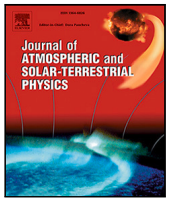
Journal of Atmospheric and Solar–Terrestrial Physics 246 (2023) 106058

Contents lists available at ScienceDirect

Journal of Atmospheric and Solar–Terrestrial Physics

journal homepage: www.elsevier.com/locate/jastp



Research paper

Quantifying Schumann resonances' variation over time through statistical differences

Manuel Soler-Ortiz *, Manuel Fernández-Ros, Nuria Novas-Castellano, Jose A. Gázquez-Parra

Engineering Dept., Carr Sacramento S/N, La Canada, 04120, Almeria, Spain

ARTICLE INFO

Keywords:
Schumann resonance
Stochastic process
Statistical distance
Kolmogorov–Smirnov distance

ABSTRACT

Schumann resonances' statistical parameters vary over time, which gives way to analyze them as a stochastic process. By splitting the Schumann resonance's records into segments and obtaining their empirical distribution function, the differences can be evaluated by calculating the Kolmogorov–Smirnov distance between them. The analysis' results allow for the characterization of the Schumann resonance's variations, along with the typical values it reaches under the chosen metric. It is shown how divergence is mitigated through sample averaging, and also how divergent samples impact on the Fast Fourier Transform algorithm.

Divergence quantification adds a layer of information for data processing. Knowing the changes experienced over time in Schumann resonances gives a way to know what kind of mathematical procedures can be applied to the signal. Quantification of signal variations over time can identify error sources in specific procedures, filter out samples unfit under certain analyses, or serve as a stop criteria for cumulative analyses.

1. Introduction

The *Schumann Resonance* (SR) is an electromagnetic resonant phenomenon occurring inside the cavity formed by the Earth's crust and the lower layers of the ionosphere, sourced by lightning activity all around the globe (Ogawa et al., 1969). Since they were finally measured in the decade of 1960 (Balser and Wagner, 1960) ten years after being theorized (Schumann, 1952), they have been steadily attracting the interest of the scientific community. Studies have proven the strong link between the first SR mode and the global tropical temperature (Williams, 1992). The relationship between the electromagnetic signal and lightning activity is well understood, with studies that mathematically infer the latter using the former (Boldi et al., 2018) up to the point that lightning monitoring through the analysis of SR records is more efficient than other lightning activity monitoring techniques (Williams et al., 2021). Other main line of research in steady development is looking for evidence of seismic activity on the earth–ionosphere resonances, which could significantly improve the early detection of this destructive phenomenon (Hayakawa, 1994; Cano-Domingo et al., 2022; Wang and Cao, 2011). Several studies have been published as well about how isolated and powerful solar and extra-solar events reaching Earth affect SRs (Nickolaenko et al., 2012; Dyrda et al., 2015; Sátori et al., 2016), and on the analysis of long periods of time to study the signal temporal patterns through its frequency and intensity variations (Galejs, 1970; Greenberg and Price, 2007; Domingo et al., 2021).

Nowadays, it is clear that the SR reflect the state of the Earth and the atmosphere (Besser, 2007). From the mathematical models developed, it can be seen how the frequency values of the SRs modes are dependent on the size of the resonant cavity and the charge distribution on the ionosphere (Barr, 1975). The layers that conform the waveguide, namely Earth's crust and the ionosphere, have an effect on their own as well. The ionosphere conductivity profile (distribution of charge as a function of height) has been a subject of study, as well as its impact on the SR (Jones, 1967), and in turn SR have been used to identify changes in the ionosphere due to external disturbances (Kudintseva et al., 2018). Earth's crust conductivity has an effect, albeit negligible, reason why it is usually ignored in most models (Kulak and Mlynarczyk, 2012). Despite the stability and periodicity of all these layers, they define mostly the signal's frequency modes.

Lightning activity, SR's main source of excitation, can be classified as a stochastic process (Artigas, 2012). Therefore, the intensity variations in the resonant signal produced by lightning can be studied from the perspective of a stochastic process as well (Surkov and Hayakawa, 2010), even more so when considering the other factors that may influence SR. Although this is not especially relevant when working on monthly or even daily timescales, on which their patterns and

* Corresponding author.
E-mail addresses: msolerortiz@ual.es (M. Soler-Ortiz), mfernandez@ual.es (M. Fernández-Ros), nnovas@ual.es (N. Novas-Castellano), jgazquez@ual.es (J.A. Gázquez-Parra).

https://doi.org/10.1016/j.jastp.2023.106058
Received 15 December 2022; Received in revised form 1 March 2023; Accepted 23 March 2023
Available online 3 April 2023

structures emerge, it has a noticeable impact when working with data segments.

This article proposes a direct, automatic way of quantifying SRs intensity variations over time under different situations using statistical methods. The proposed methodology is tested on real data and the results are evaluated to uncover any additional information that can be extracted from the procedure.

Providing SRs' variability along with the data enhances confidence in the results and helps other researchers reproduce the results by performing the analysis on samples with the appropriate characteristics. By having an independent metric to quantify the signal state, experimentally obtained criteria could be enhanced using this methodology. An example is the appropriate segment size and the number of averaged segments required for the resonances to manifest in the records (Nickolaenko and Hayakawa, 2014; Guha et al., 2017). While averaging a sufficient number of segments to ensure the resonances appear in the spectrogram is effective (as proven by all the meaningful results produced under it), advantages may arise from quantifying the stability of the signal, such as choosing the most stable segment to carry sensitive analyses or controlling the temporal resolution in frequency.

2. Methodology

This section will provide a detailed explanation of all the elements involved in the analysis. First, the most relevant characteristics of the analyzed data will be provided. Next, the available statistical methods for this process are described and their pros and cons are weighed. Following up, the specific techniques chosen to explore the signal's variability in time will be explained. Lastly, the analysis steps will be shown.

