A new way of analyzing the Schumann Resonances: a statistical approach

Manuel Soler-Ortiz   Manuel Fernández Ros

Nuria Novas Castellano

Jose A. Gázquez Parra, *Senior Member, IEEE*

January 2021

Abstract

Seven decades have passed since the Schumann Resonances (SRs) were identified. Since then, their research interest has increased, currently being a topic of significance. Nonetheless, the papers that study their nature from the frequency perspective are in clear majority regarding those who focus on the time domain. To fill this gap in the literature and further characterize the SRs, a method to perform statistical analysis on the SRs signal in the time domain has been developed. For any given segment of data, the analysis performs a Maximum Likelihood Estimation (MLE) of the statistical parameters from a group of previously selected distributions. After that, the best fit among the target distributions is selected through the application of Akaike Information Criterion (AIC). The method is tested by analyzing a month's worth of data, showing the general analysis' results and discussing the relationship between the chosen target distributions and common aspects between the time segments fitted to them. The reliability of the method's results is also discussed by looking at different aspects of the analysis. Special emphasis is put over the results being correlated with lightning activity. This correlation highlights the usefulness of the method, given the well established relationship between lightning and SRs.

Keywords Schumann Resonances, ELF spectrum, Statistical analysis, Akaike Information Criterion.

1   Introduction

In 1952, Winfried Otto Schumann mathematically predicted the resonant phenomenon in the earth-ionosphere cavity [1], and its appearance on the Extreme Low Frequency (ELF) band of the electromagnetic spectrum. Since then, the topic has gained interest over time and the Schumann Resonances (SRs) have been studied to understand their behavior and role on Earth's physical processes.

At the present time, Earth's lightning activity is accepted as the main source of SRs [2], with the inverse problem of extracting total lightning activity from its records still under research [3]. Nonetheless, other natural phenomena contribute and/or affect the signal, like solar activity [4, 5] and gamma rays [6, 7]. There are studies linking SRs with global temperature [8] and more recently with earthquakes, a topic of interest since SRs monitoring might help in their early detection [9, 10].

In the course of the last decade, researchers have shifted efforts towards developing computer simulations of the phenomenon [11]. The most extended method is using finite-difference time-domain techniques to solve SRs equations in the time domain, [12, 13] but there are other methods, like the one described in [14] where random values extracted from specific distributions are supplied to the equations that describe SRs in the time domain. Besides in simulation oriented works, literature on the topic of SRs time domain analysis is scarce, with only a few researchers taking this approach [15]. This work was developed to contribute into filling this gap.

A prevalent way to analyze time domain signals is through statistical methods. Due to Shannon's information theory [16], these are commonly used on communications [17, 18] but are also applied on other fields related to signal processing such as audio studies [19], or even epidemiology [20].

Even when some works refer to the SRs statistically [15, 21], the topic has not been addressed directly to the best of our recollection. It is common to assume that natural processes are typically Gaussian [22, 23] but this is not the case in phenomena belonging to complex systems [24]. Consequently, there is interest in providing a statistical description of SRs signal in the time domain in order to refine its understanding.

2   Purpose and analysis features

The goal of this article is to ascertain the statistical nature of the time domain aspect of the electromagnetic signals belonging to the 1 to 100Hz band, with a special focus on SRs. Given the many factors exerting influence over the medium and the signal, the nature of the measurand changes over time, causing measurements to have low reproducibility. Consequently, different aspects should be taken into account while analyzing the registers, such as the duration of analyzed data and the predominant signal on a given time.

