Study of the statistical footprint of lightning activity on the Schumann Resonance

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

December 2023

Abstract

The Schumann resonance is an electromagnetic phenomenon, a product of lightning activity inside the earth-ionosphere cavity. Five years of Schumann resonance records are analyzed by a novel methodology that segments the records into time intervals and finds the probability distribution that best describes each segment. Then patterns are extracted from the resulting time series and compared against known patterns of global lightning activity to further test the power of the methodology under study.

The Quality of Fit indices show how over 95% of the segments analyzed are properly described by the distribution that fit them best. The relationship between global lightning activity and the number of segments identified as Gaussian emerges clearly. A link between Laplacian segments and local lightning activity is explored as well. This presents transient statistical fitting as an alternative for characterizing complex phenomena by identifying different segments with a probability distribution, then identifying circumstances that segments with the same distributions have in common.

This study further validates the chosen analysis tool, showing its capacity to offer information on the activity of thunderstorms in the time segments analyzed from the Schumann resonance data. It presents an additional source of information that complements the usual techniques used to study the signal in the frequency domain.

1 Introduction

The electromagnetic resonant signal known as the *Schumann Resonance* (SR), which has a fundamental mode of $7.8\,\text{Hz}$ is a subject of study due to its central role in the electric circuit of the Earth (Williams and Mareev, 2014) and especially its relationship with atmospheric and climate phenomena. The work by Williams (1992) on which the relationship between global temperature and the intensity of the first mode was discovered is an extensively cited example. Numerous works have studied different geophysical factors through SRs, such as the state of the resonant cavity itself (Surkov et al., 2013), extracting models of the atmospheric conductivity profile (Nickolaenko et al., 2016; Kudintseva et al., 2018), tracking climate phenomena such as El Niño Southern Oscillation through SR records (Yang and Pasko, 2007; Williams et al., 2021), and finding evidence of SR variations related to other climate phenomena such as the Madden-Julian oscillation (Beggan and Musur, 2019). Studies on transient events have also been performed, such as the impact of solar flares (Shvets et al., 2017), gamma rays (Nickolaenko et al., 2012), volcanic eruptions (Nickolaenko et al., 2022; Bór et al., 2023), and especially earthquakes (Hayakawa et al., 2005; Gazquez et al., 2017; Figueredo et al., 2021), whose interest has grown in recent decades.

Despite the considerable number of factors that influence SRs, it is well established that the phenomenon that produces the frequency response of the earth-ionosphere cavity is lightning discharges around the globe (Ogawa et al., 1969). Global lightning activity is generally detected using the *Very Low Frequency* (VLF) range (Ogawa and Komatsu, 2010). In the *Extremely Low Frequency* (ELF) range, thunderstorm discharges are studied by their relationship to SR (Price, 2016). Alpha stable distributions are a common choice to represent this phenomena (Chrissan and Fraser-Smith, 2000) since they are the tool of choice in telecommunications (Volland, 1995). They are especially useful for analyzing radio electric noise in a wide range of frequencies and under different conditions Lamey et al. (2021). Models have also been proposed that solve the inverse problem of deducing the amount of lightning activity from the SR records by providing approximate solutions (Heckman et al., 1998; Nickolaenko et al., 1998). Annual variability in SR records is also a subject of study, with some analyses indicating additional links to lightning activity (Domingo et al., 2021). As a consequence, the signal patterns caused by lightning are well known.

The behavior of lightning activity in specific zones is still a topic of interest (Xu et al., 2022). Especially for each of the three thunderstorm centers (Nieckarz et al., 2009; Ouyang et al., 2015), which are the regions of the globe with the highest concentration of lightning activity, located in South America, Africa and Asia, specifically in the maritime continent. There are works in which the activity of one (Dyrda et al., 2014) or all (Prácser et al., 2019) thunderstorm centers is characterized, using different techniques with promising results.

There is abundant literature on the relationship between lightning activity and SR from different perspectives, such as mathematical modeling (Nickolaenko, 1997; Pechony and Price, 2004), experimental studies relating SR mea-

surements and lightning activity (Price, 2016; Tatsis et al., 2021), and computer models to simulate its signal or key aspects (Kudintseva et al., 2017; Perotoni, 2018; Ralchenko et al., 2015). More recently, data-driven approximations have been taken, applying forecasting methods (Cano-Domingo et al., 2023; Tulunay et al., 2008) based on neural networks of specific purpose (Cano-Domingo et al., 2022b; Tulunay et al., 2004). However, the many different events that influence the SR have prevented the scientific community from providing an accurate model that takes into account every situation. Most analytical solutions are developed for the ideal case, and even if more advanced, computer-based models may produce a satisfactory response, they are mostly focused on presenting the ELF background noise under unchanging variables The point to be made is that, despite the advances provided by the scientific community in the field, SR is a complex phenomenon affected by many natural and artificial events, and modeling it faithfully is a daunting task.

Most ELF studies are based on data extracted from the frequency domain, with only a few notable exceptions (Nieckarz et al., 2009; Ogawa and Komatsu, 2009; A. Nickolaenko, P. Colin, 2000; Cano-Domingo et al., 2022a; Gowanlock et al., 2018). We recently presented a publication based on time domain analysis using statistical methods (Soler-Ortiz et al., 2021). The philosophy behind the method is to divide the time records into segments, fit each segment to a simple model, and extract conclusions about the signal state depending on the chosen model. The idea is not new in the domain of electromagnetic signals (Alata et al., 2013). Some studies have worked to statistically characterize ELF noise (Evans and Griffiths, 1974) with the purpose of using the band for communications. Close to our previous work is the study by Chrissan and Fraser-Smith (2000), in which the best model for impulsive atmospheric noise is tested, considering different bandwidths among the ELF and VLF frequency ranges. This work use amplitude probability distributions described in terms of voltage deviation, which rely on a great amount of data to produce a faithful model but disregarding completely the temporal aspect, and also chooses the statistical model that showed the best performance on each band. On the other hand, our previous study focused on the $0\,\mathrm{Hz}$ to $100\,\mathrm{Hz}$ band and specifically in the $6\,\mathrm{Hz}$ to $40\,\mathrm{Hz}$ range where the SR gathers most of its power and works under a different paradigm. The methodology we developed finds the statistical model among a set of chosen distributions that showed the best fit for each segment. Its most interesting contribution is how the distribution that best fits the segment varied depending on the time interval from which the sample was taken, hinting a qualitative relationship between the state of the atmosphere and the chosen model.

This article expands on the relationships described in Soler-Ortiz et al. (2021), testing the advantages and capabilities of the developed technique. To do so, the analysis technique has been applied to five years of SR data. The resulting in-depth analysis establishes further relationships between the classification provided by the analysis and the known facts and patterns on SR and lightning activity. This will be done by comparing the known patterns of lightning activity with the of segments per day associated with a specific

3

distribution.

With this methodology, a direct way to identify segments where lightning activity is detectable from within a large amount of SR records is provided. This, in turn, gives way to the possibility of identifying more processes through their statistical shape, and also provides an additional level of information for the studies on this field and is complementary to the usual frequency-based approaches.

2 Methodology

The mentioned data comes from the research group's ELF measurement station in Sierra de los Filabres, Almería (Lat 37.22, Long -2.55), with a bandwidth of 100 Hz. The electronic components of the system are thoroughly described in Parra et al. (2015). The station has two sensors, one with orientation *East-West* (EW), while the other is *North-South* (NS) oriented, each reading stored separately in 30 min files on our data server, at the University of Almería. The sampling frequency of the measuring equipment is 187 Hz, which is equivalent to 336600 samples per file. Data from both channels EW and NS have been analyzed.

Its remote location away from the electrical networks required the autonomy of the system that uses batteries and solar panels. Given its remoteness, data are sent from the station to the server via a radio link.

These characteristics allow continuous data capture and storage, which makes it possible to gather the data used for this study. However, these elements may also be prone to failure. Continuous cloudy weather prevents batteries from being charged via solar power, and atmospheric phenomena may, for example, disrupt radio link communication. Maintenance or upgrade of the system can cause longer outages. In conclusion, if rigorousness is to be maximized, data imputation techniques must be taken into account.

2.1 Data imputation in time series

Time series are especially sensitive to missing values; ideally the whole register should be complete since each segment belongs to a different time frame and the information they bring is different (Khayati et al., 2020), so *Not Available* (NA) values must be estimated. This makes data imputation in time series a subject of recent studies (Afrifa-Yamoah et al., 2020; Zefreh and Torok, 2018; Zhang and Thorburn, 2021), especially for real-time analysis.