2.1. Data specifics

Two years of data serve as the baseline for this work. The data belongs to the years 2018 and 2019, and were captured by the ELF station our research group developed in 2013, and has been in continuous operation since 2015 (Parra et al., 2015). It is located in Sierra de los Filabres (Lat 37.226, Long −2.546), its closest landmark being Calar Alto astronomical observatory, Almería, Spain.

Only the data from the *East–West* (EW) channel will be used in this study. This decision comes from the fact that this channel is strongly influenced by the African thunderstorm center due to its orientation, while the effects of the other two main hotspots of lightning activity (American and Maritime continent) are mitigated (Belyaev et al., 1999). This simplifies the pattern search. The raw data, captured with a 0.1 Hz to 100 Hz bandwidth sensor at a sampling frequency of 187 Hz, has been filtered with a 4 Hz to 40 Hz band-pass filter to keep only the frequency range in which the SR is clearly measurable.

2.2. Quantifying temporal variations — theoretical framework

It is worth highlighting that this work is not meant to imply that the SR is a stochastic process, only that it can be treated as such. Evidences of many different patterns are widespread in the literature; variables that affect the SR and are stable considering the timescale of the resonant phenomenon. Ionospheric height and conductivity play a definitive role in the frequencies and average intensity of the signal (Sentman and Fraser, 1991). Disregarding anomalies caused by specific events, the ionosphere displays a daily cycle, which hardly influences the signal in the minute timescale. The solar cycle, which in turn influences the ionosphere conductivity, is another example of a stable pattern of a long period – 11 years – with a measurable effect on SR (Bozóki et al., 2021). Lastly, the major storm centers display seasonal and diurnal patterns as well, having a peak of activity between 14:00 and 15:00 at local time and being more active in the summer (Nieckarz et al., 2009). That being said, on the shorter timescale in which this article is based, the primary factor is the lightning discharges which are usually modeled as a stochastic process following the Poisson distribution (Chrissan and Fraser-Smith, 2003). By statistically comparing temporal segments of the SR the total contribution of lightning during that time is what determines the resulting statistical distribution. In consequence, by calculating the differences between several realizations of the time series mainly the stochastic source of the process is being considered.

When discussing a signal stability while modeling it as a stochastic process, each realization of the signal is observed as a statistical distribution through its histogram. To transform the time series into a certain amount of histograms two values are considered; the segment span – its duration – and the time passing from a segment's starting point to the next – the amount of time segments overlap with each other. With the former, the realization's sensitivity to transient events can be controlled, whereas the latter limits the amount of new information that each segment brings, in relation with its two neighboring segments.

Under the idea of a segment's statistical distribution being representative of the events captured by the signal in that period of time (Soler-Ortiz et al., 2021), this work will test the hypothesis of segments with low *Kolmogorov–Smirnov* (KS) distance value having similar characteristics.

Lastly, the procedure to test the signal variability over time will be measured by choosing a segment that will serve as representative, referred as the target segment, and comparing it to a specific number of neighboring segments, namely the comparison interval.

2.3. Statistical divergences and distances

There is a wide selection of generalized metrics whose purpose is to quantify the differences between statistical distributions. Among the statistical tests of goodness of fit, the Anderson–Darling test has been widely used in recent times due to its power and precision determining differences between samples (Engmann and Cousineau, 2011). The two sample test is performed by combining two observations X and Y which may contain a different number of samples. Eq. (1) shows the generalized formula for the test statistic considering, where n is the sample size of X, m is the sample size of Y, and Z is a set containing the combined and ordered samples of both X_n and Y_m. N_i represent the number of observations in X_n which value is less than or equal to the ith observation in the $Z_{(n+m)}$ set.

$$AD_D = \frac{1}{mn}\sum_{i=1}^{n+m}\left(N_i Z_{(n+m-ni)}\right)^2 \frac{1}{iZ_{(n+m-i)}} \tag{1}$$

The Anderson–Darling test is an example of widely used goodness of fit test due to its power, but as shown in Eq. (1) its setback is a more complicated formulation of its test statistic.

There are another family of tools to measure the difference between two distributions called f-divergences. These tools are theoretically developed in the field of statistical data processing (Basseville, 2013), and are a most relevant tool in information theory (Shannon, 1948), for example in the development of speech recognition techniques. Introduced in Rényi (1961), the general definition of f-divergences is based in the definition of the function $f(x)$ called generator, which has to fulfill certain conditions. It has to be finite for all $x > 0$, satisfy the equality $f(1) = 0$ and ensure continuity at $f(0)$ by fulfilling $f(0) = \lim_{t\to 0^+} f(t)$. This general definition is presented in Eq. (2), with P and Q being the *Cumulative Distribution Function* (CDF) of two distributions.

$$D_f = \int f\left(\frac{dP}{dQ}\right) dQ \tag{2}$$

Among the f-divergences, one of the best known is the Kullback–Leibler divergence (Kullback and Leibler, 1951), which is central not only to information theory, but also related to maximum likelihood estimation (Ekström, 2008) and even in optimization problems for

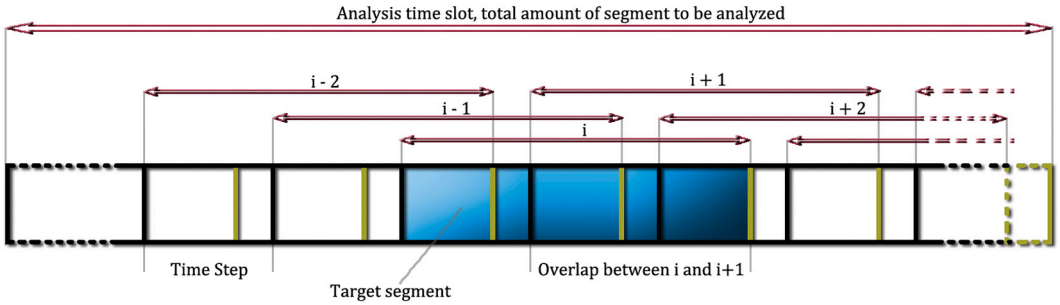


Fig. 1. Representation of time series slot, already segmented in N parts, each with a specific length and with a specific time between them.

power spectral density estimation for stochastic processes (Georgiou and Lindquist, 2003). The mathematical definition of the Kullback–Leibler divergence is presented in Eq. (3), with $P(x)$ and $Q(x)$ being discrete probability distributions.

$$KL_D = \sum P(x) \log_2 \left(\frac{P(x)}{Q(x)} \right) \tag{3}$$

That being said, the general metrics presented in this section widely used or developed in the mentioned contexts have two important drawbacks.

