at all. By its range, IMF8 of TAT approximates IMF7 for BOU and CAR, but there is no proximity in terms of median or interquartile range. Similarly, IMF9 of TAT resembles IMF8 of CAR in terms of range, but their medians are not in close agreement and their interquartile ranges even less so.

With the scrutiny of the these low frequency components, the discussion of the time-scale distribution of the IMFs from figure 7 can now be concluded. However, as previously mentioned, this particular illustration, although instructive, is incomplete. First, the box plot representation does not take the instantaneous variations of frequency into account, but renders global aggregates instead – much like the traditional Fourier methods, with the interquartile range spread in addition. This is done on purpose, with the intent of making it easier for the readership not accustomed to the HHT to create analogies with the more familiar methods (e.g. Fourier analysis, wavelets, etc.). Second and last, this particular representation is totally devoid of any information pertaining to the local amplitude, or power, or variance, of the data. With these consideration in mind, the Hilbert spectrum, a true time-frequency representation for non-linear and non-stationary data, will be discussed next. Since the goal of this exercise is to lay the groundwork for the forthcoming discussion, only the spectrum for the BOU data will be provided as an example.

The BOU Hilbert spectrum from figure 6 exhibits a high-frequency feature between 2 days and $\backsim$ 100 days, which corresponds to first five IMFs of the time-series. The instantaneous time-scales of these modes overlap (figure 7), hence the appearance on the Hilbert spectrum of a continuum instead of distinct bands. This spectral feature has relatively low power, that decreases with increasing period, as can be inferred from the sloped dent in the marginal Hilbert spectrum corresponding to this region. In the 2 days to 32 days band, amplitude modulation by the yearly cycle can be inferred from the periodic change in color, with yellow-green tones, occurring mostly during the high irradiance regime of summer, that turn blue during the winterly minima. Next, in the band between 100 and 300 days, a gap in the spectrum is apparent, as can also be inferred from the lack of support in this region for any of the BOU IMFs in figure 7. The yellow trace, corresponding to IMF6, exhibits frequency modulation around the one year period, seen as oscillations in the range of 300 to 450 days, which is also the support of this mode in the box plot of IMF time-scales. The colour of this IMF indicates that it has the highest power of all the components, as can also be inferred from the large peak on the marginal spectrum. The corresponding time-scale fluctuations are centred in 365 days, and frequency modulation is greatest during 2003 through 2005. From 2006 onwards, however, frequency modulation is less pronounced – perhaps capturing the low solar activity around the 2008 minimum in the eleven year cycle solar cycle (Hathaway, 2015). The final two low-frequency, blue-green traces on the spectrum correspond to IMF7 and IMF8. For IMF7, mode mixing is apparent through the occasional sharing of the yearly time-scale band with IMF6, between mid-2003 and 2005. IMF7 has such low power that it fails to leave an imprint on the marginal spectrum and it seems to suddenly spring into existence during summer 2003, which is in perfect agreement with its temporal representation from figure 3 (panel IMF7), whence it can be seen to have negligible amplitude during the first two and a half years. Also in agreement with its temporal depiction from figure 3 (panel IMF8), IMF8 starts out in light-green hues and slowly vanishes during mid-2007. Although this last BOU mode manages to register on the marginal spectrum through two very slight indentations around 1500 days (which is about the median period of this mode from figures 4 and 7), most of its power lies within edge effect territory, hence interpretation of these slight bumps is ambiguous at best.

Thus far, all time-series have been shown to share a high-frequency constituent between 2 days and 100 days composed of five IMFs with mean periods following a dyadic sequence, and an IMF around 365 days that captures the yearly variability. For BOU, CAR and PAY, a low power region can be found in the 100 days to 300 days band. Beyond the one year time-scale, the low-frequency variability in the 1.5 years to 6 years band is captured by another two (BOU and CAR) or three (PAY and TAT) components. The TAT data is the only time-series that has an IMF in the low power band between the high-frequency feature and the yearly cycle (median period 143.2 days).

## 5 Discussion

The previously identified features of the SSI time-series will be now be discussed in terms of their intrinsic temporal scales of variability, and physical statistical meaning.

~~Firstly, the median periods of the IMFs composing the high frequency band are revisited. It has been shown in figure 7 that they follow a dyadic repartition, that approximate the series dyadic sequence: 3.5, 7, 14, 28, 56 days. Such a doubling in frequency in IMFs has been previously reported in astrophysical and geophysical signals. When investigating three independent datasets of satellite observations of the (extraterrestrial) total solar irradiance (TSI), consistently find a similar dyadic scale progression of modes at 13.5, 27, and 54 days, statistically significant within the 95% level, that correspond to the 27-day solar rotation period and its (sub-)harmonics. find intrinsic periodicities having an average of~~ $\cancel{25^{+7}_{-2}}$ ~~and~~ $\cancel{44^{+10}_{-5}}$ ~~days five different solar proxy signals. The mean periods and the associated error bars in sub-/superscript, estimated at the half-level width of the corresponding probability histogram, were obtained by analysing the sunspot area for the whole Sun, and for the northern and southern solar hemispheres taken separately, the 10.7 cm radio flux intensity, and the helioseismic frequency shift. Compelling as it may seem, nevertheless, the imprint of a solar rotation signature on ground measurements of the SSI is highly unlikely, as~~

~~it would imply the existence of hitherto unknown physical mechanisms in Earth's atmosphere . The amplitudes of the IMFs of the TSI time-series and those of the IMFs in the SSI data differ at times by two orders of magnitude, e.g. compare figure 3 with figure 1 in . If the solar rotation signature were to be seen in the IMFs of the SSI this would require the existence of amplifying processes. and have studied the possibility of such a mechanism, and have concluded that, irrespective of the mechanisms invoked and of the amplification of the solar variability, for the past decades solar forcing is only a minor contributor and thus not able to account for most of the global warming observed in the second half of the twentieth century, which could be better explained by an increase in greenhouse gases. Further proof will be provided subsequently, this time from a signal theoretical point of view, in support of the view that it is unlikely that the solar rotation signature is captured in measurements of the SSI.~~

~~Secondly, in the 100 days to 300 days band, two of the stations, BOU and CAR, do not exhibit any variability. For PAY, the support of yearly IMF6 protrudes in this region, although its first quartile rests well below the 200 days mark. As mentioned before, the power of the portion of this IMF that extends into the high-frequency range is very small (not shown). Hence, while not totally devoid of spectral features, this band contains negligible power. A distinct mode is present at TAT in this band, whose median period of 143.2 days somehow seems to continue the dyadic sequence of the previous five modes. These findings are important for the modelling and forecasting of the SSI, as follows. On the one hand, models for BOU and CAR should not containt any power in this band, or at least filter it out. For TAT, on the other hand, any model attempting to reconstruct the SSI should ensure that the 100 days to 300 days region is not a spectral void. In section 5.2, evidence will be presented that the spectral band spanning from two days to 300 days seems to be composed mostly of coloured noises, i.e. random realisations of stochastic background processes, which can be modelled following, e.g. .~~

~~Thirdly, the median periods detected around the one year mark in all the datasets can be explained by the movement of revolution of the Earth around the Sun and the associated orbital parameters. The interpretation of these components is unambiguous, with one notable exception for the PAY time-series, whose IMF6 exhibits mode mixing, i.e. it has a total range that overlaps some of the modes in the high-frequency band. Nevertheless, it will be subsequently be shown that it is indeed these components that account for variability at the one year time-scale.~~

~~Lastly, the components indicative of low-frequency variability on time-scales greater than one year are discussed. The intrinsic time-scales found in these IMFs seem to match once more those pertaining to the so-called solar quasi-biennial oscillations, i.e. variations in the activity of the Sun exhibiting periodicities between 0.6 and 4 years . Again, identify in the modes of the five solar proxies average periods of $395^{+46}_{-46}$, $626^{+69}_{-113}$, and $1423^{+196}_{-146}$ days respectively. also report solar quasi-biennial oscillations (QBOs) with time-scales from 1 year to 4.5 as being fundamental components of the variability of solar magnetic synoptic maps. identifies a 1.68 year peak in the spectral domain by inferring cloud cover from measurements of the SSI, indicating that galactic cosmic rays, rather than solar irradiance, may induce a cloud effect. provide some tentative evidence for such an effect. Nevertheless, the same final precautions must be reiterated, similarly to the previous discussion concerning the high-frequency constituents.~~

~~It has been shown in section 2 that the four measuring stations experience different climates and exhibit differences in terms of $K_T$. Figure 7 shows that the high frequency band composed of the first five IMFs is very much alike for all stations. This~~

~~section investigates the possible relationship between local climate and dissimilarities in terms of the repartition of the IMFs 6 and more.~~

~~It can be noted that the IMF6 for both BOU and CAR has a well-defined period (figure 7), with a median of respectively 368.2 and 364.3 days and very narrow interquartile range. In addition, for both stations, the IMF6 is the mode having the greatest amplitude and by far, compared to the other modes (figure 8). The IMFs 7 and 8 for CAR have less marked periods, i.e. the interquartile ranges are greater than for IMF6, and the amplitude of each IMF is very small. These observations may be related to the high frequency of cloud-free days seen in figure 2 because in absence of clouds, the variability of the daily mean of SSI is predominantly driven by the variability of the solar irradiance received at the top-of-atmosphere during the year.~~

~~PAY and TAT need four IMFs to account for the low frequency variability, i.e. one IMF more than BOU and CAR. IMF6 in PAY has a median period of 356.6 days, close to one year (figure 7) with a large interquartile range. The median amplitude of the IMF6 is approximately half of that of BOU or CAR (figure 8) and the amplitude exhibits large variations. The median amplitude of the IMF7 is similar to that of IMF6 while the period of the IMF7 is well marked with a narrow interquartile range. This may be related to the abundance of the presence of broken clouds that render the SSI signal highly intermittent. This intermittence of the signal could, in turn, explain the mode mixing observed in IMF6 .~~

~~Similar to PAY, TAT also has a low median clearness index~~ $\cancel{\widetilde{K}_T^{\mathrm{TAT}} = 0.51}$~~, which helps explain the presence of a sixth IMF (median period: 143.2 days) between the high-frequency components and the yearly IMF7 (median period: 366.6 days). In other words, there is much more weather noise in the sub-year band at TAT than at PAY, or a lower signal-to-noise ratio of the yearly cycle. Hence, this high level of noise drives the EMD to assign a dedicated intrinsic mode for this region, as opposed to PAY, where the signal in this spectral band is assigned to the yearly IMF through mode mixing.~~

### 5.1 Discriminating deterministic signals from stochastic components in the IMFs

At this point, having identified the spectral characteristics of the SSI time-series by means of the HHT, a question arises with regard to their physical and statistical significance, namely how can one ascertain which features represent the expression of real, deterministic physical phenomena and which ones can be attributed to random realizations of background processes. Such a method~~of discriminating between the "data" and the "noise" IMFs~~, proposed by Chen et al. (2013), has been described in subsection 3.2. The procedure was applied to the first eight IMFs of all the time-series and the results are presented in figure 10 (from top to bottom BOU, CAR, PAY, and TAT). First, each time-series was re-sampled with a fractional sampling rate up to a factor of two, i.e. the original uniform time-spacing of the data, $\Delta t$, was progressively made larger and larger, as described in equation (10): $\Delta t_l = \Delta t \cdot l$, where $l \in \{1.1, 1.2, \ldots, 1.9\}$ is the re-sampling rate and is running along the horizontal axis. Next, the HHT was used to decompose the resulting time-series into IMFs and to compute their spectrum-weighted mean frequencies, following equation (9). In order to emphasize the effects of the fractional re-sampling on the spectral contents of the IMFs, these latter frequencies were then normalized by the SWMF of the original, non re-sampled data as per equation (11). For each dataset, this ratio is indicated on the y-axis, as $\widehat{\omega}_{k,l}$, with $k \in \{1 \ldots 8\}$ indicating the IMF number. It then becomes possible to follow the evolution of the normalized SWMF of each individual IMF as a function of the fractional re-sampling rate (figure 10). As the EMD is an efficient "wavelet-like" dyadic filter, it follows that the IMFs of time-series of pure noise random processes undergo a translation towards lower frequencies under fractional re-sampling. Therefore, for those IMFs whose SWMFs are not down-shifted during re-sampling, the null hypothesis that they are ~~pure noise~~ purely stochastic can be rejected, i.e. they represent meaningful signals. Stated otherwise, an IMF $k$ is deemed not to be stochastic in nature if it normalized SMWFs $\widehat{\omega}_{k,l}$ stay close to the unity line for all $l$. From figure 10 it can be observed that for all the stations, the only component that maintains quasi-constant frequency under fractional re-sampling is the mode representative of the yearly variability, i.e. IMF6 for BOU, CAR and PAY, and IMF7 for TAT. All the other IMFs experience the previously described frequency down-shifting, hence for them the null hypothesis that they are ~~composed of pure noise~~ purely stochastic in nature cannot be rejected. Since the normalized SWMFs of the yearly components clearly stray from the black dashed line in figure 10 the result that they are not stochastic in nature is unambiguous. This also indicates that the signal-to-noise ratio of these components is well above the minimum value of 0.2–0.3 required to reveal potential signals (Chen et al., 2013).