The development of imputation methods is applied mainly in research where the events studied have a time frame similar to the instrument's capture frequency (Betrie et al., 2016). There are also studies in which the importance of data imputation in aggregated data is evaluated (Seiler and Heumann, 2013). Data aggregation comes with the setback of obscuring effects that are only observable under the aggregation threshold (Cook and Weidmann, 2019). Since the purpose of this work is to study time trends - yearly, monthly, and daily

4

patterns - that can be considered long with respect to the number of data points (60 data points per hour), the data can be aggregated in periods that do not obscure the phenomena under study while mitigating the impact of NA values.

In light of this situation, NA values' propagation is prevented as long as the number of NA values present in the aggregation does not exceed a certain threshold. This is based on the aggregation of any remaining values being more informative than any other approximation that might be taken. If the aggregation is part of some trend extraction process (such as a *Moving Average* (MA)) the gap will be covered by linear interpolation.

To provide the reader with data integrity information, a color-coded bar like the one shown in Fig.1 will be displayed under the relevant figures, showing the fraction of missing values implied in the calculation of each day's value.
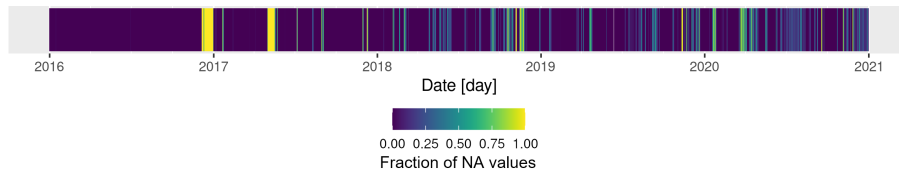

Figure 1: A color-coded bar showing the fraction of NA values involved in the averaging of each day.

2.2 Time Domain Analysis by Segmentation and Distribution Fitting

The statistical analysis tool used in this study was presented in Soler-Ortiz et al. (2021), where the intricacies of the methodology are presented, including examples of different cases, as well as the initial assessment of the activity of lightning that led to this study. To facilitate the reading of this work, some highlights of the methodology will be presented in this section.

The results of this analysis are based on the SR records filtered with a 6 Hz to 40 Hz band-pass filter, this being the frequency band that contains the first six modes of SRs (Domingo et al., 2021). The records are then divided into segments of a specified duration, and the data points are presented as histograms with a number of bins equal to the square root of the number of data points in a segment - having 112200 data points for a 10 min segment, and 11220 for 1 min segments. When showing the histograms for each segment, the bins were capped at 100 to allow the bins to be visible. For each histogram, its central moments are extracted, and then the data of the segment undergoes a fit process using Maximum Likelihood Estimates for each of the chosen target distributions. Then, the one that best fits the data is chosen using a variation of *Akaike Information Criterion* (AIC) called Akaike weights (Wagenmakers and Farrell, 2004). This transformation allows the AIC values of different fitting processes on the same data, which are estimates of information loss if the fitted

5

data were represented by the chosen model, to be compared with each other. Thus, Akaike weights are a direct representation of the amount of information loss each model incurs compared to the rest. The specific application of Akaike weights to this problem can be consulted in Soler-Ortiz et al. (2021), whereas the Akaike weights method itself is thoroughly described in Wagenmakers and Farrell (2004).

It must be clarified that the chosen distribution for a given segment does not have to describe the entire data set correctly. Consequently, a quality of fit index is defined based on the concept of QQ plots, in which the quantiles of a given data set are plotted against those of a chosen statistical distribution - choosing as many quantiles as data points a segment has. A perfect fit between the data and the distribution is recognized as a segment coincident with the identity line. As the quality index for each fitting process, the correlation coefficient value of a linear regression between the data points of the segment presented and the chosen distribution's points presented in QQ plot is chosen.

The analysis' specifics are as follows:

- The set of target distributions for this work are Logistic, Gaussian, and Laplacian, and the quality of fit for a segment to be considered as appropriately fit is still 99%.

- All segments whose quality of fit is below 99% are deemed as Unclassified for this work.

- The segment lengths for this analysis are 1 min and 10 min. The difference in duration has been related to the ability of transient phenomena to influence the shape of the distribution (Soler-Ortiz et al., 2021). In other words, shorter segments have a higher sensitivity to transient events, while longer segments require intense activity to affect their shape.

- The possibility of analyzing overlapping segments has been introduced in the analysis, which gives the possibility to smooth the transition between one sample and the next by controlling the amount of time they share. The 10 min duration segments are taken with an overlap of 9 min, to produce approximately the same amount of samples as with the 1 min segments.

Fig. 2 presents an overview of the way data is processed to produce the results presented in this work.

2.3 Methodology strengths and potential

The methodology used to produce this article's results has already been tested and validated (Soler-Ortiz et al., 2021), showing how it can process a given amount of data automatically, and how its results (in the format of the number of segments per hour whose best fit is the Gaussian distribution) show a relationship with global lightning activity.

It is through these similarities that this methodology displays an usefulness not explored through the use of other noise models such as alpha stable distributions or Hall models. Although there are research lines to monitor global lightning activity through the SR by means of spectral analysis, these are still in the early stages (Bozóki et al., 2023). This is why the possibilities offered by this qualitative relationship are the focus of this article. The main hypothesis to be explored is that SR's temporal segments background noise where noticeable global lightning activity is present display a Gaussian distribution. Otherwise, the best fit for SR background noise will be the Logistic distribution (Soler-Ortiz et al., 2021). Since both distributions feature in segments where the signal displayed is the SR background noise, those whose best fit is the Gaussian distribution will be called storm-active background noise, whereas those fitted to the Logistic Distribution will be considered steady-state background noise. Through the evidence presented in this article, it is intended to link the different distributions with different states of global lightning activity. This, in turn, would mean that by checking the distribution that fits a temporal segment of the SR best, the state of global lightning activity could be qualitatively assessed.

3   Results and Discussion

In this section, the general results of the analysis will be presented first, as they are compared with the known general behavior of the signal studied. Next, different data aggregations will be shown to inquire about the implications and usefulness of temporal segments' statistical classification. Lastly, *Autocorrelation Functions* (ACFs) will be applied to the resulting data for Gaussian occurrence to compare the patterns displayed by the classified segments with those already known by previous studies of the SR.
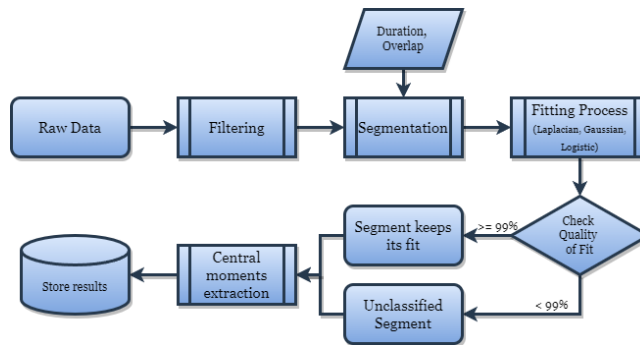

Figure 2: Flow chart of data processing for this work.

3.1 Hypothesis' assessment

The relationship between the median of Gaussian (Normal) segments per hour for the analyzed month (April 2016) and lightning activity uncovered in Soler-Ortiz et al. (2021) gives way to the current work and is explored further using the analyzed data in Fig. 3. Figs. 3a and 3b show the hourly average of Gaussian segments per year on the EW and NS channels, respectively. Fig. 3c presents the hourly average of the five years for both channels with the intensity given to the SR by each thunderstorm center, these values extracted from Nickolaenko and Hayakawa (2014), calculated using the model proposed in Nickolaenko et al. (1998).

The graphs in Fig. 3a show the same pattern, albeit with some differences between their mean values and the values they reach along the different time intervals. The same can be said of Fig. 3b. Repetition of patterns is evidence that the Gaussian occurrence is related to a physical phenomenon that affects the analyzed signal. The differences between the values reached each year reinforce the evidence as an indicator of intensity variations, since the intensity of global lightning activity is not constant throughout the years, although their daily activity patterns are. Now, in Fig. 3c, it can be identified how EW Gaussian occurrence increases with the activity of the African thunderstorm center, whereas the peak during the late hours on the NS channel follows the American thunderstorm center. However, other factors in the relationship are unclear, such as the contribution of the Asian thunderstorm center to any of the channels. In Soler-Ortiz et al. (2021) a two-hour shift was applied to the obtained values to highlight its similarities with the Maritime Continent's thunderstorm center. In light of this analysis's results, there is not enough evidence to state that the Gaussian occurrence increase at 7 UTC is caused by the Maritime Continent's thunderstorm center, so this must be explored further.

First, it must be explained that the sensitivity of the induction coils used as magnetic antennas is distributed as an eight-shaped curve with its long axis perpendicular to the core of the coil (Burrows, 1978). As a consequence, the electromagnetic activity of each storm center affects each channel in different proportions, depending on intensity, source-observer distance, and relative orientation (Nickolaenko et al., 1998). That being said, the equations that describe the horizontal magnetic field at the ELF range are mentioned in the literature (Nickolaenko, 1997) and are reproduced here for the reader's convenience in Eq. 1 and Eq. 2.