1. Although used as such, they do not qualify as metrics. For example, Kullback–Leibler is not symmetric and does not satisfy the triangle inequality. This is a problem in this context, since the lack of these or other characteristics associated with metrics would make the resulting values harder to understand.
2. They involve computationally intensive calculations. Since the goal of quantifying the SRs variations is to provide support or additional information to other analysis, introducing it as a costly algorithm defeats its purpose.

To provide a methodology able to meet the requirements set, something simpler, less computationally intensive, and more straightforward is needed. For these reasons, KS distance was chosen. This quantity measures the maximum absolute difference between the CDFs of both empirical distributions (Eq. (4), with $F(x)$ and $G(x)$ representing the CDFs of the two compared time segments) which fulfills all the previous conditions while effectively quantifying the difference between two statistical distributions, as it has been applied in other works for the same purpose (Baselice et al., 2019).

$$KS_d = sup_x(|F(x) - G(x)|) \tag{4}$$

It is important to highlight that it is the statistic of the KS test what was chosen, and not the p-value. The latter estimates from the statistic the probability of both realizations coming from the same distribution, which could be argued is a measure of similarity, but it has some setbacks such as being dependent on the number of samples.

2.4. Analysis' specifics

From the two previous subsections, it follows that the proposed analysis grants three degrees of freedom:

1. The segment's length in time.
2. The analysis' temporal resolution by means of the time each segment shares with its neighbors.
3. The time slot for which variability shall be considered, defined by the time interval from which neighboring segments will be compared against the target segment.

A graphical representation of the concepts involved in the analysis is shown in Fig. 1.

The scope and reach of a single variability test can be defined through these three variables. Increased time can be thus analyzed either by increasing the time slot or by sequentially repeating the process for all the time slots contained between two selected dates. The test itself is performed by segmenting the chosen interval of the signal after choosing the time slot size, the segment length, and the temporal distance between segments. Then, a specific segment is selected and its KS distance with all segments is calculated. The resulting curve is a representation of the statistical variation under the defined conditions.

That being said, it is important to mark that this test measures statistical variability. One can feel tempted to associate this variability with specific events, but to do so, more information is required about the circumstances when the signal was captured. This analysis measures the existing differences between the target segment's distribution and the rest of the segments' distributions in the accounted time slot. Noticeable variability points out the existence of transient events in the analyzed time slot, but the exact segment on which the transient event occurred cannot be pinpointed by this method. As stated before, the purpose of this analysis is only to determine the variability of the signal under certain conditions.

3. Results and discussion

In this section, the results of quantifying SR divergence using KS distance will be presented, in order to characterize the methodology.

At first, the analysis results will be outlined by presenting instances of the two general behaviors observed along the analyzed data. Up to some degree, when this analysis is performed with a high resolution (meaning the time difference between segments is considerably lower than the segment length), a level of base variance is appreciated, which reflects the stochastic nature of the studied process. The purpose is to locate more pronounced variations over time, noticeable over the baseline of the signal.

An example of each can be seen in the following figures. On one hand, Fig. 2 shows a stable sample with no significant trend changes and overall low variability along the whole interval whereas Fig. 3 displays significant variations over the base value.

The former serves as a baseline to quantify the standard level of randomness that the signal can exhibit, whereas the latter displays several differences when compared with the target segment. The target segment is compared with itself as part of the analysis' process and thus is identifiable for being the only data point showing KS distance of 0.0.

Although the two Figures present the analysis of different days, both share the same parameters. The analysis is performed in the interval from 15:00 to 16:00. The interval has been divided into segments of 5 min in length, with a difference of two seconds between each segment. This means the first segment goes from 15:00:00 to 15:05:00, the second segment goes from 15:00:02 to 15:05:02, and so on. The segments are identified by their starting time, which means that actually, the last analyzed segment goes from 16:00:00 to 16:05:00. In the same fashion, the target segment goes from 15:30:00 to 15:35:00.

Besides the base randomness value and the existence of trends in the statistical variation between temporal segments of a given time slot, other fact strikes as surprising. The KS distance was expected to

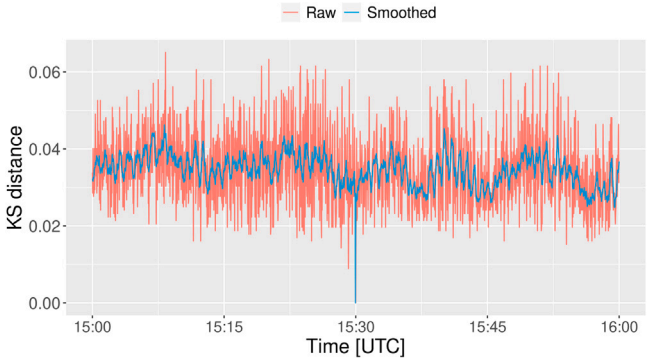


Fig. 2. Signal variability on June the 24th, 2018 from 15:00 UTC to 16:00 UTC — segments 5 min in length, 2 s difference between segments.

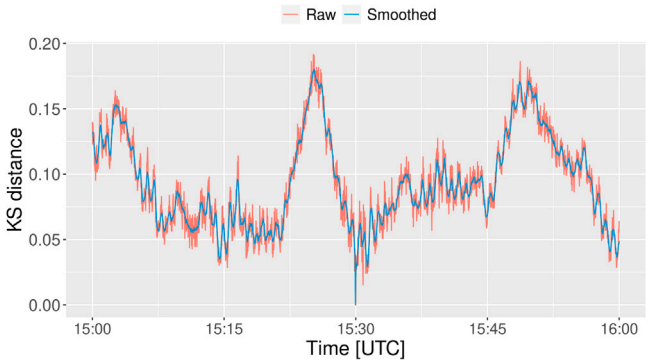


Fig. 3. Signal variability on April the 6th, 2019 from 15:00 UTC to 16:00 UTC — segments 5 min in length, 2 s difference between segments.

gradually increase from the 0.0 that indicates the target segment. This is so because the data point belonging to the comparison against a segment two minutes forward shares 3/5 of its information with the target segment. Despite this, it can be seen in Fig. 3 how segments further away (for example, the minimal values around 15:15) are more similar to the target segment than the ones that share information with the target segments, like the data points around the 15:25 peak.