At this point, several precautionary notes are compulsory. First, the rule of inference used here is *modus tollens*, i.e. the results from figure 10 do not imply that the modes who experience down-shift in their SWMFs are made up of pure noise. It will be subsequently shown that, for the first five IMFs at least, this is indeed the case; although ~~"noise-like", or~~ (quasi-)stochastic in nature, they are not completely devoid of information. Second, the result is mostly qualitative, since it is difficult to define a confidence interval owing to the adaptive nature of the null hypothesis that can account for different types of ~~noise~~models of the stochastic background. Third and last, the approach is best applied only to the high frequency modes, with respect to the data length and sampling, since by re-sampling spurious low-frequency oscillation may inadvertently be introduced (Chen et al., 2013). This is further supported by the fact that as IMF number progresses, the region where the influence of edge effects becomes important is getting larger and larger, hence only adding uncertainty to the interpretation of the results. This is also the reason why this type of analysis was only carried out on the first eight IMFs of each dataset. As a corollary, unambiguous interpretations of QBO-like components seems to be out of reach.

### 5.2 Amplitude modulation through non-linear cross-scale coupling

This section investigates whether the first five IMFs can be modelled as purely uncorrelated, random noise, or if they also contain any other form of information. To test this, the rank correlation between the yearly and sub-yearly IMFs and their envelopes, e.g the AM part in the middle panel of figure 5, has been computed for each SSI time-series. Kendall's rank correlation coefficient, $\tau$, a statistical measure of ordinal association describing how similar the orderings of the data are when ranked (Kendall, 1938), is employed here to establish whether each pair of the two variables, AM$x$ and IMF$y$ with $x, y \in \{1 \dots 7\}$, may be regarded or not as independent. $\tau = 1$ indicates perfect agreement between rankings, while $\tau = -1$ denotes perfect disagreement, i.e. one ranking is the reverse of the other; for $\tau = 0$ the two variables are statistically independent.

The resulting rank correlation coefficients and the associated $p$-values, are presented in figure 9. For each panel, the columns denote the EMD modes (IMF$x$), and the rows their amplitude envelopes (AM$y$). The background colour of each cell (AM$x$, IMF$y$) indicates the rank correlation $\tau$ between IMF$y$ and the AM part of IMF$x$ within the same dataset. The legend of the colour encoding is found on the color bar at the bottom of the figure. The associated $p$-values are presented numerically in each cell, for the sake of completeness and transparency (Wasserstein and Lazar, 2016). For BOU, CAR and

PAY, IMF6 accounts for the yearly variability of the time-series, hence the correlation matrices are $6 \times 6$ in size. For TAT, the yearly mode is IMF7, thus in this case the correlation matrix has a size of $7 \times 7$. Two conclusions can be drawn from figure 9.

Values of $\tau$ significantly different from zero, shown in red, are recorded in the last column, for all stations. These demonstrate a modulation of the amplitude of the components having sub-year time-scales, i.e. AM1…AM5, respectively AM6 for TAT, by the yearly IMF, at a statistically significant level ($p \sim 0$). The effect is most pronounced for PAY, as inferred from the darker red shades (larger rank correlation coefficients).

For the BOU and CAR datasets the first row (AM1) exhibits blue and dark blue cells for IMF3…IMF5 at the statistically significant level. This indicates a negative rank correlation. Similar, but lighter, amplitude modulation is observed on the second row (AM2), but only by IMF4 and IMF5. For the PAY series, this negative rank correlation is greatly reduced for the first row (light blue tones) and is absent in the second row. For TAT no such correlation can be observed. At this point it is interesting to note, that in a similar way to the discussion from section 5.4, the different features of the datasets from figure 9 also enable a classification of the local climate experienced by the measuring stations.

It should be mentioned that the amplitude modulation of high-frequency ~~"noise-like"~~ components by lower frequency ones is also found in sunspots number time series (Chen et al., 2013) and in multiple solar proxies (Kolotkov et al., 2015). The short term intrinsic periodicities in the solar proxies appear to be indicative of "randomly distributed dynamical processes in the solar atmosphere" that are closely related to the 11 year solar activity and therefore, unsurprisingly, the high-frequency modes are found to be modulated by this latter cycle (Kolotkov et al., 2016). But this phenomenon is not limited to solar activity signals, and has also been identified in surface air temperature records (Paluš, 2014), and time-series of the sea level (Liu et al., 2007), and may indicate cross-scale non-linear couplings (Paluš, 2014; Huang et al., 2016). ~~Following this train of thought, the term "weather noise" is adopted for~~

### 5.3 The intrinsic time-scales of variability of the SSI

Firstly, the median periods of the IMFs composing the high frequency band are revisited. It has been shown in figure 7 that they follow a dyadic repartition, that approximate the series dyadic sequence: 3.5, 7, 14, 28, 56 days. Such a doubling in frequency in IMFs has been previously reported in astrophysical and geophysical signals. When investigating three independent datasets of satellite observations of the (extraterrestrial) total solar irradiance (TSI), Lee et al. (2015) consistently find a similar dyadic scale progression of modes at 13.5, 27, and 54 days, statistically significant within the 95% level, that correspond to the 27-day solar rotation period and its (sub-)harmonics. Kolotkov et al. (2015) find intrinsic periodicities having an average of $25^{+7}_{-2}$ and $44^{+10}_{-5}$ days five different solar proxy signals. The mean periods and the associated error bars in sub-/superscript, estimated at the half-level width of the corresponding probability histogram, were obtained by analysing the sunspot area for the whole Sun, and for the northern and southern solar hemispheres taken separately, the 10.7 cm radio flux intensity, and the helioseismic frequency shift. Emery et al. (2011) also find periodicities of 5, 7, 9, 13.5, and 27 days in different radiation belt, solar wind, geomagnetic and auroral parameters. Compelling as it may seem, nevertheless, the imprint of a solar rotation signature on ground measurements of the SSI is highly unlikely, as it would imply the existence of hitherto unknown physical mechanisms in Earth's atmosphere (Thuiller, 2015). The amplitudes of the IMFs of the TSI time-series and those of the IMFs in the SSI data differ at times by two orders of magnitude, e.g. compare figure 3 with figure 1 in (Lee et al., 2015). If the solar rotation signature were to be seen in the IMFs of the SSI this would require the existence of amplifying processes. Stott et al. (2003) and Lockwood and Fröhlich (2007) have studied the possibility of such a mechanism, and have concluded that, irrespective of the mechanisms invoked and of the amplification of the solar variability, for the past decades solar forcing is only a minor contributor and thus not able to account for most of the global warming observed in the second half of the twentieth century, which could be better explained by an increase in greenhouse gases. Further proof will be provided subsequently, this time from a signal theoretical point of view, in support of the view that it is unlikely that the solar rotation signature is captured in measurements of the SSI.

Secondly, in the 100 days to 300 days band, two of the stations, BOU and CAR, do not exhibit any variability. For PAY, the support of yearly IMF6 protrudes in this region, although its first quartile rests well below the 200 days mark. As mentioned before, the power of the portion of this IMF that extends into the high-frequency ~~band defined by the~~ range is very small (not shown). Hence, while not totally devoid of spectral features, this band contains negligible power. A distinct mode is present at TAT in this band, whose median period of 143.2 days somehow seems to continue the dyadic sequence of the previous five modes. Since a similar transitional mode has also been found for two locations in Europe (Bengulescu et al., 2017), presently no explanation in terms of physical processes, such as monsoon rainy seasonality, can be proposed for IMF6 of TAT. These findings are important for the modelling and forecasting of the SSI, as follows. On the one hand, models for BOU and CAR should not contain any power in this band, or at least filter it out. For TAT, on the other hand, any model attempting to reconstruct the SSI should ensure that the 100 days to 300 days region is not a spectral void. In section 5.2, evidence will be presented that the spectral band spanning from two days to 300 days seems to be composed mostly of random realisations of stochastic background processes, which can be modelled following, e.g. (Flandrin et al., 2004a; Rilling et al., 2005; Welter and Esquef, 2013; Kolotkov et al., 2016).

Thirdly, the median periods detected around the one year mark in all the datasets can be explained by the movement of revolution of the Earth around the Sun and the associated orbital parameters. The interpretation of these components is unambiguous, with one notable exception for the PAY time-series, whose IMF6 exhibits mode mixing, i.e. it has a total range that overlaps some of the modes in the high-frequency band. Nevertheless, it will be subsequently be shown that it is indeed these components that account for variability at the one year time-scale.

Lastly, the components indicative of low-frequency variability on time-scales greater than one year are discussed. The intrinsic time-scales found in these IMFs seem to match once more those pertaining to the so-called quasi-biennial oscillations that have been observed in solar activities and proxies with periodicities between 0.6 and 4 years (Bazilevskaya et al., 2015; Kolotkov et al., 2015; Vecchio et al., 2012), as well in meteorological data like Harrison (2008) who identifies a 1.68 year peak in cloud cover or high-latitude stratospheric temperatures and geopotential heights (Labitzke and Loon, 1988). Nevertheless, within the scope of the current analysis, the interpretation of these low frequency variability components as as a real, possibly QBO-like, signal is uncertain.

### 5.4 The local climate imprint in the IMFs

It has been shown in section 2 that the four measuring stations experience different climates and exhibit differences in terms of $K_T$. Figure 7 shows that the high frequency band composed of the first five IMFs ~~in the Hilbert spectra of the SSI data.~~ is very much alike for all stations. This section investigates the possible relationship between local climate and dissimilarities in terms of the repartition of the IMFs with mode number 6 and higher.

~~Rank correlations between IMFs~~ It can be noted that the IMF6 for both BOU and CAR has a well-defined period (figure 7), with a median of respectively 368.2 and ~~their AM components for BOU~~ 364.3 days and very narrow interquartile range. In addition, for both stations, the IMF6 is the mode having the greatest amplitude and by far, compared to the other modes (figure 8). The IMFs 7 and 8 for CAR have less marked periods, i.e. the interquartile ranges are greater than for IMF6, and the amplitude of each IMF is very small. These observations may be related to the high frequency of cloud-free days seen in figure 2 because in absence of clouds, the variability of the daily mean of SSI is predominantly driven by the variability of the solar irradiance received at the top-of-atmosphere during the year.