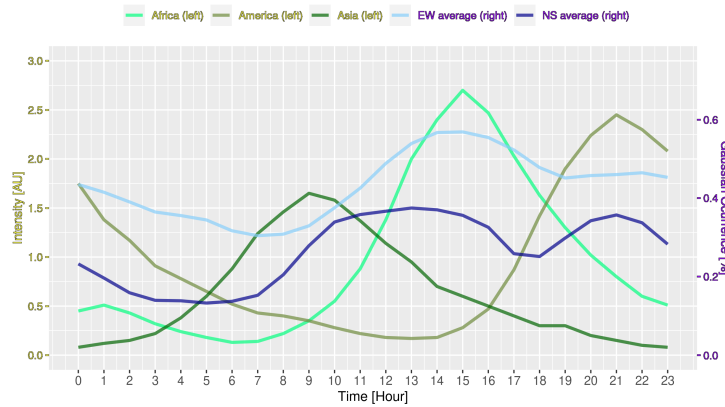$$H_x(t) = H_{NS} = \sum_{k=-\infty}^{\infty} A_k \cdot g(t - t_k) \cdot sin(B_k) \tag{1}$$

$$H_y(t) = H_{EW} = \sum_{k=-\infty}^{\infty} A_k \cdot g(t - t_k) \cdot cos(B_k) \tag{2}$$

With A_k being the amplitude of the kth lightning pulse, $g(t)$ the waveform of the magnetic field components and B_k the kth source bearing. Given the data
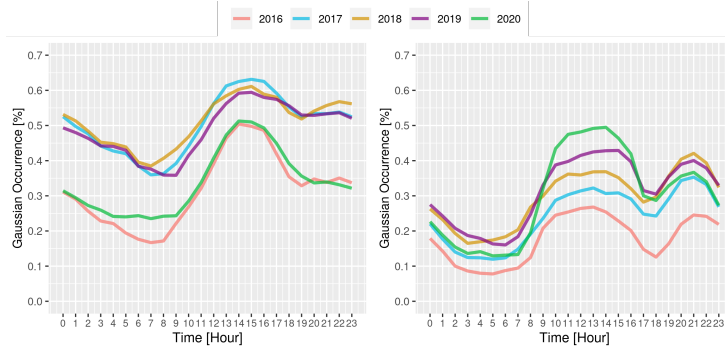
(a) EW channel

(b) NS channel

(c) Hourly average storm center activity (left axis) extracted from Nickolaenko and Hayakawa (2014) and calculated using the model proposed in Nickolaenko et al. (1998) with the hourly average of Gaussian occurrence (right axis) for EW and NS channels.

Figure 3: Overview of hourly averaged fraction of Gaussian classified segments for each year (1 min length), which for the remainder of the article will be called Gaussian occurrence. The similarity between trends of different years and their stability over the years for both channels (Figs. 3a and 3b) clearly indicates the periodicity of Gaussian occurrence. In the comparison between the average occurrence by channel for the entire data set with thunderstorm activity (Fig. 3c) the effects of Africa in the EW channel and of America in NS could be perceived.

9

Table 1: Average geographical coordinates of the thunderstorm centers and resulting coefficients to calculate each channel intensity.

	Coordinates	Azimuth	EW coef.	NS coef.
African	10º S, 28º E	6.14 Mm, 142.44°	0.793	0.61
American	4º S, 66º W	7.98 Mm, 249.98°	0.342	0.94
Asian	2.5º S, 120º E	13.01 Mm, 70.99°	0.326	0.945

depicted in Fig. 3, the above equations imply that the NS and EW channels are affected by the total field intensity proportionally to the sine or cosine of the source bearing, respectively. In Nickolaenko et al. (1998) there is also a model for the monthly location of thunderstorm centers from which the average yearly location of each can be calculated. With them, the azimuth of each location with respect to our observatory can be calculated, giving us the coefficient by which intensity is modified for each channel. The results are shown in Table 1.

The distance to the thunderstorm center is also relevant since the intensity measured by the sensors depends on the distance from the thunderstorm center (Nickolaenko, 1997). The top left graph of Figure 1 of the cited work presents the power loss in the electromagnetic field according to distance. According to Table 1 the distance of the African (6.14 Mm) and Asian (13.01 Mm) thunderstorm centers seem to be at opposite sides of the curve, so theoretically their reception value should be the same. The American thunderstorm center's distance (7.98 Mm) it is higher in the curve but the small resolution of the figure makes us refrain from making any estimates. The calculated coefficients are then applied to the intensity data and presented in Fig. 4. Each sub-figure features thunderstorm center intensity modified to match the level at which is received by each sensor as well as Gaussian occurrence from the records of each channel; (4a) shows the EW channel and 4b displays the data for the NS channel. After this transformation, the similarities between Gaussian occurrence and thunderstorm intensity are further highlighted.

It can be seen in Fig. 4a how the maximum value of Gaussian occurrence matched the maximum value of the intensity of the African thunderstorm center, which supports the main hypothesis, since this storm center contributes fully to the signal. The contribution of the other two is considerably lower, with the American thunderstorm center having a peak intensity halving that of the African's. The way Gaussian Occurrence flattens at the end of the period seems to be due to an interaction between both thunderstorms, but to properly define it, further research is needed. Finally, the real contribution of the Asian thunderstorm center to EW intensity is reduced by 68 % due to its geographical position in relation to the measurement station is almost four times lower than the American contribution, thus explaining how its effect is barely observable in the Gaussian occurrence curve.

On the other hand, Fig. 4b shows American thunderstorm center as its main contributor. This is consistent with the similarities displayed between the peaks of Gaussian occurrence and American thunderstorm intensity. It also
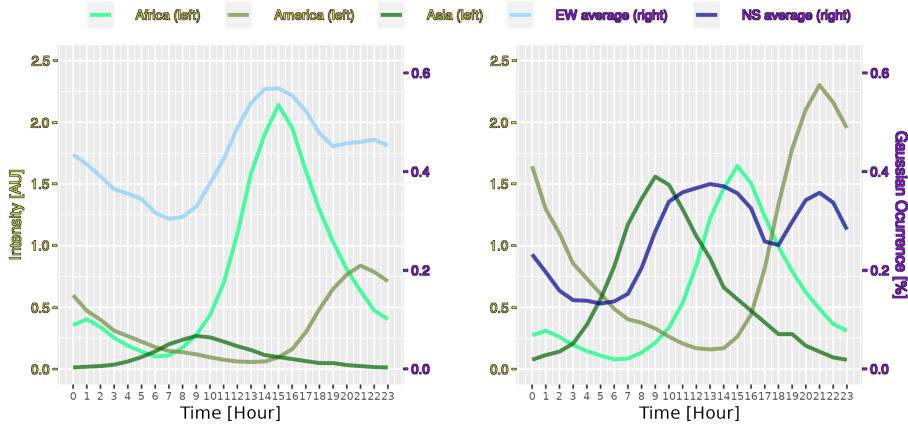
(a) EW channel   (b) NS channel

Figure 4: Thunderstorm intensity (left axis) and Gaussian occurrence (right axis) comparison with location correction. After applying the correction, some of the differences in Gaussian occurrence between the channels are contextualized, such as Gaussian occurrence peaks matching max intensity peaks, the dominant effect of African thunderstorm center in EW channel, and the plateau in NS's Gaussian occurrence corresponding with the similar intensity of Africa and Asia.

shows a reduced contribution from Africa, which has almost the same peak value as the intensity supplied by the Asian thunderstorm center. This justifies the plateau shape displayed by the NS curve from 10 UTC to 16 UTC, which has a one-hour difference between the hours when Asian and African thunderstorm centers reach their peaks (9 UTC and 15 UTC, respectively). However, some inconsistencies should be further explained, such as the Gaussian occurrence of the NS channel having lower average value than EW's despite having a higher intensity contribution overall, as well as the effects of American thunderstorm center in EW channel, or the delay between the rising edge of NS Gaussian occurrence and Asian thunderstorm center's. The answers to further refine these values may be found watching the main wave propagation paths from our observatory, as depicted in Fig. 5.