As explained in the previous section, the relevant source of variations at this temporal resolution is lightning activity. Therefore, the variations shown in both Figs. 2 and 3 can be related to differences in thunderstorm activity over time. On one hand, the minimal variations displayed by Fig. 2 can be considered as a stable period of lightning activity, since all the analyzed segments are similar. On the other hand, the variations presented in Fig. 3 mean that during the analyzed hour, lightning activity was inconsistent.

One should keep in mind that the level of variation between the target segment at any other segment does not have any physical meaning by itself; it only quantifies how big are the differences between the compared segment and the target segment. Thus, one can safely assert that segments with different KS distance values can be attached to different states, but to assign physically meaningful states, one should resort to additional data. For example, in the cases presented above the information revealed by the figures is related to lightning activity, and if it was either stable or unstable. To actually relate the variations' results with the state of thunderstorm activity additional information is needed; at the very least, the global lightning count during the interval defined by the target segment.

Moving onwards, the major point of interest in testing the presented methodology is to offer some insight into the general trend changes such as the ones displayed in Fig. 3. Consequently, all the results will be smoothed as displayed in the figure. It is worth mentioning how the

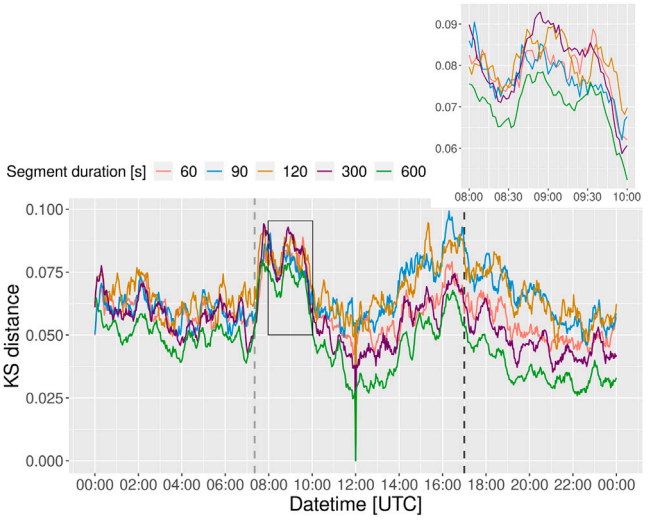


Fig. 4. Signal variability on December the 27th, 2018 — variable duration, 2 min difference between segments. The dashed lines mark the time of sunrise (light gray) and sunset (dark gray).

smoothed line still reaches a 0.0 value for the target segment; this is intended to make it identifiable at a glance.

To further characterize the resulting curve's shape and how the methodology's degrees of freedom affect the results, the parameters of the analysis must be accounted for. It is trivial that the time between segments affects the analysis resolution. Common sense drives the hypothesis of segment duration mitigating the average variation value for a given period, but a simple test reveals that this is not the case. Fig. 4 displays an analysis result where the length of the segments has been modified while keeping the step size (namely, the amount of time between a segment and the next) constant.

While it is true that modifying segment duration affects the average variation, their relationship is not direct. Fig. 4 shows how the segment duration that peaks in average variation are 90 and 120 s. It is also worth noticing how, on top of the lower average values displayed by the 600 s segments' analysis, it shows a heightened sensitivity to difference detection. Between 08:00 and 10:00 AM, average KS distance is doubled for the 600 s segments, reaching the values shared by the rest of the curves, while the comparative gain on the rest is not that high. On that same interval, it is the 300 s segments' curve that peaks in variance, as can be clearly seen in the inline zoom featured in the Figure. The variance increase for all segment lengths matching sunrise (light gray dashed line) is worth noticing, as well as the falling edge of a variance peak matching sunset (dark gray dashed line).

To provide further insight, Fig. 5 displays a similar procedure applied on a day with more pronounced variations.

In this case, the analysis with a segment duration of 120 s displays the highest average variation, especially noticeable in the second half of the day where the signal is more stable. Despite the differences appreciated between all curves from the target segment to sunset (Dark gray dashed line), after sunset the pattern of all curves display similar trends. This similarity is higher when comparing them by pairs, especially between 60 s (Salmon curve) and 90 s (Blue curve) and between 300 s (Purple curve) and 600 s (Green curve). Once again, a variability peak's rising edge matches sunrise although no peak matches sunset.

In this initial stage of methodology testing, the reason why the average variability does not linearly decrease with segment duration can only be the subject of speculation. It may be due to the value of the rest of the parameters or related to some disturbance or behavior present on the records. The former requires extensive testing since the many tests realized during the development of this work yielded no

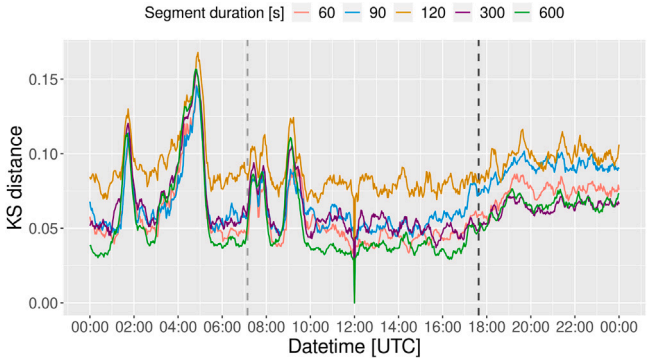


Fig. 5. Signal variability on February the 4th, 2018 — variable duration, 2 min difference between segments. The dashed lines mark the time of sunrise (light gray) and sunset (dark gray).