PAY and TAT need four IMFs to account for the low frequency variability, i.e. one IMF more than BOU and CAR. IMF6 in PAY has a median period of 356.6 days, close to one year (figure 7) with a large interquartile range. The median amplitude of the IMF6 is approximately half of that of BOU or CAR (figure 8) and the amplitude exhibits large variations. The median amplitude of the IMF7 is similar to that of IMF6 while the period of the IMF7 is well marked with a narrow interquartile range. This may be related to the abundance of the presence of broken clouds that render the SSI signal highly intermittent. This intermittence of the signal could, in turn, explain the mode mixing observed in IMF6 (Huang et al., 2003).

Similar to PAY, TAT also has a low median clearness index $\widetilde{K}_T^{\mathrm{TAT}} = 0.51$, which helps explain the presence of a sixth IMF (median period: 143.2 days) between the high-frequency components and the yearly IMF7 (median period: 366.6 days). In other words, the amplitudes of the stochastic components in the sub-year band are higher at TAT than at PAY, or, conversely, there is a lower signal-to-noise ratio of the yearly cycle. Hence, this high power of the background drives the EMD to assign a dedicated intrinsic mode for this region, as opposed to PAY, where the signal in this spectral band is assigned to the yearly IMF through mode mixing.

## 6 Conclusion and outlook

To sum up, the HHT analysis of decennial time-series of daily means of measurements of the SSI, from distinct BSRN stations has revealed the following: the presence of a high-frequency band (2-100 days) ~~"weather noise"~~ consisting of quasi-stochastic IMFs that have been shown to be amplitude-modulated by the yearly cycle; a low power spectral spectral band in the 100 days to 300 days region; a well-defined spectral peak at the one year mark accounting for the yearly variability; and multiple ~~"QBO"~~ QBO-like components whose character has been, inconclusively, attributed to quasi-stochastic random processes.

This separation of the (quasi-)periodic components of the signal from the apparently random realizations of a ~~noisy~~ stochastic background has been shown to significantly augment accuracy in time-series modelling (Rios and de Mello, 2013). Our findings can be thus directly used to improve models for estimating SSI from satellite images or forecasts of the SSI.

We have shown that the adaptive Hilbert-Huang Transform (HHT) is a versatile tool in analysing SSI data-sets, exhibiting significant non-linearity and non-stationarity. First, we have employed it to extract the intrinsic modes of variability of the SSI at distinct time-scales. Second, the HHT has been used to discriminate between the deterministic yearly cycle and the quasi-stochastic ~~"weather noise"~~high-frequency components. The same methodology could also be employed on different geophysical signals, such as wind speed time-series, river discharge datasets, etc.

When modelling climate processes as dynamical systems with low-frequency oscillations and noise effects, Chekroun et al. (2011) have shown that "even the 'approximately right' noise can help, rather than hinder". Here, we have provided a recipe not only for extracting, but also for characterizing the ~~"weather noise"~~ stochastic high-frequency constituents of long-term time-series of the SSI. Indications with respect to modelling these quasi-stochastic components have also been provided. With respect to SSI forecast models, it is exactly this ~~"weather noise" component~~ spectral region that is the focus of attention (Ehnberg and Bollen, 2005; Hoff and Perez, 2010; Marquez and Coimbra, 2013). Inman et al. (2013) venture as far as stating that "the accuracy of the solar irradiance forecasting models depends almost exclusively on the ability to forecast the stochastic component". In this light, the recipe for ~~noise~~ discriminating the realizations of random background processes that we have put forth can be seen as one of the more significant contributions of our paper.

We have also proposed that a classification of the measuring stations according to climate and/or solar insolation conditions may be possible, based on the Hilbert spectral features of the data. Thus, one future research pathway could consist in creating a catalogue of the variability of the solar resource, at different time-scales, on a global scale via satellite estimates of the SSI. Current meteorological re-analyses are too noisy in their estimates of the SSI to form the basis for such a catalogue (Boilley and Wald, 2015). In terms of solar power production, the low-frequency variability data would aid with policy and investment decisions, while short-term variability would be of interest from a monitoring, operations and engineering perspective.

## 7 Code availability

The software used for this study, comprising general EMD and HSA routines is publicly available online, as follows:

- The fast EMD routine used in this study is provided by Wang et al. (2014) and can be downloaded at:
  http://rcada.ncu.edu.tw/FEEMD.rar

- Methods pertaining to Hilbert spectral analysis are part of a general HHT toolkit provided by Wu and Huang (2009) and can be downloaded at:
  http://rcada.ncu.edu.tw/Matlab%20runcode.zip

- The code for the ICEEMD(AN) algorithm (Colominas et al., 2014) is provided by María Eugenia Torres on her personal webpage, and can be downloaded at:
  http://bioingenieria.edu.ar/grupos/ldnlys/metorres/metorres_files/ceemdan_v2014.m

## 8 Data availability

The raw BSRN datasets employed in this study are made available by König-Langlo et al. (2015). Zip archives containing the data can be found at:
https://doi.pangaea.de/10.1594/PANGAEA.852720.

*Author contributions.* All authors contributed equally to this work.

*Competing interests.* The authors declare no competing interests.

*Disclaimer.* N/A.

*Acknowledgements.* The authors wish to thank Patrick Flandrin from Ecole Normale Supérieure de Lyon, France, and Gerard Thuiller from Laboratoire Atmosphères, Milieux, Observations Spatiales in Guyancourt, France, for the fruitful conversations and their insightful comments that have sparked the development of this study. Dmitrii Kolotkov from the University of Warwick, United Kingdom, is also acknowledged for the personal communications pertaining to the stochastic nature of the ~~"weather noise" band. "~~ high-frequency variability band.

## References


BSRN station listing, http://tinyurl.com/px3dhdz, Accessed Dec. 09, 2015.

Alberti, T., Lepreti, F., Vecchio, A., Bevacqua, E., Capparelli, V., and Carbone, V.: Natural periodicities and Northern Hemisphere–Southern Hemisphere connection of fast temperature changes during the last glacial period: EPICA and NGRIP revisited, Climate of the Past, 10, 1751–1762, doi:10/f6n8p6, 2014.

Alberti, T., Piersanti, M., Vecchio, A., De Michelis, P., Lepreti, F., Carbone, V., and Primavera, L.: Identification of the different magnetic field contributions during a geomagnetic storm in magnetospheric and ground observations, Annales Geophysicae, 34, 1069–1084, doi:10/f9ft2h, 2016.

Bazilevskaya, G., Broomhall, A.-M., Elsworth, Y., and Nakariakov, V.: A combined analysis of the observational aspects of the quasi-biennial oscillation in solar magnetic activity, in: The Solar Activity Cycle, pp. 359–386, Springer, doi:10/bcz3, 2015.

Beer, J., Vonmoos, M., and Muscheler, R.: Solar variability over the past several millennia, Space Science Reviews, 125, 67–79, doi:10/djw2kj, 2006.

Bengulescu, M., Blanc, P., Boilley, A., and Wald, L.: Do modelled or satellite-based estimates of surface solar irradiance accurately describe its temporal variability?, Advances in Science and Research, 14, 35–48, doi:10/b37n, 2017.

Bengulescu, M., Blanc, P., and Wald, L.: On the temporal variability of the surface solar radiation by means of spectral representations, Advances in Science and Research, 13, 121–127, doi:10/bp6g, 2016a.

Bengulescu, M., Blanc, P., and Wald, L.: Characterizing temporal variability in measurements of surface solar radiation and its dependence on climate, Energy Procedia, 97, 164 – 171, doi:10/bt46, European Geosciences Union General Assembly 2016, EGU Division Energy, Resources & the Environment (ERE), 2016b.

Blanc, P., Coulaud, C., and Wald, L.: Yearly changes in surface solar radiation in New Caledonia, Advances in Science and Research, 12, 1–4, doi:10/brvs, 2015.

Boilley, A. and Wald, L.: Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface, Renewable Energy, 75, 135–143, doi:10/f259gv, 2015.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of Essential Climate Variables in support of climate research, applications, and policy, Bulletin of the American Meteorological Society, 95, 1431–1443, doi:10/bcq2, 2014.

Calif, R., Schmitt, F. G., Huang, Y., and Soubdhan, T.: Intermittency study of high frequency global solar radiation sequences under a tropical climate, Solar Energy, 98, 349–365, doi:10/7xg, 2013.

Chekroun, M. D., Kondrashov, D., and Ghil, M.: Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation, Proceedings of the National Academy of Sciences, 108, 11 766–11 771, doi:10/bpt5kk, 2011.

Chen, X., Wang, M., Zhang, Y., Feng, Y., Wu, Z., and Huang, N. E.: Detecting signals from data with noise: theory and applications, Journal of the Atmospheric Sciences, 70, 1489–1504, doi:10/bb6d, 2013.

Cohen, L.: Time-frequency distributions-a review, Proceedings of the IEEE, 77, 941–981, doi:10/d2hp7f, 1989.

Colominas, M. A., Schlotthauer, G., Torres, M. E., and Flandrin, P.: Noise-assisted EMD methods in action, Advances in Adaptive Data Analysis, 4, doi:10/9gp, 2012.

Colominas, M. A., Schlotthauer, G., and Torres, M. E.: Improved complete ensemble EMD: A suitable tool for biomedical signal processing, Biomedical Signal Processing and Control, 14, 19–29, doi:10/7tp, 2014.

Coskun, C., Oktay, Z., and Dincer, I.: Estimation of monthly solar radiation distribution for solar energy system analysis, Energy, 36, 1319–1323, doi:10/cwbf42, 2011.

Duffy, D. G.: The application of Hilbert-Huang transforms to meteorological datasets, Journal of Atmospheric and Oceanic Technology, 21, 599–611, doi:10/b9td58, 2004.

Ehnberg, J. S. and Bollen, M. H.: Simulation of global solar radiation based on cloud observations, Solar Energy, 78, 157–162, doi:10/chvffx, 2005.

Emery, B. A., Richardson, I. G., Evans, D. S., Rich, F. J., and Wilson, G. R.: Solar rotational periodicities and the semiannual variation in the solar wind, radiation belt, and aurora, Solar Physics, 274, 399–425, doi:10/fj4ksd, 2011.

Flandrin, P.: personal communication, 2015.

Flandrin, P. and Gonçalves, P.: Empirical mode decompositions as data-driven wavelet-like expansions, International Journal of Wavelets, Multiresolution and Information Processing, 2, 477–496, doi:10/c9cswc, 2004.

Flandrin, P., Gonçalvès, P., and Rilling, G.: EMD equivalent filter banks, from interpretation to applications, in: Hilbert-Huang Transform and Its Applications, pp. 57–74, World Scientific Pub Co Pte Lt, doi:10/dxqw6v, 2005.

Flandrin, P., Rilling, G., and Gonçalves, P.: Empirical mode decomposition as a filter bank, Signal Processing Letters, IEEE, 11, 112–114, doi:10/c9g7q3, 2004a.

Flandrin, P., Gonçalvès, P., and Rilling, G.: Detrending and denoising with empirical mode decompositions, in: 2004 12th European Signal Processing Conference, pp. 1581–1584, 2004b.

Franzke, C.: Multi-scale analysis of teleconnection indices: climate noise and nonlinear trend analysis, Nonlinear Processes in Geophysics, 16, 65–76, doi:10/fsgrzt, 2009.

Franzke, C.: Nonlinear trends, long-range dependence, and climate noise properties of surface temperature, Journal of Climate, 25, 4172–4183, doi:10/fx8xk3, 2012.