This map serves to illustrate another variable to be taken in consideration. The average central location of the storm centers has been used to estimate the proportional intensity that reaches each channel, but global lightning activity is never localized in one point; it spans whole continents. That is why the main wave propagation paths help to understand how and when lightning intensity affects each channel. For example, the NS channel's main wave propagation path passing through the Maritime continent is almost tangential, but the thunderstorm center extends as far as India, which is close enough to the
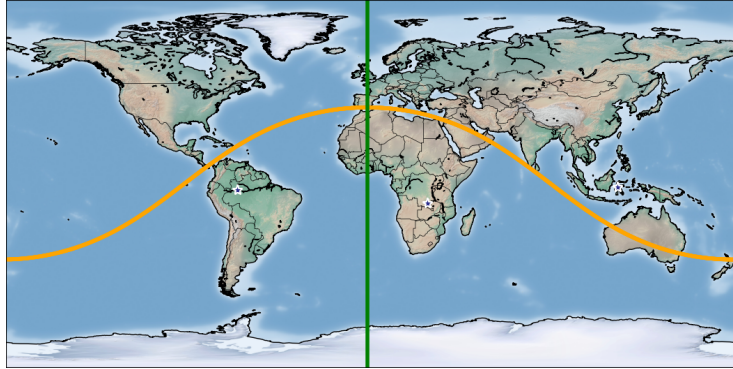
Figure 5: Main wave propagation paths from our observatory for the EW channel (green line) and the NS channel (yellow line) with the average geographic centers of the thunderstorms in each continent marked with a blue-white stars.

path to be quite influential. Considering the evidence discussed so far, the delay of Gaussian occurrence rising edge in our observatory's records from the NS channel in relationship with Asian lightning activity could be caused by the detection of Indian lightning activity, which would happen a couple hours later than the Maritime continent. The purpose of Fig. 5 is to offer the reader some perspective about other factor that influence SR recording, which means that it should be considered to have an effect on Gaussian occurrence in SR segments, considering their prospective relationship with lightning activity.

The evidence shed by this rough approximation is enough to consider SR segments displaying a Gaussian distribution as an indication of lightning activity on the globe. However, more research is needed to formalize the relationship through mathematical models based on real data. This method presents a way to process data in bulk while identifying time segments in SR background noise that are storm-active, or related to thunderstorm activity. This allows researchers to automatically classify their records and to have a general idea of the phenomena that occurred within them.

3.2 Overall Results

Fig. 6 gives an overview of the analysis performed, showing the results using segments of 1 min duration.

Fig. 6a shows the results for the EW channel and Fig. 6b for the NS channel. Both show the daily average of each target distribution's occurrence as the length of the segment of the appropriate color, with each vertical segment being relative to one day. Laplacian distributions are scarce but still present in both channels, appearing more notably during the months ranging from November to April.

On the days when Laplacian segments are present, Logistic segments occurrence is also higher, for example, in the last trimester of 2018. This is especially

12
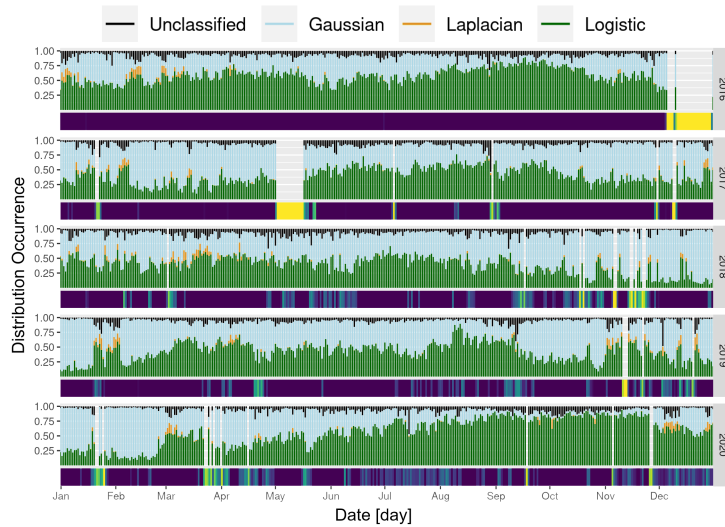
(a) EW channel
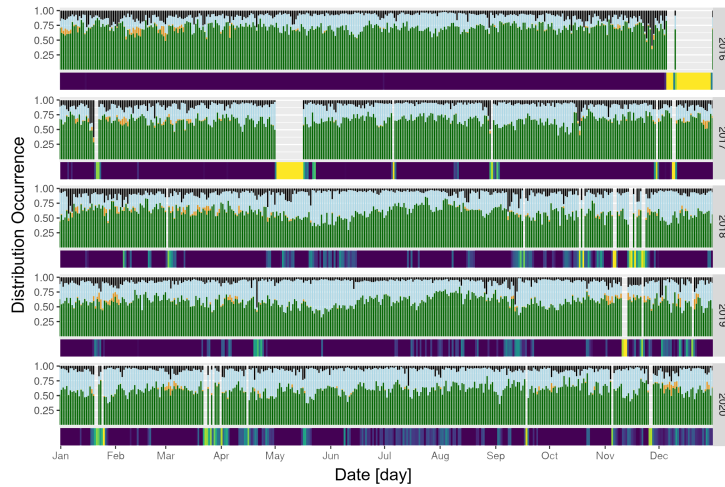
(b) NS channel

Figure 6: Daily average of distribution occurrence for all the analyzed data (1 min segments). Each vertical bar represents a day, with the % of each distribution occurrence filling the bar with its color proportionally. At first glance, it can be seen that the Gaussian occurrence (light blue) is more common in the EW channel (Fig. 6a) than in the NS channel (Fig. 6b). Its seasonal pattern is also more pronounced.

noticeable in Fig. 6a, where they reach values similar to the ones shown in Fig. 6b.

The preliminary conclusion regarding Laplacian segments is that they are mostly related to strong lightning discharges closer to the sensor. This is driven by the fact that the same behavior is displayed on both channels when several Laplacian segments are concentrated on the same day. The sensing capacity of the magnetic coils that serve as sensors overlaps only at a close range. Inspection of Fig. 6 confirms that despite presenting different data, Laplacian segments appear on the same days on both channels and with sizes that imply similar occurrence as well. To further evaluate this hypothesis, the hourly averaged amount of segments for each distribution on each channel has been correlated between them, both by different years and for the whole analysis. The results are in Table 2.

Table 2: Yearly correlation coefficient of comparing distribution occurrence between channels. All p-values below $0.001\,\%$

	2016	2017	2018	2019	2020	All years
Gaussian	0.27	0.24	0.37	0.43	0.38	0.35
Logistic	0.25	0.14	0.23	0.29	0.32	0.27
Laplacian	0.89	0.88	0.89	0.95	0.90	0.90

The high correlation values are notable evidence for Laplacian segments being related to close and/or strong lightning activity.

Regarding the other two distributions, Gaussian segments show a higher correlation value than Logistic segments. The significant correlation values displayed by Gaussian segments are evidence of thunderstorms affecting both sensors in a specific proportion, and shows how this relationship can be extracted from the statistically analyzed records.

Having a target distribution group of three and ascertaining that Laplacian manifests only occasionally, Logistic and Gaussian segments can be considered complementary. This complementarity is further evidence of SR background noise - the base state of the Schumann resonance and hence, the most common - answering to both statistical classifications, being Gaussian on the storm-active periods and Logistic on the steady-state periods. Both Logistic and Gaussian distribution's presence is distributed throughout the whole year, but the number of segments classified as Logistic experiences a higher variation in Fig. 6a than in Fig. 6b. Since the overall intensity of lightning activity captured by the EW sensor is higher than what the NS sensor receives (see Fig. 4), observing a similar behavior in Gaussian occurrence is another sign of its link with lightning activity.

Fig. 7 displays the results of the analysis for April of all the years analyzed for both channels (Fig. 7b for NS and Fig. 7a for EW). In this case, each vertical segment represents an hour. Data integrity is absent, since at this level of aggregation it gives no additional information; the amount of hours for which a data point is produced with a fraction of the data is scarce, given the format

14

of analyzed data.

The asymmetry between both channels is clear here as well; in Fig. 7a the fluctuation between Gaussian and Logistic segments is more noticeable than in Fig. 7b. Also, Gaussian occurrence is way higher in the former than in the latter. In all years, a daily variation can be appreciated. This adds further weight to the hypothesis explored in this article, since thunderstorm centers have, among others, a daily cycle (Ogawa et al., 1969; Prácser et al., 2019). Daily and seasonal trends will be explored in detail later on.

It can be seen again how the presence of segments classified as Laplacian occurs in both channels at once and always comes with an increase of Unclassified segments, although a sudden increase of Unclassified segments with small or nil Laplacian occurrence can be observed as well.

On hourly averaged records, more extreme cases are observed, during which most hours appear as Unclassified (2016-04-20, 2017-04-18, 2018-04-27, and 2020-04-19 are examples of this). The reason is the same as expressed previously; these are mostly Laplacian segments that, by not crossing the quality of fit threshold, fell into this category. Laplacian segments that do not exceed the quality threshold are classified as such for containing isolated among a segment that is mostly background noise. This can be caused by close lightning activity, as was proposed for the rest of the Laplacian segments caused by the so-called ELF flashes (Ogawa et al., 1966), such as the segment shown in Fig. 8, but could also be due to a captured Q-burst in a segment which is mostly background noise. Having the ability to tag Q bursts automatically would be an extremely useful feature for an automated analysis methodology, but so far all attempts to differentiate between these two cases among Unclassified segments have been unsuccessful.