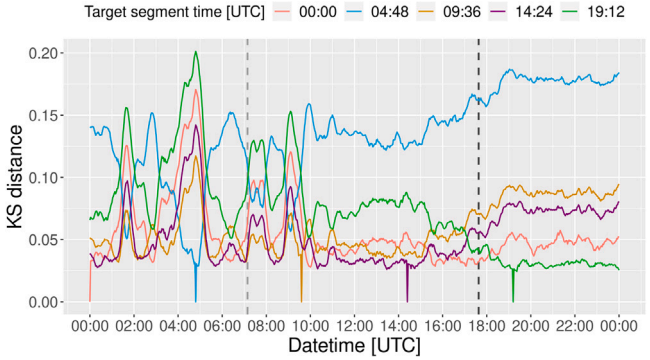


Fig. 6. Signal variability on February the 4th, 2018 — 10 min length, 2 min difference between segments. The dashed lines mark the time of sunrise (light gray) and sunset (dark gray).

specific evidence. The latter is suggested by how in Fig. 4 the curves of different segments display the maximum variability at different periods. It will require a detailed study of a period well characterized with information about the additional variables that define the SR (lightning activity, ionosphere state, etc.).

To consider the weight of the choice of the target segment, an analysis has been performed multiple times targeting different segments inside the same interval. An example of this is shown in Fig. 6 performed on the same day as Fig. 5 which displays pronounced variations during the early hours of the day while staying stable for the rest of the day.

Once again, the temporal location of each target segment is given by the moment where the curve value in the y-axis is 0.0, as well as highlighted in the Figure's legend. An interesting detail is how the analysis for the target segment of 04:48 UTC (Blue curve) is matched with a maximum variation point for the analysis of 19:12 UTC (Green curve). This causes these two curves to almost mirror each other; despite certain differences, the same peaks and troughs that appeared in Fig. 5 can be appreciated. In the same fashion, the rest of the segments follow the green curve's tendency during most of the first half (up to 10:00 UTC), although the maximum value they reach during each peak has different magnitudes. In the second half, the 04:48 UTC segment (Blue curve) is the one with the highest divergence.

Among the rest there are trend shifts worth noticing. The 19:12 UTC segment keeps its variability values above the other three until approximately 17:00 UTC, where it is surpassed by the 09:36 UTC analysis (Orange curve). It is also interesting how the variability of the 00:00 UTC, 09:36 UTC, and 14:24 UTC analyses (Salmon, orange, and purple curves respectively) stop having a clear, common trend during the 10:00 UTC to 15:00 UTC section. From this point onward,

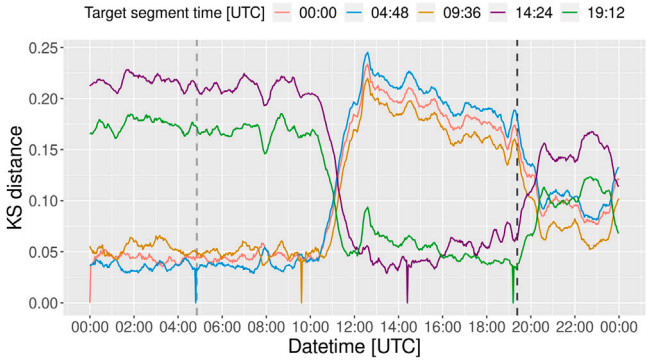


Fig. 7. Signal variability on June the 2nd, 2018 — 10 min length, 2 min difference between segments. The dashed lines mark the time of sunrise (light gray) and sunset (dark gray).

the orange and purple curves match each other and mimic the trend marked by the 04:48 UTC variability. Just as well, there are similarities between the variability of these three segments and that from the 00:00 UTC segment after 19:00 UTC.

The way different segment's variability curves share similar trends along specific sections points out certain level of independence from the chosen segment. In other words, while there are differences in absolute variability between any chosen segment and a segment of the signal, the general similarity between trends means variability over time can be appreciated choosing any given segment. Even in segments containing isolated events such as the one at 04:48 UTC (blue curve) the trend is present, albeit inverted. The variability between the signal and this target segment shows the possibilities of choosing an odd segment; the variability is high through most of the curve, giving reasons to expect similarities between the target segments and those segments whose variability is lower.

In Fig. 7 the results of an analysis under the same parameters are displayed, but in this case, it shows variation distributed in three clear zones.

The maximum/minimum value inversion behavior is similar to the one displayed in Fig. 6 but, in this case, no curve clearly mirrors the other. Most of them tend to follow the same tendency (especially in the section between 12:00 UTC and 21:00 UTC) with only the curve with the lowest average value (Segment in 14:24 UTC, with the purple curve) per section following a different tendency. The sudden shift that can be appreciated around 11:00 UTC is of interest, since it marks a boundary between the nighttime and daytime realizations. The most likely explanation should be the ionosphere change at sunrise, but since the time when the inversion happens is six hours after sunrise (gray dashed line) takes away the weight of the hypothesis. It is also worth noticing the absence of sudden variations around sunrise (light gray dashed line), as observed in the rest of Figures. Nonetheless, all curves increase their steepness around sunset (dark gray dashed line).

Once again it can be seen that total variation can be characterized by any segment, regardless of its location along the time slot. The behaviors displayed by different segments point once again to different states contained within each temporal slot.

A variation overview for all the analyzed data is shown in Fig. 8. All chosen data was subjected to hourly analysis, using segments of 10 min in length and with a time difference of 2 min between each other. All the results are hourly averaged by month and the resulting mean and statistical deviation by hours is displayed here, for both years.