Gabor, D.: Theory of communication. Part 1: The analysis of information, Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 93, 429–441, doi:10/2rz, 1946.

Harrison, R. G.: Discrimination between cosmic ray and solar irradiance effects on clouds, and evidence for geophysical modulation of cloud thickness, in: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 464-2098, pp. 2575–2590, The Royal Society, doi:10/dqsd4x, 2008.

Hathaway, D. H.: The solar cycle, Living Reviews in Solar Physics, 12, 1–87, doi:10/bjxj, 2015.

Hoff, T. E. and Perez, R.: Quantifying PV power output variability, Solar Energy, 84, 1782–1793, doi:10/dm7spw, 2010.

Huang, N. E. and Shen, S. S. P.: Hilbert–Huang Transform and Its Applications. Second Edition, World Scientific, doi:10/bqvn, 2014.

Huang, N. E. and Wu, Z.: A review on Hilbert-Huang transform: Method and its applications to geophysical studies, Reviews of Geophysics, 46, doi:10/fr2hk2, 2008.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454, 903–995, doi:10/dr5ngd, 1998.

Huang, N. E., Wu, M.-L. C., Long, S. R., Shen, S. S. P., Qu, W., Gloersen, P., and Fan, K. L.: A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 459, 2317–2345, doi:10/c8wsrv, 2003.

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X., and Blank, K.: On instantaneous frequency, Advances in Adaptive Data Analysis, 1, 177–229, doi:10/dw9j3k, 2009.

Huang, N. E., Chen, X., Lo, M.-T., and Wu, Z.: On Hilbert spectral representation: a true time-frequency representation for nonlinear and nonstationary data, Advances in Adaptive Data Analysis, 3, 63–93, doi:10/b7m8xt, 2011.

Huang, N. E., Hu, K., Yang, A. C., Chang, H.-C., Jia, D., Liang, W.-K., Yeh, J. R., Kao, C.-L., Juan, C.-H., Peng, C. K., et al.: On Holo-Hilbert spectral analysis: a full informational spectral representation for nonlinear and non-stationary data, Phil. Trans. R. Soc. A, 374, 20150 206, doi:10/bhhc, 2016.

Inman, R. H., Pedro, H. T., and Coimbra, C. F.: Solar forecasting methods for renewable energy integration, Progress in energy and combustion science, 39, 535–576, doi:10/brhc, 2013.

Kendall, M. G.: A new measure of rank correlation, Biometrika, 30, 81–93, doi:10/fxdpmd, 1938.

Kolotkov, D., Broomhall, A.-M., and Nakariakov, V.: Hilbert–Huang transform analysis of periodicities in the last two solar activity cycles, Monthly Notices of the Royal Astronomical Society, 451, 4360–4367, doi:10/bcxn, 2015.

Kolotkov, D., Anfinogentov, S. A., and Nakariakov, V. M.: Empirical mode decomposition analysis of random processes in the solar atmosphere, Astronomy and Astrophysics, 592, A153, doi:10/br7c, 2016.

König-Langlo, G., Driemel, A., Raffel, B., and Sieger, R.: BSRN snapshot 2015-09, links to zip archives, doi:10/brqd, 2015.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen-Geiger climate classification updated, Meteorologische Zeitschrift, 15, 259–263, doi:10/c37229, 2006.

Labitzke, K. and Loon, H. V.: Associations between the 11-year solar cycle, the QBO and the atmosphere. Part I: the troposphere and stratosphere in the northern hemisphere in winter, Journal of Atmospheric and Terrestrial Physics, 50, 197 – 206, doi:10/fvq5q6, 1988.

Lauret, P., Perez, R., Aguiar, L. M., Tapachès, E., Diagne, H. M., and David, M.: Characterization of the intraday variability regime of solar irradiation of climatically distinct locations, Solar Energy, 125, 99–110, doi:10/bcrf, 2016.

Lee, J. N., Cahalan, R. F., and Wu, D. L.: The 27-day rotational variations in total solar irradiance observations: From SORCE/TIM, ACRIMSAT/ACRIM III, and SOHO/VIRGO, Journal of Atmospheric and Solar-Terrestrial Physics, 132, 64–73, doi:10/6q4, 2015.

Lee, T. and Ouarda, T. B. M. J.: Prediction of climate nonstationary oscillation processes with empirical mode decomposition, Journal of Geophysical Research: Atmospheres, 116, D06 107, doi:10/dm6zvs, 2011.

Liu, Y., San Liang, X., and Weisberg, R. H.: Rectification of the bias in the wavelet power spectrum, Journal of Atmospheric and Oceanic Technology, 24, 2093–2102, doi:10/dv9zq4, 2007.

Lockwood, M. and Fröhlich, C.: Recent oppositely directed trends in solar climate forcings and the global mean surface air temperature, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 463, 2447–2460, doi:10/bkg3th, 2007.

Marquez, R. and Coimbra, C. F.: Proposed metric for evaluation of solar forecasting models, Journal of solar energy engineering, 135, 011 016, doi:10/brhb, 2013.

Medvigy, D. and Beaulieu, C.: Trends in daily solar radiation and precipitation coefficients of variation since 1984, Journal of Climate, 25, 1330–1339, doi:10/dqkcvg, 2012.

Moghtaderi, A., Flandrin, P., and Borgnat, P.: Trend filtering via empirical mode decompositions, Computational Statistics & Data Analysis, 58, 114–126, doi:10/c4nqzz, 2013.

Nagovitsyn, Y. A.: A nonlinear mathematical model for the solar cyclicity and prospects for reconstructing the solar activity in the past, Astronomy Letters, 23, 742–748, http://tinyurl.com/gm6c4u9, 1997.

Ohmura, A., Gilgen, H., Hegner, H., Müller, G., Wild, M., Dutton, E. G., Forgan, B., Fröhlich, C., Philipona, R., Heimo, A., König-Langlo, G., McArthur, B., Pinker, R., Whitlock, C. H., and Dehne, K.: Baseline surface radiation network (BSRN/WCRP): new precision radiometry for climate research, Bulletin of the American Meteorological Society, 79, 2115–2136, doi:10/bvpw97, 1998.

Pachauri, R. K., Allen, M., Barros, V., Broome, J., Cramer, W., Christ, R., Church, J., Clarke, L., Dahe, Q., Dasgupta, P., et al.: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, http://epic.awi.de/37530/, 2014.

Paluš, M.: Multiscale atmospheric dynamics: cross-frequency phase-amplitude coupling in the air temperature, Physical review letters, 112, 078 702, doi:10/bhpj, 2014.

Rilling, G., Flandrin, P., and Gonçalves, P.: On empirical mode decomposition and its algorithms, in: Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, vol. 3, pp. 8–11, Grado (Italy), http://tinyurl.com/k3z2yv3, 2003.

Rilling, G., Flandrin, P., and Gonçalves, P.: Empirical mode decomposition, fractional Gaussian noise and Hurst exponent estimation, in: Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on, vol. 4, pp. iv/489–iv/492 Vol. 4, IEEE, doi:10/fmnjq6, 2005.

Rios, R. A. and de Mello, R. F.: Improving time series modeling by decomposing and analyzing stochastic and deterministic influences, Signal Processing, 93, 3001–3013, doi:10/bdvd, 2013.

Rios, R. A. and de Mello, R. F.: Applying Empirical Mode Decomposition and mutual information to separate stochastic and deterministic influences embedded in signals, Signal Processing, 118, 159–176, doi:10/bdvc, 2016.

Rios, R. A., Parrott, L., Lange, H., and de Mello, R. F.: Estimating determinism rates to detect patterns in geospatial datasets, Remote Sensing of Environment, 156, 11–20, doi:10/bdvf, 2015.

Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., and Zhang, T.: Assessment of BSRN radiation records for the computation of monthly means, Atmospheric Measurement Techniques, 4, 339–354, doi:10/dpqsc8, 2011.

Schlotthauer, G., Torres, M. E., Rufiner, H. L., and Flandrin, P.: EMD of Gaussian white noise: effects of signal length and sifting number on the statistical properties of intrinsic mode functions, Advances in Adaptive Data Analysis, 1, 517–527, doi:10/b9b7st, 2009.

Schroedter-Homscheidt, M., Delamare, C., Heilscher, G., Heinemann, D., Hoyer, C., Meyer, R., Toggweiler, P., Wald, L., and Zelenka, A.: The ESA-ENVISOLAR project: experience on the commercial use of Earth observation based solar surface irradiance measurements for energy business purposes, in: Solar Energy Resources Management for Electricity Generation, edited by Dunlop, E. D., Wald, L., and Šúri, M, pp. 111–124, Nova Science Publishers, http://tinyurl.com/hpf8d5g, 2006.

Solé, J., Turiel, A., and Llebot, J. E.: Using empirical mode decomposition to correlate paleoclimatic time-series, Natural Hazards and Earth System Sciences, 7, 299–307, doi:10/d4n2m3, 2007.

Stott, P. A., Jones, G. S., and Mitchell, J. F.: Do models underestimate the solar contribution to recent climate change?, Journal of Climate, 16, 4079–4093, doi:10/bzdgx4, 2003.

Tary, J. B., Herrera, R. H., Han, J., and van der Baan, M.: Spectral estimation—What is new? What is next?, Reviews of Geophysics, 52, 723–749, doi:10/f259g3, 2014.

Thuiller, G.: personal communication, 2015.

Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, Bulletin of the American Meteorological society, 79, 61–78, doi:10/bhbwhf, 1998.

Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P.: A complete ensemble empirical mode decomposition with adaptive noise, in: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 4144–4147, IEEE, 2011.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's global energy budget, Bulletin of the American Meteorological Society, 90, 311–323, doi:10/d3xccg, 2009.

Vecchio, A., Capparelli, V., and Carbone, V.: The complex dynamics of the seasonal component of USA's surface temperature, Atmospheric Chemistry and Physics, 10, 9657–9665, doi:10/cssb3t, 2010.

Vecchio, A., Laurenza, M., Meduri, D., Carbone, V., and Storini, M.: The dynamics of the solar magnetic field: polarity reversals, butterfly diagram, and quasi-biennial oscillations, The Astrophysical Journal, 749, 27, doi:10/bcz5, 2012.

Wahab, M. A., El-Metwally, M., Hassan, R., Lefevre, M., Oumbe, A., and Wald, L.: Assessing surface solar irradiance and its long-term variations in the northern Africa desert climate using Meteosat images, International Journal of Remote Sensing, 31, 261–280, doi:10/ck533g, 2010.

Wang, G., Chen, X.-Y., Qiao, F.-L., Wu, Z., and Huang, N. E.: On intrinsic mode function, Advances in Adaptive Data Analysis, 2, 277–293, doi:10/b466fq, 2010.

Wang, Y.-H., Yeh, C.-H., Young, H.-W. V., Hu, K., and Lo, M.-T.: On the computational complexity of the empirical mode decomposition algorithm, Physica A: Statistical Mechanics and its Applications, 400, 159–167, doi:10/3nz, 2014.

Wasserstein, R. L. and Lazar, N. A.: The ASA's statement on *p*-values: context, process, and purpose, The American Statistician, 70, 129–133, doi:10/bc4d, 2016.

Welter, G. S. and Esquef, P. A. A.: Multifractal analysis based on amplitude extrema of intrinsic mode functions, Physical Review E, 87, 032 916, doi:10/brmr, 2013.

Wu, Z. and Huang, N. E.: A study of the characteristics of white noise using the empirical mode decomposition method, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 460, 1597–1611, doi:10/d9p67j, 2004.

Wu, Z. and Huang, N. E.: Ensemble empirical mode decomposition: a noise-assisted data analysis method, Advances in Adaptive Data Analysis, 01, 1–41, doi:10/fdnzfq, 2009.