After checking the general results presented in Figs. 6 and 7, it is logical to assume that three out of four categories of this analysis can be considered as a state in the SRs records.

- The steady-state background noise of the resonance, identified with the Logistic distribution.

- Intervals containing at least one high amplitude peaks, which are the segments classified as Laplacian.

- Segments of storm-active SR background noise, meaning that they are linked to global lightning activity, for whose the Gaussian distribution is the best fit.

- Unclassified segments are those who do not belong to either of the three categories, although some of the phenomena ending in this group are recognized. Examples are periodic disturbances, Q-bursts, or ELF flashes.

By taking advantage of the rest of the parameters extracted by the analysis, the Unclassified category could be refined. An example of this is how high pulses can be located by looking for segments with high kurtosis and low statistical
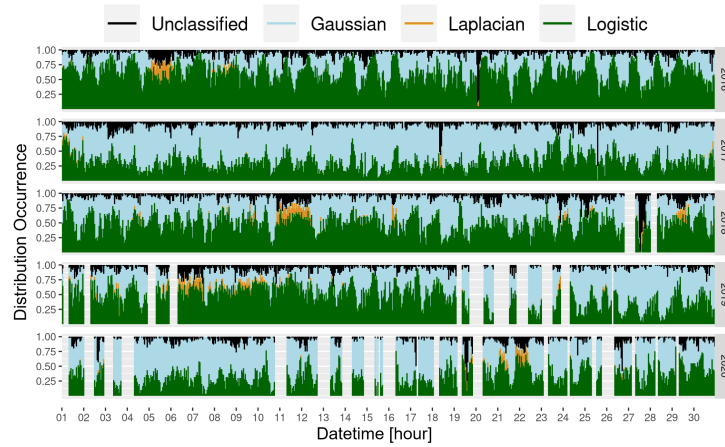
(a) EW channel
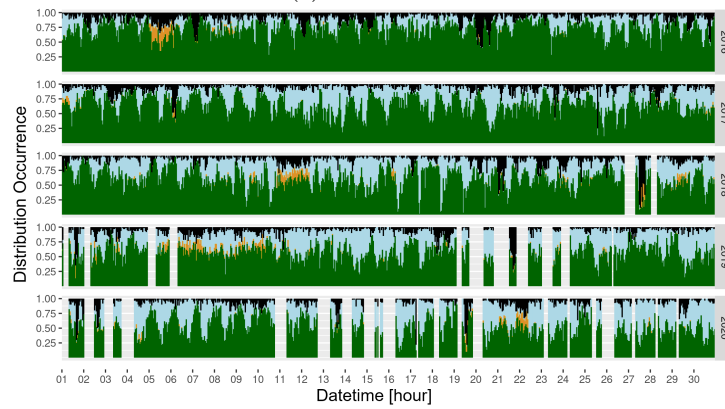
(b) NS channel

Figure 7: Hourly average of distribution occurrence in April of each year (1 min segments). A daily pattern between Gaussian (light blue) and Logistic (green) can be appreciated. It is worth highlighting the Laplacian (orange) concentration in both channels at the same time, such as from 5 to 7 April 2016 and from 11 to 12 April 2018.
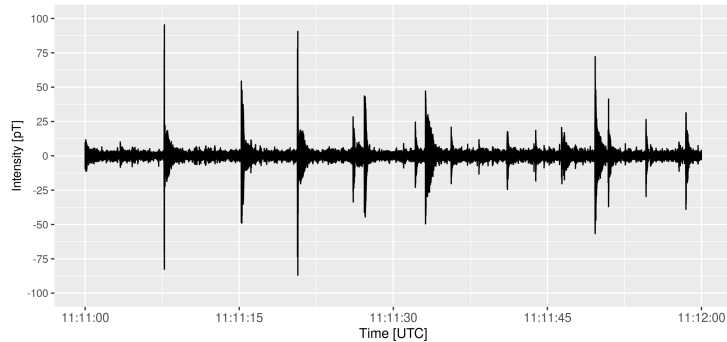
Figure 8: Unclassified segment from 2018-04-27, displaying high amplitude pulses.

deviation. Such a filter was used to get the segment displayed in Fig. 8. The same can be said about the rest of categories; in the same way that the best fit distribution describes a segment, its statistical parameters may serve to narrow down specific events under a certain category.

This and other analyses should be carried on to reach a definitive conclusion; its interest justified by the presented evidence of the chosen statistical model for a given segment being a way to identify the signal's state.

3.3   Gaussian occurrence analysis

Understanding the effects of segment length in classification results is a way to better understand the relationship between the fraction of segments classified as Gaussian and lightning activity. To do so, the results of two analysis with different segment lengths - 10 min segments with overlap of 9 min and 1 min segments without overlap - will be compared, plotting the results together along with the most significant trends, extracted from the raw data using *Moving Average* (MA) functions.

First, Fig. 9 presents the daily fraction of Gaussian segments captured by the NS channel. Some similarities and differences between the results can be highlighted under different segment lengths.

- 1 min segments' average value is higher than in the 10 min ones. It seems that the former rarely reaches 0, whereas for the latter it is a common value.

- Despite the difference between average values, the seasonal trends of both analyzes show the same trend. However, excursions in the 1 min trend are more pronounced than in the 10 min trend. This is the reason why on the former more peaks and troughs can be observed than on the latter.

- Regarding the monthly trend, the peaks in the 10 min trend have always a counterpart in the 1 min one. It is not so clear if the relationship works
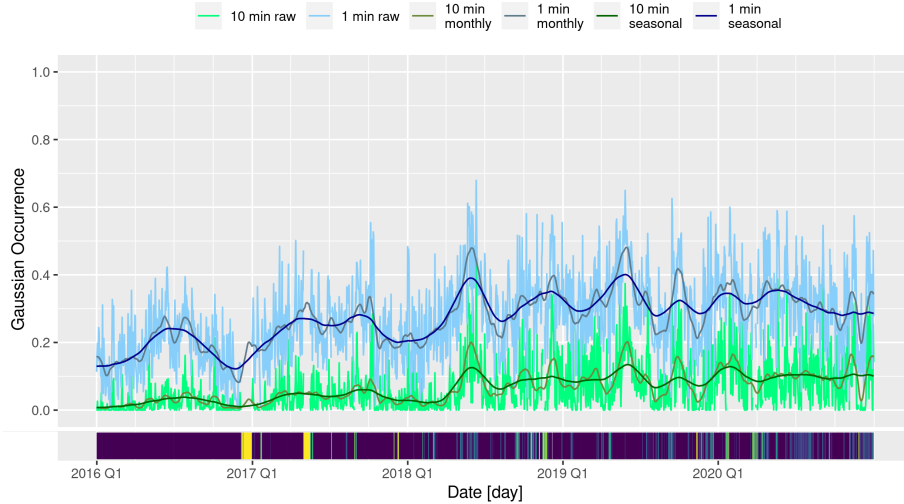
17

Figure 9: Comparison between classified data from channel NS with 1 min and 10 min segments, displaying monthly and seasonal trends. It is difficult to visually identify any trends, the comparison between different analyses show a higher average value for 1 min segments. Despite that, the trends of both curves are similar.

in the opposite direction, especially in sections where the average value of the 10 min trend is too low, such as in 2016 and 2017.

Next, Fig. 10 presents the data from the EW channel in the same format used for Fig. 9. The same observations highlighted for the NS data hold in this case, but there are also two additional aspects to consider.

- First, the maximum values are higher than on the NS channel, reaching values over 0.95 around New Years Eve in 2018 and peaking over 0.8 on occasion.

- On the EW channel, the 10 min segment, the raw peak value overlaps with the 1 min peak, sometimes even going over it. However, its average value is still lower; 10 min reaching the same value as 1 min occurs only for high occurrence values.

After reviewing the results of both channels, one of the main differences between the two lengths analyzed was their average value and their peak value. The former seems to be directly related to the analysis features. Generally speaking, trend changes are more noticeable in the 1 min segments than in 10 min.