To pursue the main goal while addressing the issues mentioned above, our

analysis splits the analyzed data into segments. Then, among a set of previously chosen distributions, it finds the one that fits each segment best. The measurement system and data format will be described below, and then the analysis method will be explained in detail.
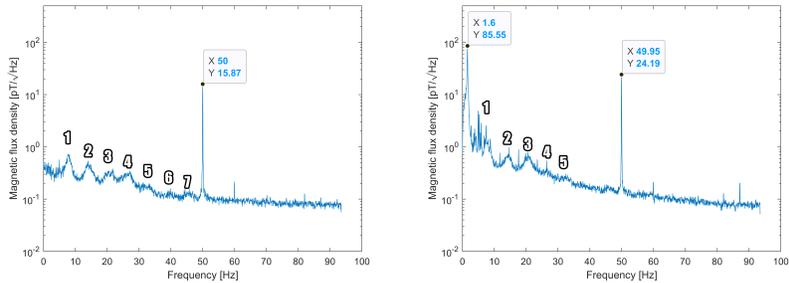
(a) Capture from 15-Apr-2016 02:23:05 to 02:33:05 with no unexpected significant disturbances. The data point shows the value of the European power grid signal.

(b) Capture from 15-Apr-2016 11:42:51 to 11:52:51 with low frequency noise and the European power grid signal, both peaks characterized with datapoints.

Figure 1: Two 10 minute captures in the frequency domain, with the SRs resonance modes numbered.

2.1 ELF station and data files

The data used in this study comes from our ELF station in Sierra de los Filabres, its closest landmark being Calar Alto astronomical observatory (Lat 37.226, Long -2.546), Almería, Spain. Its remote location minimizes interference from man-made signals. The main sensing equipment installed in the station is described and characterized in [25]. It is worth highlighting the presence of two sensing channels fed by high inductance coils acting as sensors, one oriented North - South (NS) and the other East - West (EW).

The resulting measurements are relayed via radio link to a server in the University of Almería, where data from each channel is stored separately in 30 minute files, with a sampling frequency of 187Hz (336600 samples per file) and a bandwidth ranging from 1 to 100Hz. Before this analysis, the data was calibrated to remove any non-linearity caused by the sensors.

2.2 General aspects of the analysis

There are other signals in the 1 to 100Hz band that mask the SRs. Particularly, the 50Hz signal from Europe power grid and the 60Hz signal from America and Asia are always present. Other common disturbance appears on the low frequency range of the sensors' bandwidth, caused by the wind. The effect of these unwanted signals in the SRs' spectra can be appreciated between Fig. 1a and 1b. The power grid signals are present in both, and even if low frequency

3

interference only appears in certain segments, its effect might extend to interfere with the first mode of the SRs, as 1b.

It follows that to properly determine the characteristics of the SRs, these unwanted signals and artefacts should be filtered from the data. This is even more important to ascertain the statistical nature of the signal.

In the frequency spectra, the resonant modes can be observed regardless of the power of these unwanted signals. In Fig. 1a the first five modes are clearly observed, with the sixth and the seventh being also noticeable. Even in 1b with its low frequency interference, the first mode is perceivable. Nonetheless, when presenting time series as histograms, the obtained distribution is related with the signal component of highest intensity. This can be seen in Fig. 2, which shows the histogram related to the spectra depicted in Fig. 1.

Fig. 2a presents the histogram of the signal which spectra is shown in Fig. 1a, where the 50Hz signal of the European power grid is dominant. The predominance of a sine wave is marked by the appearance of two maxima in the histogram, implying an average peak-to-peak value of 10pT. In Fig. 2b the histogram is clearly bell shaped. It is by means of the signal spectrum (Fig. 1b) that the histogram can be associated with low frequency disturbances caused by the wind. The spectrum also shows how low frequency is over three times more powerful than the power grid signal, thus effectively masking it.

This representation serves to illustrate the importance of filtering in this work. After filtering a signal its temporal response changes, and so does its statistical distribution. To properly characterize the signal measured by our sensors before it can be analyzed, different filters will be applied to the raw data, and a whole analysis will be performed for each filtered signal. The applied filters are:

1. None. To characterize the signal as it was captured, Raw unfiltered data is analyzed first.

2. A 60th order High Pass Filter (HPF) with a 4Hz cutoff frequency, in order to remove the low frequency noise caused by wind.

3. A 50Hz notch filter to remove the European power grid effect.

4. Both previous filters applied at once, where SRs should be properly represented.

5. A Band Pass Filter (BPF) with 36Hz bandwidth, from 6Hz to 40Hz, containing the first fourth modes of the SRs, which represent most of the SRs power.