Wu, Z. and Huang, N. E.: On the filtering properties of the empirical mode decomposition, Advances in Adaptive Data Analysis, 2, 397–414, doi:10/bscvxv, 2010.

Wu, Z., Huang, N. E., and Chen, X.: Some considerations on physical analysis of data, Advances in Adaptive Data Analysis, 3, 95–113, doi:10/cftghv, 2011.

Yordanov, G. H., Saetre, T. O., and Midtgard, O.-M.: 100-millisecond resolution for accurate overirradiance measurements, Photovoltaics, IEEE Journal of, 3, 1354–1360, doi:10/7v8, 2013.

Zeng, Z., Yang, H., Zhao, R., and Meng, J.: Nonlinear characteristics of observed solar radiation data, Solar Energy, 87, 204–218, doi:10/bcr8, 2013.

## List of Figures



## List of Tables

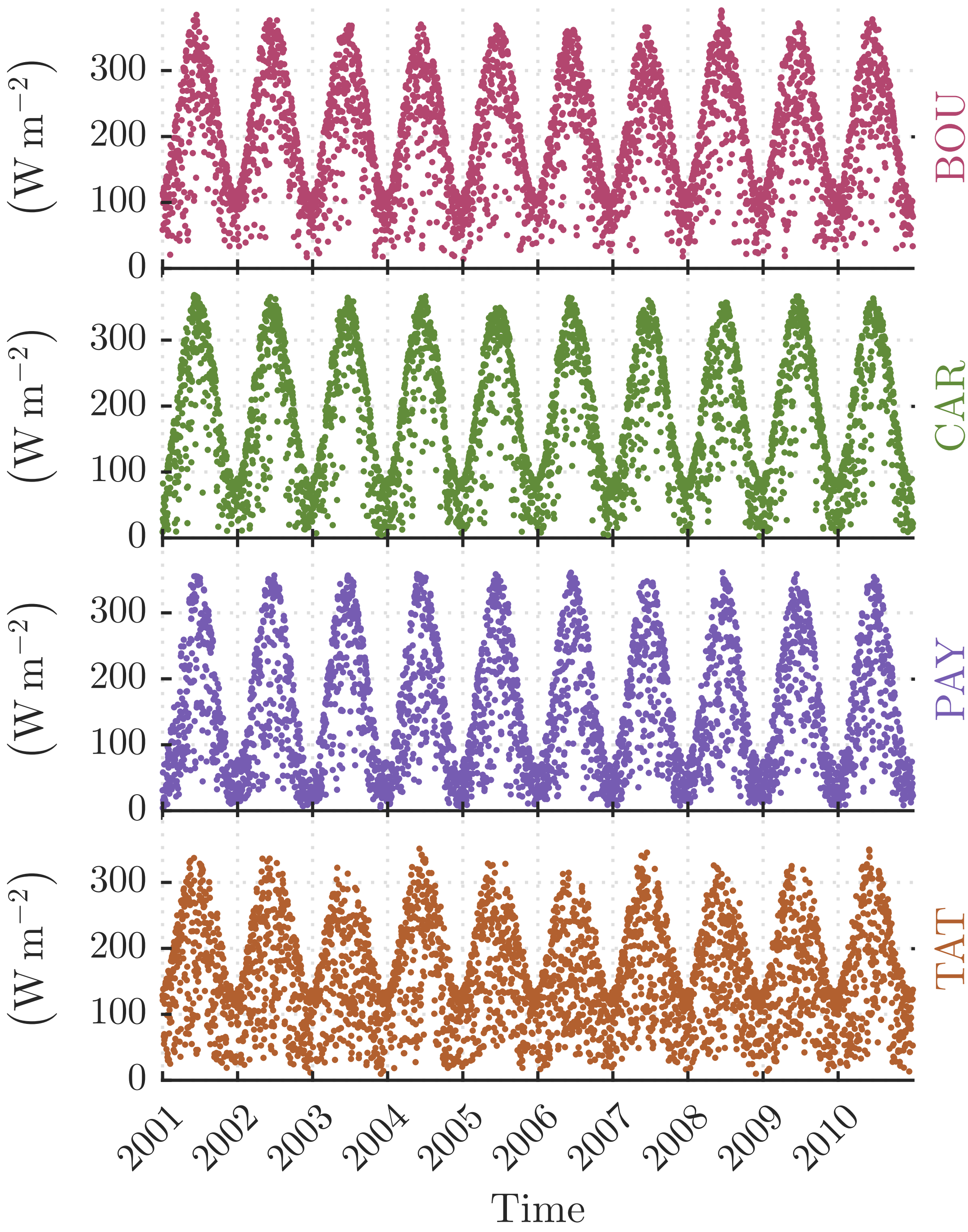


**Figure 1.** The four decennial SSI time-series investigated in this study, spanning 2001 through 2010. From top to bottom: BOU, CAR, PAY, and TAT. Each point corresponds to a daily mean of SSI. Time markers on the abscissa indicate the start of the corresponding year.

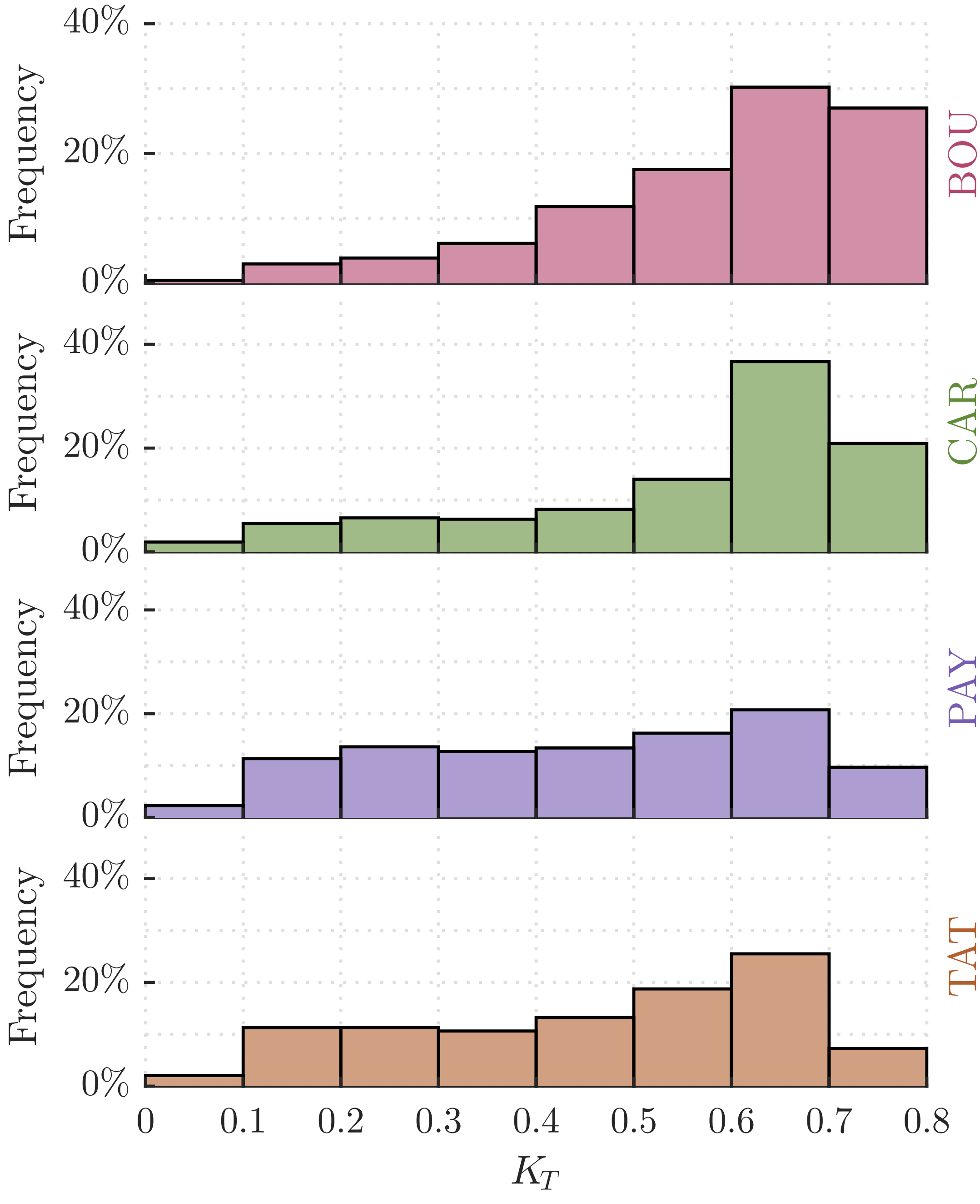


**Figure 2.** Histograms of the daily clearness index $K_T$ over the decennial time-span in frequency in percent. From top to bottom: BOU, CAR, PAY, and TAT.

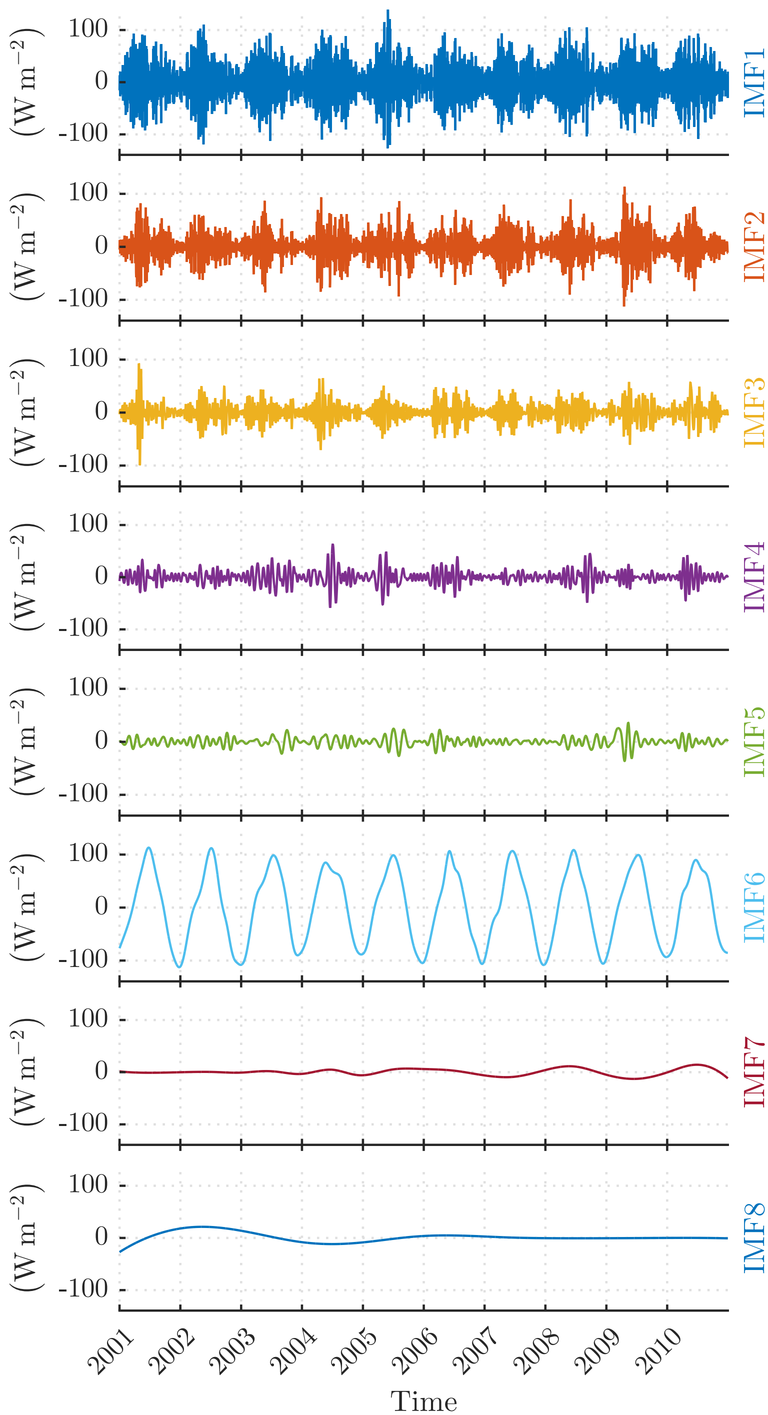


**Figure 3.** The eight IMFs obtained by decomposing the BOU time-series, from top to bottom IMF1...IMF8. The panels plot SSI (ordinate) versus time (abscissa). Time markers on the horizontal axes indicate January 1$^{st}$ of the corresponding year. The zero-centred oscillatory nature of the modes can be clearly seen. Also apparent is the local time-scale increase with mode number.