The ELF signature of a lightning event can have a duration of a whole second. Under the evidence obtained so far, which links Gaussian segments with global
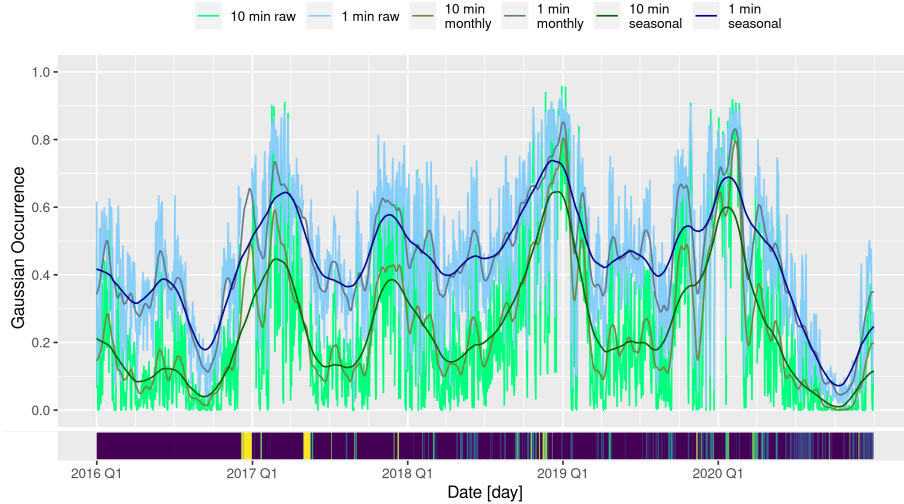
18

Figure 10: Comparison between classified data from channel EW with 1 min and 10 min segments, displaying monthly and seasonal trends. The yearly pattern of the African Thunderstorm center appears clearly in both analyses, due to its major influence in this channel. Some instances of 10 min segments can be seen overcoming the value of 1 min segments.

lightning activity, it is reasonable to assume that the presence of a certain number of lightning events is what modifies the SR background noise from steady-state to storm-active. Considering that enough lightning events over a time period will make it statistically identifiable with the Gaussian distribution, it is clear that longer segments will need more events to be identified as Gaussian. The only situation in which a Gaussian occurrence in 10 min can overcome the value displayed in the 1 min segments is when the number of lightning events is enough to affect a considerable number of 10 min segments. Consequently, 1 min segments display a higher selectivity than 10 min, with the latter reacting strongly on what, under the hypothesis of this work, must be considered periods of high lightning intensity. Evidence of this can be seen in Fig. 11; under the hourly average of the results. It can be seen that the peak values of the 10 min analysis occur invariably at noon, when the African storm center peaks in activity. It can be appreciated as well how in the 1 min analysis Gaussian occurrence rarely reaches 0.0, whereas in the 10 min analysis it is a common value.

On the other hand, during the high-occurrence hours, the number of Gaussian segments is higher for the 10 min than for the 1 min. This consequence comes from the 9 min overlap applied to the 10 min segments since a specific event will appear in 10 different segments, while in the 1 min length segments (no overlap), each event affects only one segment. In conclusion, even when the particular impact of the analysis' parameters is not quantified, it is clear
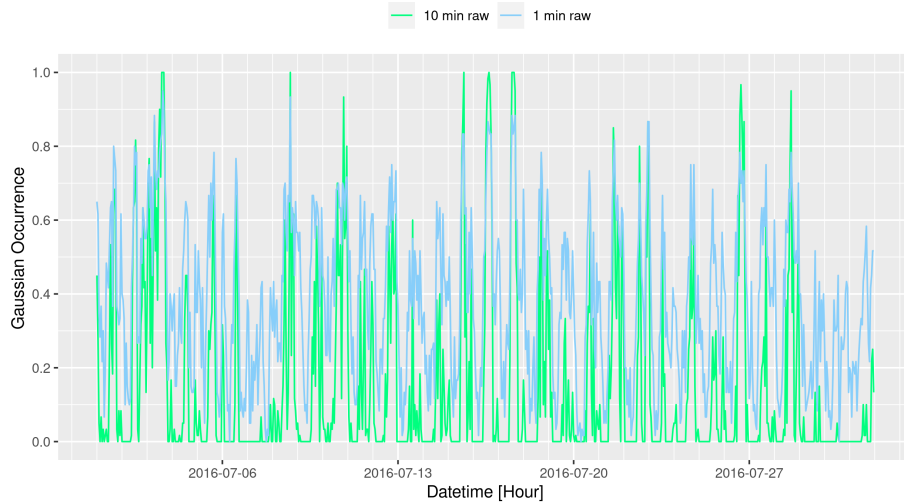
19

Figure 11: Hourly averaged fraction of Gaussian segments for July 2016, EW channel, with 1 min and 10 min segments. The daily pattern emerges clearly, with 10 min taking more extreme values but 1 min analysis displaying a more balanced behavior.

that by controlling segment length and overlap, the researcher can modify the selectivity of the method.

With respect to the significance of the peak values, it can be seen how the occurrence value in the NS channel is never very high, and its distribution throughout the temporal records is sparse enough to not produce higher peak values on the 10 min records. The EW channel, in turn, shows much higher values for 1 min segment values and the fact that 10 min segments peaks go over them implies that on the days of high events occurrence the events are cluttered, causing a higher number of events identification.

This behavior coincides once again with the recorded lightning intensity of the storm centers after considering the source bearing with respect to the observatory.

3.4   Pattern analysis

To further analyze the daily trends displayed by Gaussian occurrence, an analysis using *Autocorrelation Function*s (ACFs) with a window of 24 hours has been applied to the fractions of Gaussian occurrence from the EW channel, choosing four months for all the years analyzed. The ACF function applied to the data is of discrete form, as shown in 3.

$$R_{yy}(\ell) = \sum_{n \in Z} y(n)\overline{y(n-\ell)} \tag{3}$$

20

The results are presented in Fig. 12.

The ACF results display the existence of a daily trend in every case, although there are some deviations and the possible reasons are worth discussing. First, the results shown in Fig. 12a clearly indicate an underlying trend present in addition to the daily one. However, the daily cycle can still be appreciated in the results for the years 2018 and 2020. Fig. 12b is still affected by a different trend, but the daily oscillations of the autocorrelation coefficient that hint at a daily trend can be identified, even in the years 2017 and 2019 where the differences in average value between periods clearly hint at the presence of another trend. The presence of a daily trend can be confirmed for all years in Fig. 12c, with only 2019 deviating slightly from the general pattern, while in Fig. 12d the daily trend for November 2019 is greatly distorted and in 2018 is barely recognizable.

The explanation of these differences could be attributed once again to lower frequency patterns present in the signal, but by comparing the data from different years it can be argued how the season to which the data belong should be accounted for. Fig. 12c shows the most visible trends in July. This is the month when African thunderstorm activity is at its lowest on average (Albrecht et al., 2016). This would imply that minimal activity also translates into minimal variability, so even when lightning activity is at its lowest (see Fig. 10) its activity is regularly synchronized with the sun cycle, and the ACF shows regular peaks with high autocorrelation values. Arguing along the same lines for April (Fig. 12b) and November (Fig. 12d), these months are theoretically on the falling and rising edges of activity, respectively. By following the previous train of thought, it can be considered that the months which are closer to the period of minimal thunderstorm activity will show this daily trend clearly than the months who are further away.

As with the original SR records, the number of nuances and differences can be overwhelming, and more relationships could be found with deeper studies.

4   Conclusions

For this study, a recently developed methodology has been applied to 5 years of data. Through statistical analysis, expert knowledge of the storm centers' seasonal trends, and of the Schumann Resonance, the latter signal has been segmented and categorized using statistical models. In this article, we consider the implications behind the classification of a specific segment in each model. Through this exploration, the utility of the procedure has been proven. After testing the connections between each of the statistical distributions and the different states of lightning activity, there should be interest in exploring what other information can be extracted from this modeling strategy. Throughout the exploration of the relationship between Gaussian occurrence and lightning activity, this methodology shows promise of being one of the most efficient ways to assess the state of lightning activity in an automated way, to the best of our knowledge.