First, it clearly states the main advantage of averaging a signal such as the SR in monthly segments; the average KS distance stays almost ever below 0.06, being most usually around 0.04. It points out how averaging reduces significantly the variation, allowing many requirements to be fulfilled. That being said, it is interesting how

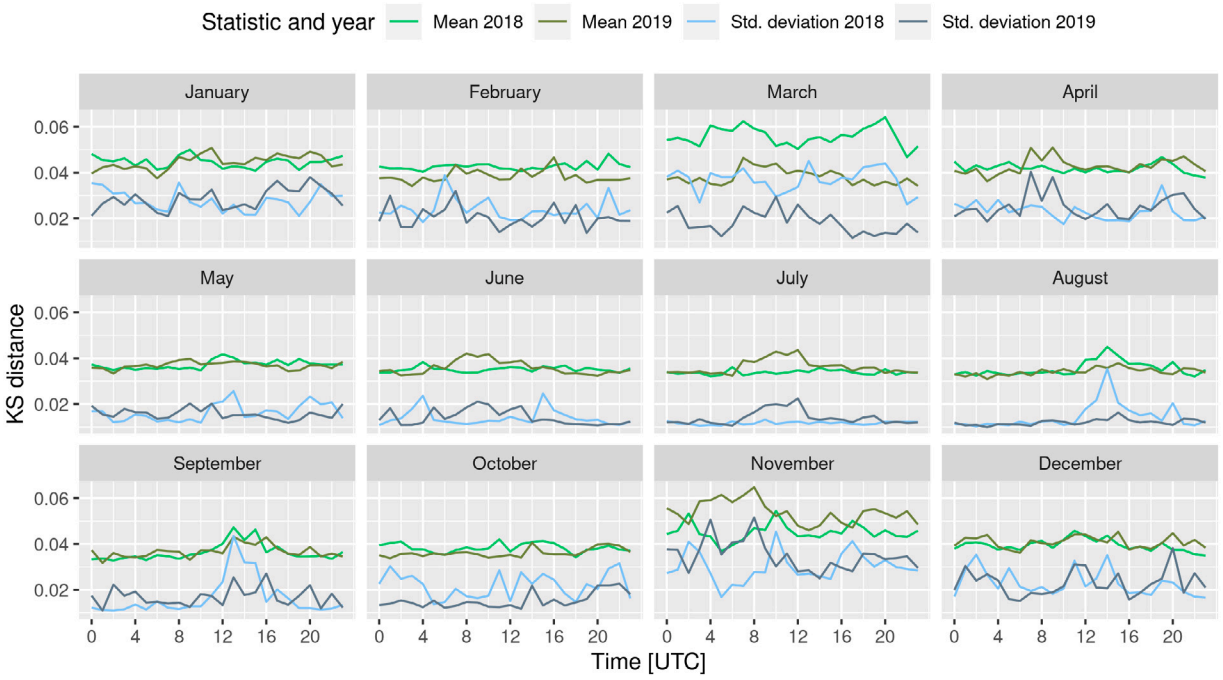


Fig. 8. General results, showing the hourly mean and statistical deviation of each month, for the two analyzed years.

the statistical deviation is more sensitive than the average, the latter following the trend of the former almost to a fault but acquiring higher relative values as well. Lastly, it is worth noticing how the variations, although still reduced, are more pronounced in the first and last quadrimester of the year, whereas in the middle quadrimester the signal is more stable. This is consistent with the properties of the African thunderstorm cycle, which reaches a minimum activity level during late spring and summer. On the other hand, the peak of daily activity for this thunderstorm center happens at 15:00 UTC (Nickolaenko et al., 1998), but the variation analysis does not show a clear maximum at this time, or even around it.

Lastly, a simple test is performed to present in a practical way the effects of this measurement, its results displayed in Fig. 9.

In this figure, the results of a variation analysis can be seen, with the variation's excursions showing how the time interval is one where the signal changed noticeably in small amounts of time. Below, three sets of frequency spectra calculated with the FFT algorithm using the same signal segments involved in the variation analysis are shown. These have 7 s length and 7 s between each other, closely matching the recommendations from Nickolaenko and Hayakawa (2014). Nonetheless, the FFTs has been calculated using different sets of segments. For instance, the spectrum colored in orange has been calculated using the segments with low variation (segments whose variation is under the orange line, a total of 52 segments) whereas the blue one comes from those with high variation (the 58 segments whose variation places them over the blue line). Finally, the pale red spectrum was calculated using the first 58 segments of the hour.

Nickolaenko and Hayakawa (2014) state how the first four modes are clearly visible after averaging 7 minutes' worth of segments, and recognizable after 2 min. It is understood that this rule of thumb is applicable to clear segments with no interference. In this test, the FFTs were calculated using almost 7 min of data, (albeit a bit less for the segments with lower values of variation) but, due to the noisy nature of the segment, the SR modes are hard to notice in the FFT calculated just by averaging the spectra of the segments sequentially. On the contrary, the other two spectra show the first three modes, showing how by choosing segments with a similar variation the number of segments to produce a clear spectrum is reduced. That being said, the spectrum produced by using similar segments (IE those with lower values of variation) is clearer than the one calculated using the divergent segment, despite having fewer samples. An explanation can be reasoned under the constraints of this analysis. There is no doubt that segments with a low KS distance will be similar to the target segment, and by extension similar to each other. On the other hand, segments with a high variation are different from the target segment, but they are not necessarily similar between them.

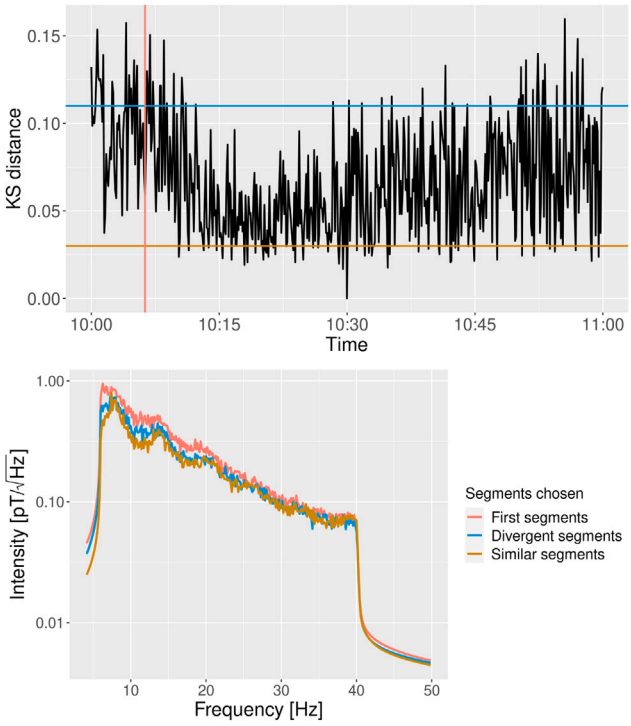


Fig. 9. Practical analysis to present the effects of variation, showing the signal variability for April the 20th, 2018 with segments length of 7 s, 7 s difference between segments, and the FFTs calculated from different subsets of the sample.