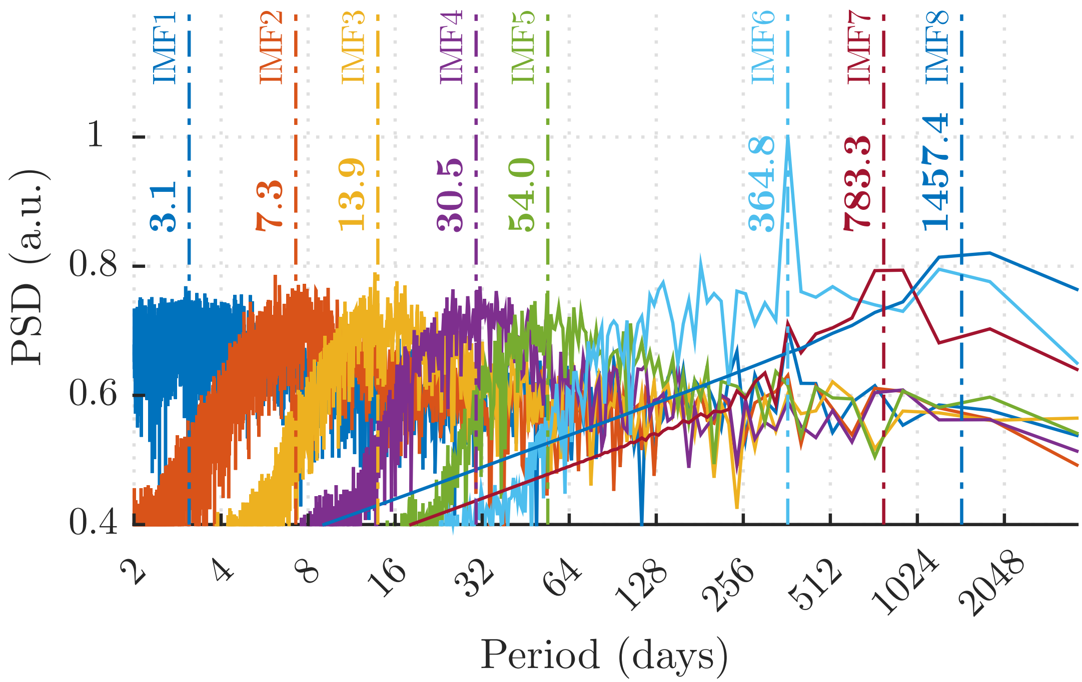


**Figure 4.** The power spectral density (PSD) of the eight IMFs for BOU (solid line) on a logarithmic scale normalized with respect to the power of the highest spectral peak. The period, or inverse frequency, runs on the abscissa in a base-2 logarithm. The individual spectra are shown in the same colours as the IMFs from figure 3, from left to right: IMF1. . . IMF8. The Fourier estimates of the median periods, marked along the dash-dotted lines, are seen to increase with mode number. Notable features are the prominent spectral peak of IMF6 at $\sim 365$ days corresponding to the yearly cycle and the apparent dyadic repartition of the time-scales for IMF1. . . IMF5.

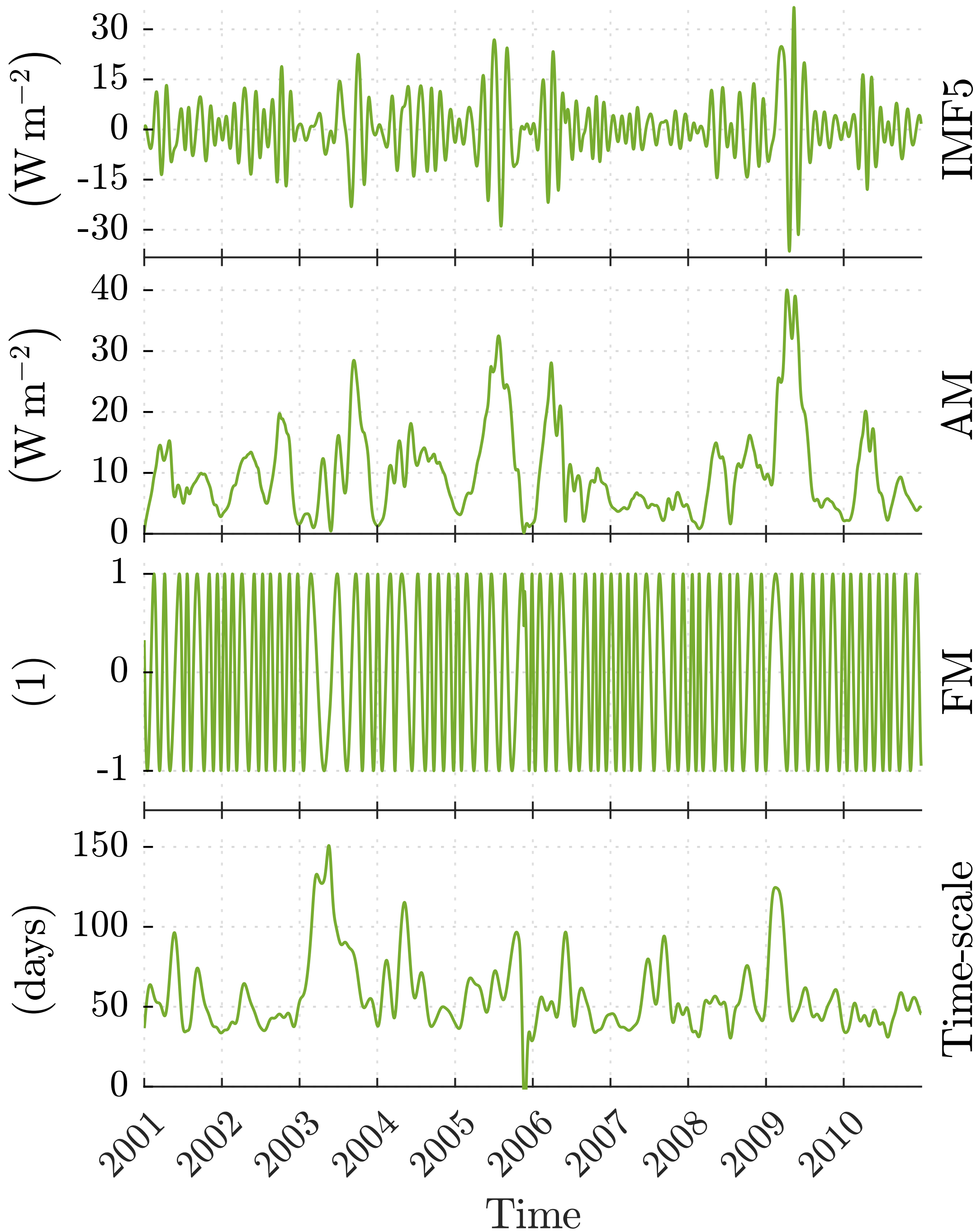


**Figure 5.** Hilbert spectral analysis of the fifth IMF of the BOU time-series. The intrinsic mode function (IMF5 ~~, top~~ panel) is the product of its constituent slowly-varying amplitude modulation part (AM ~~, middle~~ panel) and of its rapidly-changing frequency-modulation component (FM panel). The time-varying local time-scale, ~~bottom~~ extracted from the FM component, is also depicted (Time-scale panel). Time markers on the abscissa denote the beginning of the corresponding year.

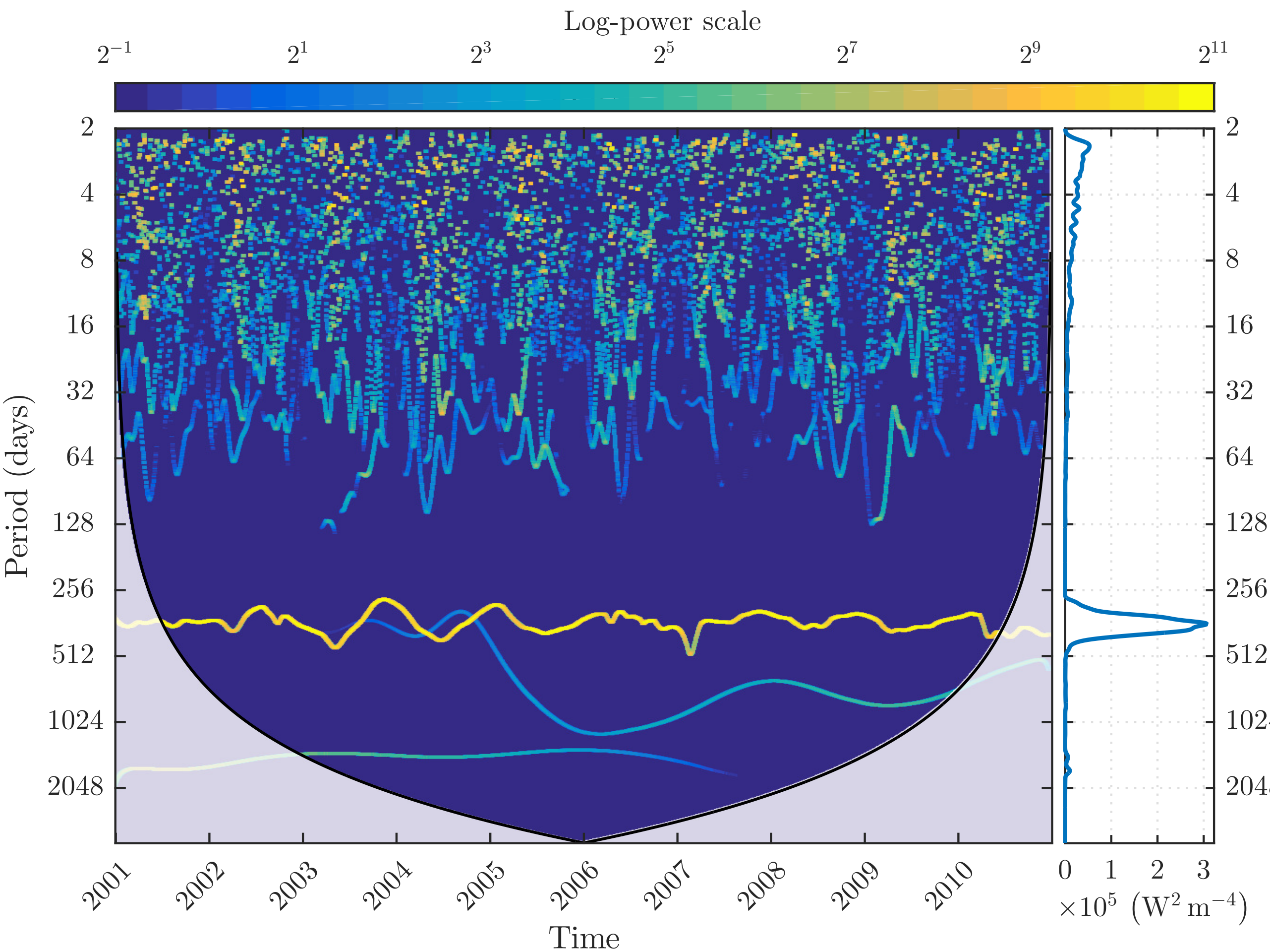


**Figure 6.** The Hilbert spectrum (left panel) of the ten year time-series of SSI for BOU, spanning 2001 through 2010. Pixel colour encodes power (logarithmic scale colour bar on top) at each instant (abscissa) and each scale (ordinate). Time markers on the horizontal axis denote the start of the corresponding year. The white-out area indicates the regions where edge effects become significant. The Hilbert marginal spectrum in the right panel is the time-integrated version, i.e. line-by-line sum, of the Hilbert spectrum and indicates the amount of power at each scale.