The similarities between the occurrence of the distribution over a period

(a) January   (b) April
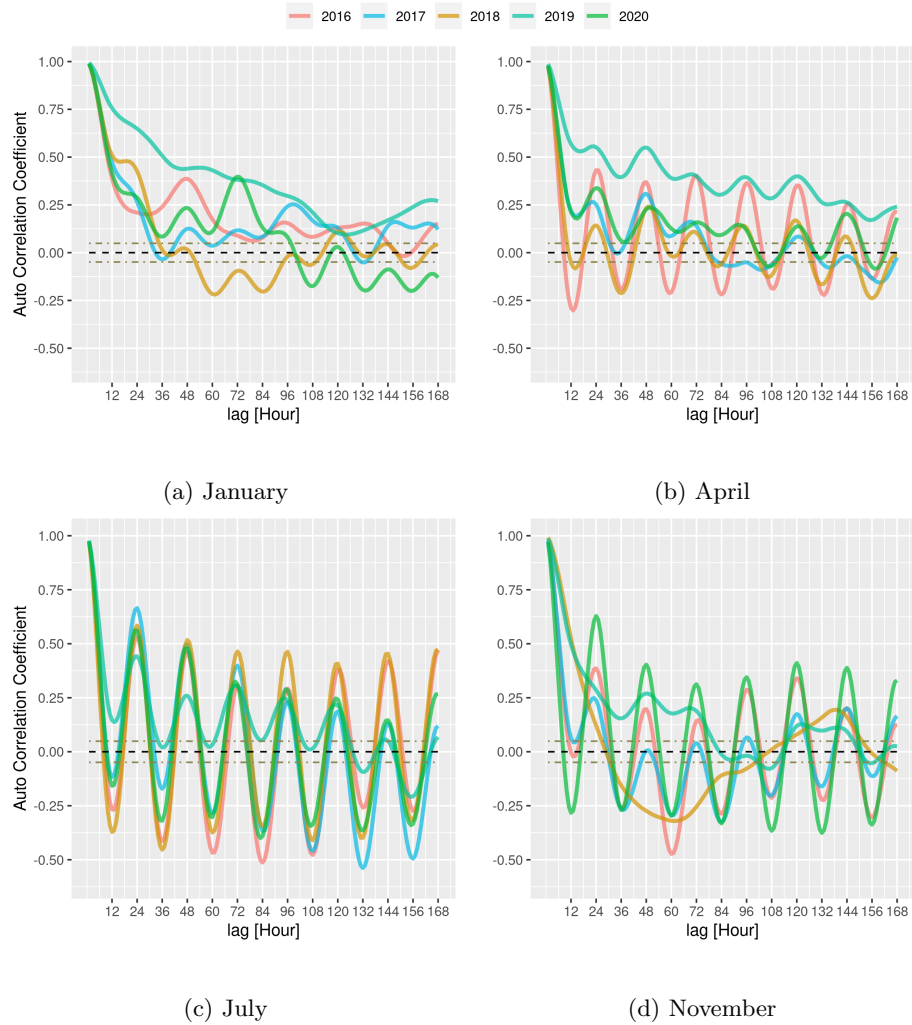
(c) July   (d) November

Figure 12: ACFs results for the hourly averaged data of Gaussian occurrence on different months on all the analyzed years, with 7 days of maximum lag. The daily pattern is observable in most years for all the chosen months, although other underlying trends may be perceived. The daily patterns are more homogeneous throughout the years in the month of July (Fig. 12c) where lightning activity is at its lowest.

of time and the different trends known for the signal under study point to the validity of these relationships and the main argument to demonstrate the utility of the methodology. Specifically, Laplacian segments have been confirmed to be related to strong and/or close lightning activity, and the relationship between Gaussian segments and global lightning activity is clearly shown. The behavior of monthly and daily trends in the fraction of Gaussian segments per hour has also been studied, which settles them as an alternative way to study lightning in the ELF range. However, more research is needed to compare the distribution occurrence with geotagged lightning activity records to fully validate these hypotheses and further explore the quantitative information that can be extracted from the analysis results. Just as well, the segments deemed Unclassified for this study reveal the necessity of refining the system in that direction, looking for ways to extract valuable data from those segments that do not get past the demanded quality of fit.

This analysis proves to be a complement to commonly used frequency domain information when searching for correlations and studying the different aspects of the Schumann resonance. The categorization provided by the analysis is already a great advantage, allowing researchers to automatically process a huge number of registers and get a reasonable estimate of the events that occur during it. Inversely, knowing the parameters of the distribution displayed by a particular event, an automatic search could be performed on the analyzed data to confirm if the distribution parameters represent the event.

Acknowledgments

References

A. Nickolaenko, P. Colin, D. I. (2000). Hurst exponent derived from natural terrestrial radio noise in Schumann resonance band. *Geophysical Research Letters*, *27*(19), 3185–3188. 10.1029/1999GL000015.

Afrifa-Yamoah, E., Mueller, U. A., Taylor, S. et al. (2020). Missing data imputation of high-resolution temporal climate time series data. *Meteorological Applications*, *27*(1), e1873. 10.1002/met.1873.

Alata, O., Olivier, C., and Pousset, Y. (2013). Law recognitions by information criteria for the statistical modeling of small scale fading of the radio mobile channel. *Signal processing*, *93*(5), 1064–1078. 10.1016/j.sigpro.2012.11.016.

Albrecht, R. I., Goodman, S. J., Buechler, D. E. et al. (2016). Where are the lightning hotspots on earth? *Bulletin of the American Meteorological Society*, *97*(11), 2051–2068. 10.1175/BAMS-D-14-00193.1.

Beggan, C. D., and Musur, M. A. (2019). Is the madden–julian oscillation reliably detectable in schumann resonances? *Journal of Atmospheric and Solar-Terrestrial Physics*, *190*, 108–116. 10.1016/j.jastp.2019.05.009.

Betrie, G. D., Sadiq, R., Tesfamariam, S. et al. (2016). On the issue of incomplete and missing water-quality data in mine site databases: Comparing three imputation methods. *Mine Water and the Environment*, *35*(1), 3–9. 10.1007/s10230-014-0322-4.

Bór, J., Bozóki, T., Sátori, G. et al. (2023). Responses of the ac/dc global electric circuit to volcanic electrical activity in the hunga tonga-hunga ha'apai eruption on 15 january 2022. *Journal of Geophysical Research: Atmospheres*, (p. e2022JD038238). 10.1029/2022JD038238.

Bozóki, T., Sátori, G., Williams, E. et al. (2023). Day-to-day quantification of changes in global lightning activity based on schumann resonances. *Journal of Geophysical Research: Atmospheres*, (p. e2023JD038557). 10.1029/2023JD038557.

Burrows, M. L. (1978). *ELF communications antennas*. 5. P. Peregrinus Stevenage.

Cano-Domingo, C., Castellano, N. N., Fernandez-Ros, M. et al. (2022a). Segmentation and characteristic extraction for schumann resonance transient events. *Measurement*, *194*, 110957. 10.1016/j.measurement.2022.110957.

Cano-Domingo, C., Stoean, R., Joya, G. et al. (2023). A machine learning hourly analysis on the relation the ionosphere and schumann resonance frequency. *Measurement*, *208*, 112426. 10.1016/j.measurement.2022.112426.

Cano-Domingo, C., Stoean, R., Novas-Castellano, N. et al. (2022b). On the prospective use of deep learning systems for earthquake forecasting over schumann resonances signals. *Engineering Proceedings*, *18*(1), 15. 10.3390/engproc2022018015.

Chrissan, D., and Fraser-Smith, A. (2000). A comparison of low-frequency radio noise amplitude probability distribution models. *Radio Science*, *35*(1), 195–208. 10.1029/1999RS900085.

Cook, S. J., and Weidmann, N. B. (2019). Lost in aggregation: Improving event analysis with report-level data. *American Journal of Political Science*, *63*(1), 250–264. 10.1111/ajps.12398.

Domingo, C. C., Ros, M. F., Castellano, N. N. et al. (2021). Diurnal and seasonal results of the schumann resonance observatory in sierra de filabres, spain. *IEEE Transactions on Antennas and Propagation*, *69*(10), 6680–6690. 10.1109/TAP.2021.3069537.

Dyrda, M., Kulak, A., Mlynarczyk, J. et al. (2014). Application of the schumann resonance spectral decomposition in characterizing the main african thunderstorm center. *Journal of Geophysical Research: Atmospheres*, *119*(23), 13–338. 10.1002/2014JD022613.

Evans, J., and Griffiths, A. (1974). Design of a sanguine noise processor based upon world-wide extremely low frequency (elf) recordings. *IEEE transactions on communications*, *22*(4). 10.1109/TCOM.1974.1092215.

Figueredo, P. S., Ortega, B. M., Pazos, M. et al. (2021). Schumann resonance anomalies possibly associated with large earthquakes in mexico. *Indian Journal of Physics*, *95*, 1959–1966. 10.1007/s12648-020-01865-6.

Gazquez, J. A., Garcia, R. M., Castellano, N. N. et al. (2017). Applied engineering using schumann resonance for earthquakes monitoring. *Applied Sciences*, *7*(11), 1113. 10.3390/app7111113.

Gowanlock, M., Li, J. D., Rude, C. M. et al. (2018). Parallel optimization of signal detection in active magnetospheric signal injection experiments. *Computers & Geosciences*, *114*, 107–116. 10.1016/j.cageo.2018.01.020.

Hayakawa, M., Ohta, K., Nickolaenko, A. P. et al. (2005). Anomalous effect in schumann resonance phenomena observed in japan, possibly associated with the chi-chi earthquake in taiwan. In *Annales Geophysicae* (pp. 1335–1346). Copernicus GmbH volume 23. 10.5194/angeo-23-1335-2005.

Heckman, S., Williams, E., and Boldi, B. (1998). Total global lightning inferred from schumann resonance measurements. *Journal of Geophysical Research: Atmospheres*, *103*(D24), 31775–31779. 10.1029/98JD02648.