This test serves its purpose by proving the relevancy of segment variation in signals whose variations over time allow the interpretation of the signal as a stochastic process, such as the SR. In this case, the procedure has shown a way of choosing a set of time segments that provide the best resolution of the signal spectra. Given the nature of the calculation, the opposite can happen if the segment chosen as the target segment was especially odd since the procedure would consider it similar to the rest of odd segments. Once again, it must be remarked how the utility of variation calculation may be boosted if the characteristics of the target segment are properly identified.

4. Conclusions

In this work, we have presented a methodology based on the use of a fairly common metric, namely the Kolmogorov–Smirnov distance, to quantify the variation between temporal segments of the Schumann Resonances. The possibilities and results of the methodology are explored showing the effect of modifying its different parameters.

To the best of our knowledge, this procedure has not been widely applied to stochastic processes. Its interest and applications are multiple; going from ensuring that the chosen data fulfills similarity criteria, to the fine-tuning of analysis' parameters such as sample size or number of repetitions. This interest has been explored by showing how the spectra calculated using segments of similar variation display the Schumann resonance modes more clearly than a regular spectra, calculated by picking up the first N segments in a sequential fashion.

Through the application of this tool to the Schumann resonance records, an average variation value of 0.04 between segments containing background noise has been observed. In the same fashion, the base variation on the background noise has been estimated, ranging from 0.02 to 0.06. When a given time slot displays variation, a typical maximum value of 0.25 is observed, although a few results reach higher values.

Specific situations where the analysis was applied have been shown, and through them its potential. The possibility of detecting low variation segments of time along with the expected relationship with lightning activity displayed by the results makes this methodology an interesting complement to more in-depth analyses. Among the future developments of this line of research, it would be possible to pair this methodology with additional data about phenomena that may influence the Schumann Resonances, such as lightning activity, ionospheric charge or seismic activity. This may give way to infer different states on specific variables, knowing the state of the target segment, or perceiving the presence of undetected transient events by comparing averaged temporal variations between the variations obtained from a specific segment.

CRediT authorship contribution statement

Manuel Soler-Ortiz: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Manuel Fernández-Ros:** Resources, Data Curation, Writing – review & editing, Supervision. **Nuria Novas-Castellano:** Resources, Data Curation, Writing – review & editing, Visualization. **Jose A. Gázquez-Parra:** Resources, Data Curation, Writing – Review & Editing, Project Administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

The Ministry of Economics and Competitiveness of Spain financed this work, under Project TEC2014-60132-P, in part by Innovation, Science and Enterprise, Andalusian Regional Government and in part by the European Union FEDER Program and CIAMBITAL Group. By I+D+I Project UAL18-TIC-A025-A, Proyecto Puente 2021/001, the University of Almeria. We also thank the Andalusian Institute of Geophysics.

References

Artigas, S.E.S., 2012. Stochastic modeling of lightning occurrence by nonhomogeneous poisson process. In: 2012 International Conference on Lightning Protection. ICLP, IEEE, pp. 1–7.

Balser, M., Wagner, C., 1960. Observations of earth–ionosphere cavity resonances. Nature 188 (4751).

Barr, R., 1975. An approximate method for the evaluation of the ELF reflection coefficients of an inhomogeneous, anisotropic daytime ionosphere and its application to the solution of the earth-ionosphere waveguide mode equation. J. Atmos. Terr. Phys. 37 (11), 1405–1412.

Baselice, F., Ferraioli, G., Pascazio, V., Sorriso, A., 2019. Denoising of MR images using Kolmogorov-Smirnov distance in a non local framework. Magn. Reson. Imaging 57, 176–193.

Basseville, M., 2013. Divergence measures for statistical data processing—An annotated bibliography. Signal Process. 93 (4), 621–633.

Belyaev, G., Schekotov, A.Y., Shvets, A., Nickolaenko, A., 1999. Schumann resonances observed using poynting vector spectra. J. Atmos. Sol.-Terr. Phys. 61 (10), 751–763.

Besser, B.P., 2007. Synopsis of the historical development of Schumann resonances. Radio Sci. 42 (2), http://dx.doi.org/10.1029/2006RS003495.

Boldi, R., Williams, E., Guha, A., 2018. Determination of the global-average charge moment of a lightning flash using Schumann resonances and the LIS/OTD lightning data. J. Geophys. Res.: Atmos. 123 (1), 108–123.

Bozóki, T., Sátori, G., Williams, E., Mironova, I., Steinbach, P., Bland, E.C., Koloskov, A., Yampolski, Y.M., Budanov, O.V., Neska, M., et al., 2021. Solar cycle-modulated deformation of the earth–Ionosphere cavity. Front. Earth Sci. 9, 689127.

Cano-Domingo, C., Stoean, R., Novas-Castellano, N., Fernandez-Ros, M., Joya, G., Gázquez-Parra, J.A., 2022. On the prospective use of deep learning systems for earthquake forecasting over Schumann resonances signals. Eng. Proc. 18 (1), 15.

Chrissan, D., Fraser-Smith, A., 2003. A clustering Poisson model for characterizing the interarrival times of sferics. Radio Sci. 38 (4), 17 1–17 14.

Domingo, C.C., Ros, M.F., Castellano, N.N., Parra, J.A.G., 2021. Diurnal and seasonal results of the Schumann resonance observatory in Sierra de Filabres, Spain. IEEE Trans. Antennas and Propagation 69 (10), 6680–6690.