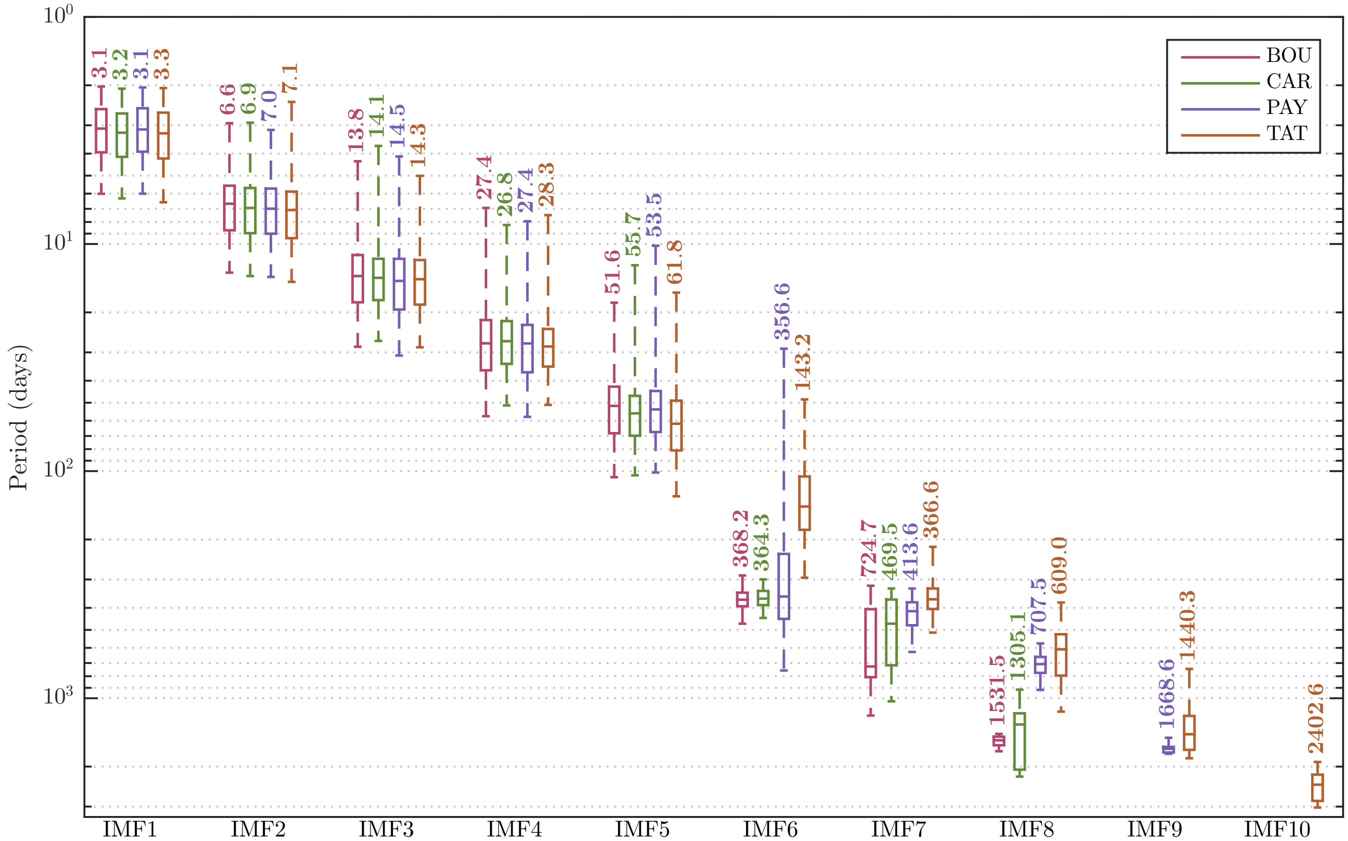


**Figure 7.** Box plot of the instantaneous time-scales of the IMFs for the four stations. The top and the bottom edges of the boxes represent the first ($Q1$) and, respectively, the third ($Q3$) quartiles. The bars inside boxes denote the second quartile ($Q2$), i.e. the median. The whisker length is set at at most 1.5 times the interquartile range, i.e. $1.5 \times (Q3 - Q1)$, hence the whiskers roughly correspond to $\pm 2.7$ standard deviations, or equivalently $\backsim 99\%$ of the data, assuming normal distribution. The median for each box is expressed numerically above the lower whiskers. Outliers are omitted. Numeric values for all the statistical descriptors are shown in table 2.

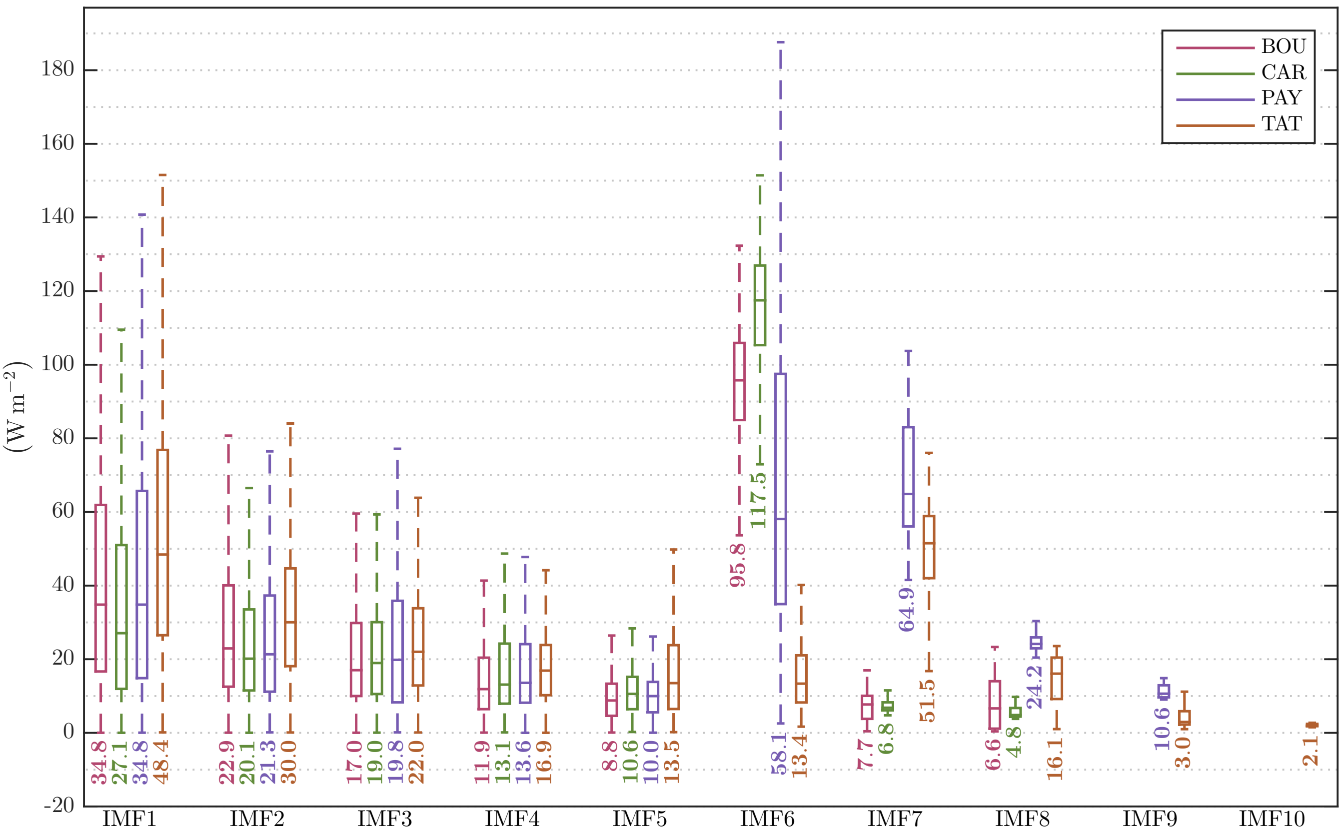


**Figure 8.** Box plot of the instantaneous amplitudes of the IMFs for the four stations. The bottom and the top edges of the boxes represent the first ($Q1$) and, respectively, the third ($Q3$) quartiles. The bars inside boxes denote the second quartile ($Q2$), i.e. the median. The whisker length is set at at most 1.5 times the interquartile range, i.e. $1.5 \times (Q3 - Q1)$, hence the whiskers roughly correspond to $\pm 2.7$ standard deviations, or equivalently $\backsim 99\%$ of the data, assuming normal distribution. The median for each box is expressed numerically below the lower whiskers. Outliers are omitted.

~~The drift of normalized SWMF $\widehat{\omega}_{k,l}$ (ordinate) of IMF $k$, with $k \in \{1 \ldots 8\}$, as a function of the re-sampling rate $l$ (abscissa) for the four time-series. From top to bottom: BOU, CAR, PAY, and TAT. The black dashed diagonal depicts the behaviour of a pure noise time-series under an ideal dyadic filter. For all datasets, the only mode that maintains a quasi-constant frequency under fractional re-sampling is the IMF associated with the yearly cycle, i.e. IMF6 for BOU,CAR, and PAY and IMF7 for TAT. In all the other IMFs quasi-stochastic behaviour is apparent, through frequency down-shifting towards the next lower octave, approximately following the dashed line.~~