Once filtered, a data file is split to analyze each of the resulting segments. Since optimal segment duration to analyze the ELF spectrum has not been quite discussed in the literature [26], a range of durations was chosen, specifically 10 minutes, 5 minutes, 1 minute and 20 seconds. This is done under the notion that longer segments will contain less transitory events than background noise, making them representative for the quasi-stationary state of the signal. On the

4
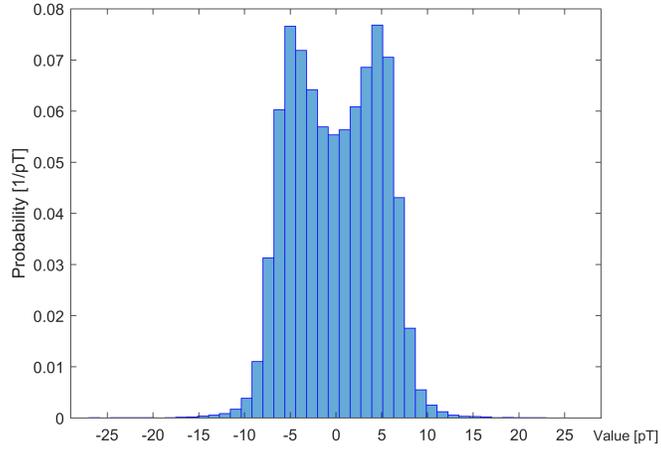
(a) Distribution from the sample depicted in Fig. 1a.
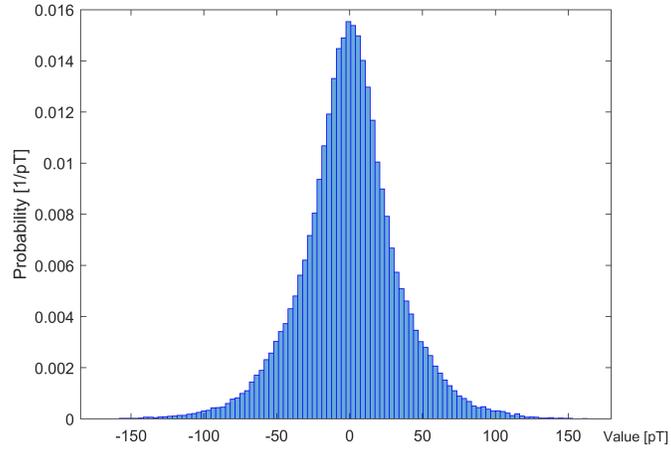
(b) Distribution from the sample depicted in Fig. 1b

Figure 2: Histograms of 10-minute segments.

other hand, short duration segments could contain only transitory events, even when most of them will be populated by background noise. Therefore, shorter segments might be useful to identify specific transitory events, or even locate odd phenomena. This choice of ranges means to explore the duality explained above.

From here onwards, a data stream of certain duration subjected to an specific filtering process will be referred to as a segment, in order not to confuse it with samples, which are used to reference the data points a segment or other kind of

register has.

2.3 Analysis description

To determine the statistical nature of the captured signal, our analysis finds the statistical distribution that fits each segment best, choosing from a set of target distributions. The steps are as follows:

1. For each target distribution, its Maximum Likelihood Estimates (MLE) in relation to the analyzed segment are calculated to obtain the parameters of this distribution family that fits the data best.

2. Akaike Information Criterion (AIC) is applied to all distribution candidates, calculated above.

3. Akaike weights [27] are calculated to be able to compare between the distribution candidates. Through them the candidate that fits the data best is chosen.

Given the above, it is clear that one of the key points of this work is the set of target distributions. After carefully inspecting the available literature, the only documented fact found was how the probability distribution of the ELF spectrum shows heavy tails [15]. Besides that, SRs are thought to be properly represented by the Normal (or Gaussian) distribution. The few works that mentions the previous statement do it without including any references [22], and when studied, only anecdotal analyses are performed [23].