Dyrda, M., Kulak, A., Mlynarczyk, J., Ostrowski, M., 2015. Novel analysis of a sudden ionospheric disturbance using Schumann resonance measurements. J. Geophys. Res. Space Phys. 120 (3), 2255–2262.

Ekström, M., 2008. Alternatives to maximum likelihood estimation based on spacings and the Kullback–Leibler divergence. J. Statist. Plann. Inference 138 (6), 1778–1791.

Engmann, S., Cousineau, D., 2011. Comparing distributions: the two-sample Anderson-darling test as an alternative to the Kolmogorov-Smirnoff test. J. Appl. Quant. Methods 6 (3).

Galejs, J., 1970. Frequency variations of Schumann resonances. J. Geophys. Res. 75 (16).

Georgiou, T.T., Lindquist, A., 2003. Kullback-Leibler approximation of spectral density functions. IEEE Trans. Inform. Theory 49 (11), 2910–2917.

Greenberg, E., Price, C., 2007. Diurnal variations of ELF transients and background noise in the Schumann resonance band. Radio Sci. 42 (2), http://dx.doi.org/10.1029/2006rs003477.

Guha, A., Williams, E., Boldi, R., Sátori, G., Nagy, T., Bór, J., Montanya, J., Ortega, P., 2017. Aliasing of the Schumann resonance background signal by sprite-associated Q-bursts. J. Atmos. Sol.-Terr. Phys. 165.

Hayakawa, M., 1994. Electromagnetic Phenomena Related to Earthquake Prediction. Terra Sci. Pub. Comp., Tokyo, 1994.

Jones, D.L., 1967. Schumann resonances and ELF propagation for inhomogeneous, isotropic ionosphere profiles. J. Atmos. Terr. Phys. 29 (9), 1037–1044.

Kudintseva, I., Galuk, Y.P., Nickolaenko, A., Hayakawa, M., 2018. Modifications of middle atmosphere conductivity during sudden ionospheric disturbances deduced from changes of Schumann resonance peak frequencies. Radio Sci. 53 (5), 670–682.

Kulak, A., Mlynarczyk, J., 2012. ELF propagation parameters for the ground-ionosphere waveguide with finite ground conductivity. IEEE Trans. Antennas and Propagation 61 (4), 2269–2275.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Nickolaenko, A., Hayakawa, M., 2014. Schumann resonance for tyros: Essentials of global electromagnetic resonance in the earth-ionosphere cavity. p. 1, 348. http://dx.doi.org/10.1007/978-4-431-54358-9.

Nickolaenko, A.P., Kudintseva, I.G., Pechony, O., Hayakawa, M., Hobara, Y., Tanaka, Y.T., 2012. The effect of a gamma ray flare on Schumann resonances. Ann. Geophys. 30 (9), http://dx.doi.org/10.5194/angeo-30-1321-2012.

Nickolaenko, A., Sátori, G., Zieger, B., Rabinowicz, L., Kudintseva, I., 1998. Parameters of global thunderstorm activity deduced from the long-term Schumann resonance records. J. Atmos. Sol.-Terr. Phys. 60 (3), 387–399.

Nieckarz, Z., Zięba, S., Kułak, A., Michalec, A., 2009. Study of the periodicities of lightning activity in three main thunderstorm centers based on Schumann resonance measurements. Mon. Weather Rev. 137 (12), 4401–4409.

Ogawa, T., Tanaka, Y., Yasuhara, M., 1969. Schumann resonances and worldwide thunderstorm activity: —Diurnal variations of the resonant power of natural noises in the earth-ionosphere cavity—. J. Geomagn. Geoelectr. 21 (1), 447–452. http://dx.doi.org/10.5636/jgg.21.447.

Parra, J.A., Ros, M.F., Castellano, N.N., Salvador, R.M., 2015. Techniques for Schumann resonance measurements: A comparison of four amplifiers with a noise floor estimate. IEEE Trans. Instrum. Meas. 64 (10), 2759–2768. http://dx.doi.org/10.1109/TIM.2015.2420376.

Rényi, A., 1961. On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions To the Theory of Statistics, Vol. 4. University of California Press, pp. 547–562.

Sátori, G., Williams, E., Price, C., Boldi, R., Koloskov, A., Yampolski, Y., Guha, A., Barta, V., 2016. Effects of energetic solar emissions on the earth–ionosphere cavity of Schumann resonances. Surv. Geophys. 37, 757–789.

Schumann, W.O., 1952. Über die strahlungslosen Eigenschwingungen einer leitenden Kugel, die von einer Luftschicht und einer Ionosphärenhülle umgeben ist. Zeitschrift Fur Naturforschung - Section A J. Phys. Sci. 7 (2), http://dx.doi.org/10.1515/zna-1952-0202.

Sentman, D., Fraser, B., 1991. Simultaneous observations of Schumann resonances in California and Australia: Evidence for intensity modulation by the local height of the D region. J. Geophys. Res. Space Phys. 96 (A9).

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3).

Soler-Ortiz, M., Ros, M.F., Castellano, N.N., Parra, J.A.G., 2021. A new way of analyzing the Schumann resonances: A statistical approach. IEEE Trans. Instrum. Meas. 70, 1–11.

Surkov, V., Hayakawa, M., 2010. Schumann resonances excitation due to positive and negative cloud-to-ground lightning. J. Geophys. Res.: Atmos. 115 (D4).

Wang, Y., Cao, Q., 2011. Analysis of seismic electromagnetic phenomena using the FDTD method. IEEE Trans. Antennas and Propagation 59 (11), 4171–4180.

Williams, E.R., 1992. The Schumann resonance: A global tropical thermometer. Science 256 (5060).

Williams, E., Bozóki, T., Sátori, G., Price, C., Steinbach, P., Guha, A., Liu, Y., Beggan, C., Neska, M., Boldi, R., et al., 2021. Evolution of global lightning in the transition from cold to warm phase preceding two super El Niño events. J. Geophys. Res.: Atmos. 126 (3), e2020JD033526.