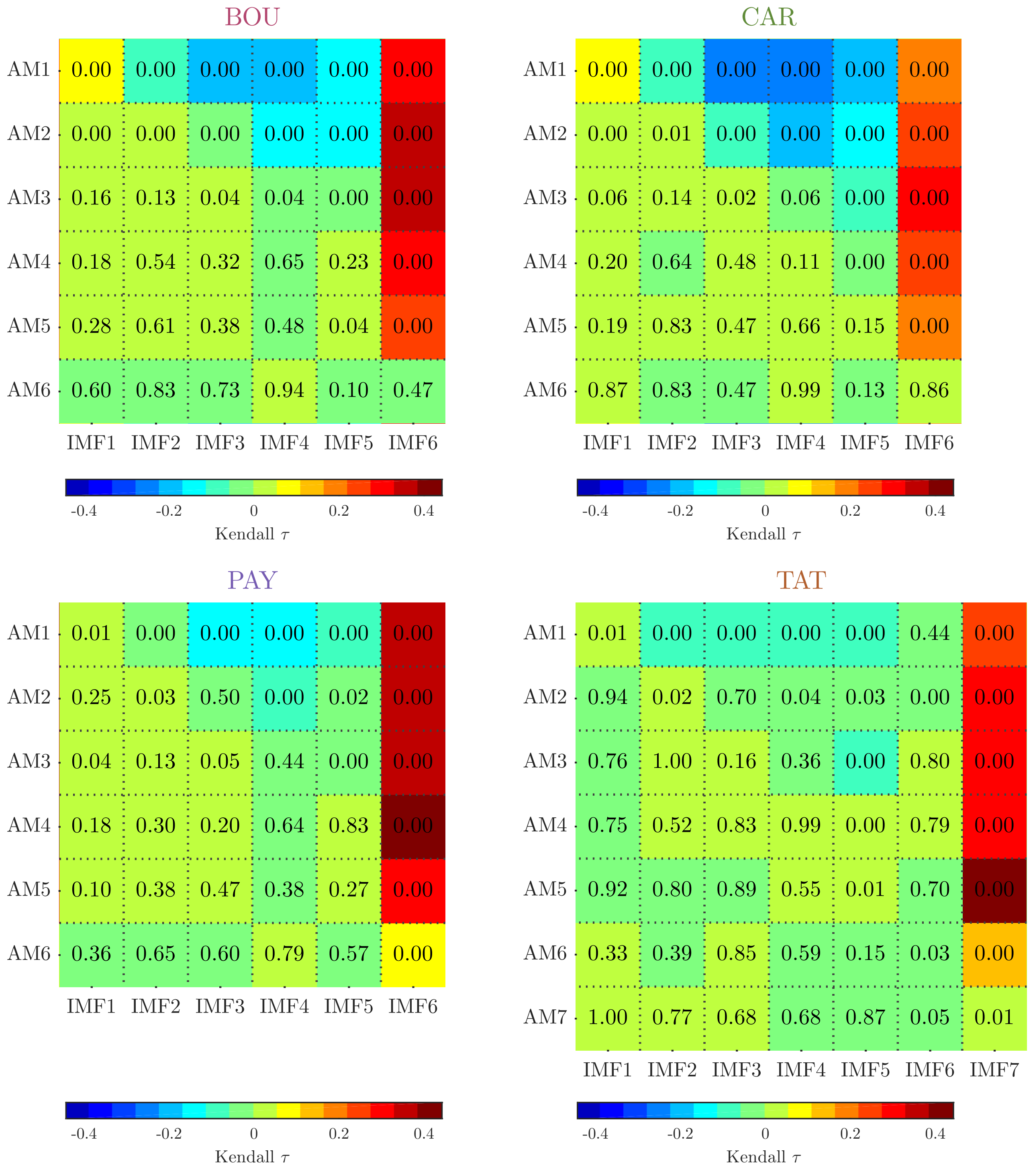


**Figure 9.** Rank correlations between IMFs and their AM components for BOU (top left), CAR (top right), PAY (bottom left) and TAT (bottom right). Kendall's rank correlation coefficient $\tau$ is colour-coded according to the colour bar on the bottom. IMFs run vertically, along the columns, and their AM components run horizontally, along the rows. The numeric values within the cells are the associated $p$-values.

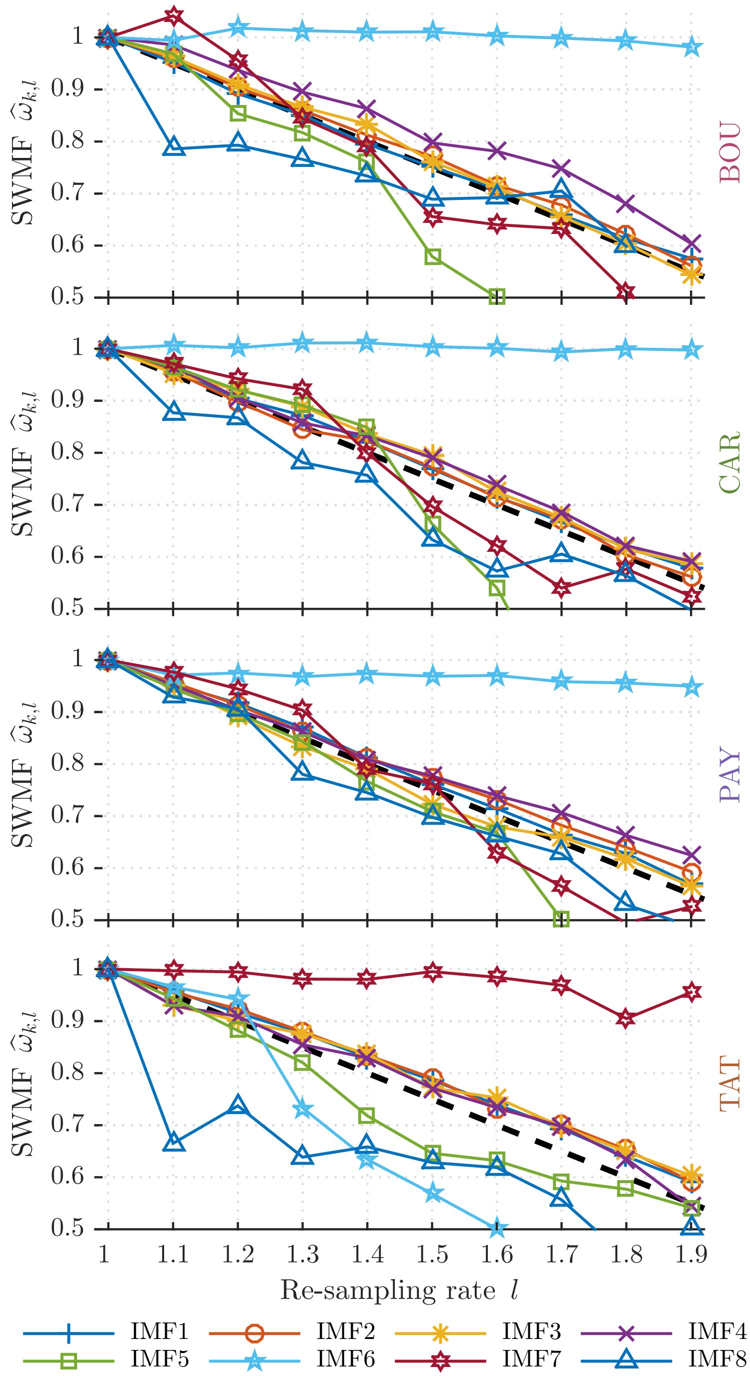


**Figure 10.** The drift of normalized SWMF $\widehat{\omega}_{k,l}$ (~~top left~~ordinate) of IMF $k$, ~~CAR~~ with $k \in \{1 \dots 8\}$, as a function of the re-sampling rate $l$ (~~top right~~abscissa) for the four time-series. From top to bottom: BOU, CAR, PAY~~(bottom left)~~, and TAT~~(bottom right)~~. ~~Kendall's rank correlation coefficient $\tau$~~ The black dashed diagonal depicts the behaviour of a pure noise time-series under an ideal dyadic filter. For all datasets, the only mode that maintains a quasi-constant frequency under fractional re-sampling is ~~colour-coded according to~~ the ~~colour bar on~~ IMF associated with the ~~bottom~~yearly cycle, i.~~IMFs run verticall~~ye. IMF6 for BOU,~~along the columns~~CAR, and ~~their AM components run horizontally, along the rows~~PAY and IMF7 for TAT. ~~The numeric values within~~ In all the ~~cells are~~ other IMFs quasi-stochastic behaviour is apparent, through frequency down-shifting towards the ~~associated $p$-values~~next lower octave, approximately following the dashed line.

**Table 1.** Ground measurement stations listing.

| Code | Location* | | Latitude† | Longitude† | Climate‡ |
|---|---|---|---|---|---|
| BOU | Boulder | (US) | 40.0500 | -105.0070 | BSk |
| CAR | Carpentras | (FR) | 44.0830 | 5.0590 | Csa |
| PAY | Payerne | (CH) | 46.8150 | 6.9440 | Cfb |
| TAT | Tateno | (JP) | 36.0581 | 140.1258 | Cfa |

* Country codes according to ISO 3166-1 alpha-2

† Positive north for latitude and positive east for longitude, following ISO 19115

‡ Köppen-Geiger climate classification according to Kottek et al. (2006)

**Table 2.** Statistical descriptors of the instantaneous time-scales of the IMFs, expressed in days.

| Station | Descriptor* | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 | IMF7 | IMF8 | IMF9 | IMF10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BOU | | | | | | | | | | | |
| | Lower whisker | 2.0 | 2.9 | 4.3 | 6.9 | 18.1 | 288.1 | 319.6 | 1435.5 | | |
| | First quartile | 2.5 | 5.5 | 11.2 | 21.6 | 42.5 | 342.9 | 405.4 | 1476.0 | | |
| | Second quartile | 3.1 | 6.6 | 13.8 | 27.4 | 51.6 | 368.2 | 724.7 | 1531.5 | | |
| | Third quartile | 3.9 | 8.7 | 18.1 | 36.0 | 68.1 | 393.7 | 807.6 | 1611.1 | | |
| | Upper whisker | 6.0 | 13.4 | 28.3 | 57.3 | 106.3 | 469.5 | 1192.1 | 1710.9 | | |
| CAR | | | | | | | | | | | |
| | Lower whisker | 2.1 | 2.9 | 3.7 | 8.2 | 12.4 | 299.7 | 328.6 | 916.4 | | |
| | First quartile | 2.7 | 5.7 | 11.6 | 21.8 | 46.6 | 337.0 | 367.6 | 1165.4 | | |
| | Second quartile | 3.2 | 6.9 | 14.1 | 26.8 | 55.7 | 364.3 | 469.5 | 1305.1 | | |
| | Third quartile | 4.1 | 9.0 | 17.7 | 33.7 | 69.8 | 388.9 | 716.2 | 2062.3 | | |
| | Upper whisker | 6.3 | 13.8 | 26.7 | 51.4 | 104.4 | 443.0 | 1031.8 | 2212.8 | | |
| PAY | | | | | | | | | | | |
| | Lower whisker | 2.0 | 3.1 | 4.1 | 7.9 | 10.2 | 28.9 | 328.8 | 573.7 | 1493.0 | |
| | First quartile | 2.5 | 5.7 | 11.6 | 22.7 | 44.4 | 231.5 | 378.1 | 658.0 | 1637.0 | |
| | Second quartile | 3.1 | 7.0 | 14.5 | 27.4 | 53.5 | 356.6 | 413.6 | 707.5 | 1668.6 | |
| | Third quartile | 3.9 | 9.0 | 19.4 | 36.7 | 67.3 | 447.4 | 477.2 | 772.9 | 1733.5 | |
| | Upper whisker | 6.0 | 14.0 | 31.0 | 57.7 | 101.5 | 755.0 | 625.3 | 918.9 | 1757.2 | |
| TAT | | | | | | | | | | | |
| | Lower whisker | 2.1 | 2.4 | 5.0 | 7.5 | 16.4 | 48.3 | 215.4 | 378.8 | 742.8 | 1908.4 |
| | First quartile | 2.6 | 5.9 | 11.8 | 23.7 | 48.9 | 105.6 | 328.7 | 522.8 | 1196.3 | 2169.2 |
| | Second quartile | 3.3 | 7.1 | 14.3 | 28.3 | 61.8 | 143.2 | 366.6 | 609.0 | 1440.3 | 2402.6 |
| | Third quartile | 4.2 | 9.4 | 18.5 | 34.6 | 81.1 | 181.3 | 404.7 | 795.0 | 1687.8 | 2831.9 |
| | Upper whisker | 6.5 | 14.7 | 28.5 | 51.1 | 129.0 | 294.6 | 513.9 | 1145.5 | 1837.4 | 3029.0 |

* A boxplot illustration of the statistical descriptors is shown in figure 7.