Consequently, a set of target distributions should be chosen with their kurtosis being above the Normal distribution value. The chosen distributions are the following.

- *Normal (or Gaussian) distribution:* With no rigorous studies being performed to the best of our knowledge, the statements about SRs being described by the Normal distribution should be tested. Therefore, it is a clear choice for the analysis' set of target distributions. Its Probability Density Function (PDF) is given by Equation 1.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1}$$

With σ being the distribution's standard deviation and μ its mean.

- *Rician (or Rice) distribution:* This distribution is typically used to model line of sight scatter. A storm cloud footprint could be considered circular, with the probability of lightning strike being a bivariate normal and the observer being at certain distance from the center of the storm. If the above holds true, this phenomena could be modeled by a Rician distribution. Its PDF is presented in Equation 2, with s being the distribution's

6

noncentrality parameter, σ its scale parameter, and I_0 is the zero-order modified Bessel function of the first kind.

$$f(x) = \frac{x}{\sigma^2} e^{\frac{-(x^2+s^2)}{2\sigma^2}} I_0 \left(\frac{xs}{\sigma^2} \right) \tag{2}$$

- *Laplace distribution:* This distribution is interesting for two reasons. It is one of the symmetric distributions with higher kurtosis. Besides, its distinct peak has been observed on histograms containing high amplitude samples. The relevant parameters of its PDF (Equation 3) are the location parameter μ and the scale parameter b.

$$f(x) = \frac{1}{2b} e^{\frac{-|x-\mu|}{b}} \tag{3}$$

- *Logistic distribution:* This distribution's main features makes it suitable for this study. While being bell shaped, it has a higher kurtosis value than the Normal distribution. Its PDF is featured in Equation 4, with μ and s being its location and scale parameters, respectively.

$$f(x) = e^{-\frac{(x-\mu)}{s}} \frac{1}{s \left(1 + e^{-\frac{(x-\mu)}{s}}\right)^2} \tag{4}$$

An edge case that should be monitored is the segments whose best fit distribution do not properly describe it. This happens because the best fit obtained through AIC means that the resulting distribution is the one that better explains the data among the target distributions. This does not guarantee that the resulting distribution appropriately explains the data. Therefore, a way to quantify the Quality of Fit (QoF) is needeed.

A quantile-quantile plot or Q-Q plot is a graphical method that plots the quantiles of a dataset against those of a statistic distribution. Since they are used to evaluate the relationship between the dataset and the distribution, it was decided to rely on them in order to ascertain the QoF of each segment. This QoF coefficient is obtained by performing a linear regression between the data points of the segment and the distribution as presented in a Q-Q plot. Using regression has the advantage of providing a widely used and easy to understand value which in turn helps deciding an acceptable threshold to accept or reject the fit.

This QoF coefficient has the additional advantage of giving an estimate of the method's accuracy for each distribution family under each filtering process. Therefore, it provides a way to evaluate each sample's fit quantitatively and interpret the results accordingly. This is appropriate for this early stage of the algorithm's development, leaving a more rigorous uncertainty analysis for the moment when the algorithm advantages have been developed further, and its flaws diminished.

To maximize the confidence in the results, only segments with a regression coefficient above 99% are considered to be appropriately described by their best fit distribution.

An edge case that must be covered are Bimodal distributions, due to sinusoidal signals' interference as the one in Fig. 2a. If uncontrolled, these segments will be inappropriately classified.

It is worth highlighting that the interest for the Bimodal distribution comes only as an edge case. In the current development stage of the algorithm, the main objective is to exclude Bimodal segments before the classification process begins. Some issues arise while trying to fit segments to Bimodal distributions, which is why no data about them is stored by our algorithm.

To detect the segments with an accused degree of Bimodality (meaning those whose two maxima are apart enough from each other to be inappropriately clasified), a graphical method was developed.

1. The segment's histogram contour is calculated by generating a data point per histogram bin, centered on the bin top.

2. The resulting function is split in two by the data point closer to the histogram's median.

3. The maximum value of each half is found and analyzed to check if the abscissa value is within a predefined interquantilic range.