Khayati, M., Arous, I., Tymchenko, Z. et al. (2020). Orbits: online recovery of missing values in multiple time series streams. *Proceedings of the VLDB Endowment*, *14*(3), 294–306. 10.14778/3430915.3430920.

Kudintseva, I., Galuk, Y. P., Nickolaenko, A. et al. (2018). Modifications of middle atmosphere conductivity during sudden ionospheric disturbances deduced from changes of schumann resonance peak frequencies. *Radio Science*, *53*(5), 670–682. 10.1029/2018RS006554.

Kudintseva, I., Nikolayenko, S., Nickolaenko, A. et al. (2017). Synthesis of schumann resonance background signal in time domain. *International Journal of Electronics and Applied Research (IJEAR)*, *4*(1), 1–23.

Lamey, R., Jha, S. N., Adouani, I. et al. (2021). Optimization of hf receiving antenna with horizontal polarization based on itu noise model. In *2021 IEEE 19th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM)* (pp. 1–2). IEEE. 10.1109/ANTEM51107.2021.9519049.

Nickolaenko, A. (1997). Modern aspects of schumann resonance studies. *Journal of Atmospheric and Solar-Terrestrial Physics*, *59*(7), 805–816. 10.1016/S1364-6826(96)00059-4.

Nickolaenko, A., and Hayakawa, M. (2014). *Schumann resonance for tyros: Essentials of global electromagnetic resonance in the earth-ionosphere cavity*. 10.1007/978-4-431-54358-9.

Nickolaenko, A., Kudintseva, I., Pechony, O. et al. (2012). The effect of a gamma ray flare on schumann resonances. In *Annales Geophysicae* (pp. 1321–1329). Copernicus GmbH volume 30. 10.5194/angeo-30-1321-2012.

Nickolaenko, A., Sátori, G., Zieger, B. et al. (1998). Parameters of global thunderstorm activity deduced from the long-term schumann resonance records. *Journal of Atmospheric and Solar-Terrestrial Physics*, *60*(3), 387–399. 10.1016/S1364-6826(97)00121-1.

Nickolaenko, A., Schekotov, A. Y., Hayakawa, M. et al. (2022). Electromagnetic manifestations of tonga eruption in schumann resonance band. *Available at SSRN 4051361*, . 10.1016/j.jastp.2022.105897.

Nickolaenko, A. P., Galuk, Y. P., and Hayakawa, M. (2016). Vertical profile of atmospheric conductivity that matches Schumann resonance observations. *SpringerPlus*, *5*(1), 1–12. 10.1186/s40064-016-1742-3.

Nieckarz, Z., Zikeba, S., Kułak, A. et al. (2009). Study of the periodicities of lightning activity in three main thunderstorm centers based on schumann resonance measurements. *Monthly weather review*, *137*(12), 4401–4409. 10.1175/2009MWR2920.1.

Ogawa, T., and Komatsu, M. (2009). Q-bursts from various distances on the earth. *Atmospheric Research*, *91*(2-4), 538–545. 10.1016/j.atmosres.2008.04.013.

Ogawa, T., and Komatsu, M. (2010). Propagation velocity of vlf em waves from lightning discharges producing q-bursts observed in the range 10–15 mm. *Atmospheric research*, *95*(1), 101–107. 10.1016/j.atmosres.2009.08.015.

Ogawa, T., Tanaka, Y., Miura, T. et al. (1966). Observations of natural elf and vlf electromagnetic noises by using ball antennas. *Journal of geomagnetism and geoelectricity*, *18*(4), 443–454. 10.5636/jgg.18.443.

Ogawa, T., Tanaka, Y., and Yasuhara, M. (1969). Schumann Resonances and Worldwide Thunderstorm Activity: —Diurnal Variations of the Resonant Power of Natural Noises in the Earth-Ionosphere Cavity—. *Journal of geomagnetism and geoelectricity*, *21*(1), 447–452. 10.5636/jgg.21.447.

Ouyang, X.-Y., Xiao, Z., Hao, Y.-Q. et al. (2015). Variability of schumann resonance parameters observed at low latitude stations in china. *Advances in Space Research*, *56*(7), 1389–1399. 10.1016/j.asr.2015.07.006.

Parra, J. A., Ros, M. F., Castellano, N. N. et al. (2015). Techniques for Schumann Resonance Measurements: A Comparison of Four Amplifiers with a Noise Floor Estimate. *IEEE Transactions on Instrumentation and Measurement*, *64*(10), 2759–2768. 10.1109/TIM.2015.2420376.

Pechony, O., and Price, C. (2004). Schumann resonance parameters calculated with a partially uniform knee model on earth, venus, mars, and titan. *Radio science*, *39*(5), 1–10. 10.1029/2004RS003056.

Perotoni, M. B. (2018). Eigenmode prediction of the schumann resonances. *IEEE Antennas and Wireless Propagation Letters*, *17*(6), 942–945. 10.1109/LAWP.2018.2825398.

Prácser, E., Bozóki, T., Sátori, G. et al. (2019). Reconstruction of global lightning activity based on schumann resonance measurements: Model description and synthetic tests. *Radio Science*, *54*(3), 254–267. 10.1029/2018RS006772.

Price, C. (2016). ELF electromagnetic waves from lightning: The schumann resonances. *Atmosphere*, *7*(9). 10.3390/atmos7090116.

Ralchenko, M., Svilans, M., Samson, C. et al. (2015). Finite-difference time-domain modelling of through-the-earth radio signal propagation. *Computers & Geosciences*, *85*, 184–195. https://doi.org/10.1016/j.cageo.2015.09.018.

Seiler, C., and Heumann, C. (2013). Microdata imputations and macrodata implications: Evidence from the ifo business survey. *Economic Modelling*, *35*, 722–733. 10.1016/j.econmod.2013.08.032.

Shvets, A., Nickolaenko, A., and Chebrov, V. (2017). Effect of solar flares on the schumann-resonance frequences. *Radiophysics and Quantum Electronics*, *60*(3), 186–199. 10.1007/s11141-017-9789-8.

Soler-Ortiz, M., Ros, M. F., Castellano, N. N. et al. (2021). A new way of analyzing the schumann resonances: A statistical approach. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–11. 10.1109/TIM.2021.3073435.

Surkov, V., Nosikova, N., Plyasov, A. et al. (2013). Penetration of schumann resonances into the upper ionosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, *97*, 65–74. 10.1016/j.jastp.2013.02.015.

Tatsis, G., Sakkas, A., Christofilakis, V. et al. (2021). Correlation of local lightning activity with extra low frequency detector for schumann resonance measurements. *Science of The Total Environment*, *787*, 147671. 10.1016/j.scitotenv.2021.147671.

Tulunay, Y., Altuntas, E., Tulunay, E. et al. (2008). A case study on the elf characterization of the earth–ionosphere cavity: Forecasting the schumann resonance intensities. *Journal of atmospheric and solar-terrestrial physics*, *70*(2-4), 669–674. 10.1016/j.jastp.2007.08.042.

Tulunay, Y., Tulunay, E., and Senalp, E. (2004). The neural network technique—-1: a general exposition. *Advances in Space research*, *33*(6), 983–987. 10.1016/j.asr.2003.06.008.

Volland, H. (1995). *Handbook of atmospheric electrodynamics* volume 2. CRC Press.

Wagenmakers, E.-J., and Farrell, S. (2004). Aic model selection using akaike weights. *Psychonomic bulletin & review*, *11*(1). 10.3758/bf03206482.

Williams, E., Bozóki, T., Sátori, G. et al. (2021). Evolution of global lightning in the transition from cold to warm phase preceding two super el niño events. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2020JD033526. 10.1029/2020JD033526.

Williams, E., and Mareev, E. (2014). Recent progress on the global electrical circuit. *Atmospheric Research*, *135*, 208–227. 10.1016/j.atmosres.2013.05.015.

Williams, E. R. (1992). The schumann resonance: A global tropical thermometer. *Science*, *256*(5060), 1184–1187. 10.1126/science.256.5060.1184.

Xu, M., Qie, X., Pang, W. et al. (2022). Lightning climatology across the chinese continent from 2010 to 2020. *Atmospheric Research*, (p. 106251). 10.1016/j.atmosres.2022.106251.

Yang, H., and Pasko, V. P. (2007). Power variations of schumann resonances related to el nino and la nina phenomena. *Geophysical research letters*, *34*(11). 10.1029/2007GL030092.

Zefreh, M. M., and Torok, A. (2018). Single loop detector data validation and imputation of missing data. *Measurement*, *116*, 193–198. 10.1016/j.measurement.2017.10.066.

Zhang, Y., and Thorburn, P. J. (2021). A dual-head attention model for time series data imputation. *Computers and Electronics in Agriculture*, *189*, 106377. 10.1016/j.compag.2021.106377.