4. The sample is labeled as Bimodal if both maxima are outside of the chosen interquantilic range.

The aforementioned range should be carefully chosen. First and foremost, it must have a proper level of selectivity to reject the segments showing enough bimodality. On the other hand, it should not reject segments that, despite having certain bimodality, could still be classified. Through many tests and visual inspection of the results, the range between the 3rd and the 5nd octile has been determined as appropriate.

3   Results and Discussion

The previously explained method was applied to ELF data from April 2016, obtained in the previously described station in Sierra de los Filabres.

Table 1 consolidates the outcome of the whole analysis process, grouped in sub-tables by segment duration. Each row on each sub-table presents the amount of segments identified as any of the possible distributions (columns) after applying the specified filter to the data. This value is expressed under Ratio. It indicates the ratio of segments whose best fit is the distribution in the column title to the total amount of analyzed segments. Next to each ratio of positives, the QoF results of the distribution's segments are presented. Also in %, the QoF results are calculated as the ratio of the amount of that distribution's segments whose QoF is over 99% to the total amount of segments classified under that distribution. Bimodal distribution column does not come with a QoF ratio since its classification is not the product of a fitting process.

These results are described as follows.

1. Bimodal distributions are common under Raw and HPF processes. Its ratio of positives is mostly constant regardless of segment duration. Under the other three filters they are almost absent, with its ratio reaching 1% only in 20 seconds segments.

2. Normal distribution performs well regardless of segment duration or filter applied to the data, its ratio of accepted segments (QoF > 99%) being over 80% in almost every case. Its ratio of positives is only significant under HPF in the 10 and 5 minute duration segments, but experiences an overall increase in the lower duration segments, specially under BPF.

3. Rician distribution has the lowest ratio of positives of all target distributions, having only noteworthy presence on 20 second segments. Its segments show a low amount of accepted segments on all the raw and HPF analysis. Also, Their QoF ratio steadily decreases for the 50Hz notch filtered data as segments got shorter. Nonetheless, under the HPF and notch and BPF processes its acceptance is high overall.

4. Laplace distribution displays the lowest acceptance ratio, the maximum

Table 1: Distribution fitting results.

10 MINUTE SEGMENTS, TOTAL 8640 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	24.35	3.97	83.97	0.72	51.61	32.45	37.27	38.51	63.21
HPF	38.01	16.20	86.71	2.96	31.64	3.28	62.54	39.55	78.67
50Hz notch	0	1.09	100	0.05	100	60.63	31.79	38.24	57.51
HPF and notch	0	0.03	100	0		7.84	51.99	92.13	78.54
BPF	0	3.70	100			3.50	68.87	92.80	93.02

5 MINUTE SEGMENTS, TOTAL 17280 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	24.34	6.15	88.90	1.56	61.85	26.62	32.65	41.33	66.00
HPF	38.41	16.36	87.05	3.44	38.99	2.75	62.11	39.04	77.65
50Hz notch	0.2	2.80	99.38	0.33	100	50.60	27.06	46.08	60.92
HPF and notch	0	0.34	100	0.01	100	6.93	52.72	92.73	79.09
BPF	0	6.52	100	0.01	100	3.04	63.43	90.43	92.14

1 MINUTE SEGMENTS, TOTAL 86400 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	25.77	13.14	90.60	12.71	76.76	15.47	11.96	32.91	59.20
HPF	38.63	17.65	90.87	8.05	56.17	1.78	47.43	33.89	77.61
50Hz notch	4.62	10.98	93.26	9.28	78.83	31.69	8.42	43.42	54.07
HPF and notch	0.21	4.49	100	0.62	100	5.71	31.33	88.96	82.18
BPF	0.31	21.68	100	1.58	100	2.2	41.15	74.23	90.62

20 SECOND SEGMENTS, TOTAL 259200 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	31.21	11.94	79.89	23.44	58.63	11.34	1.62	22.07	38.24
HPF	39.04	18.32	93.16	14.99	64.93	1.23	26.84	26.42	76.97
50Hz notch	14.29	10.76	77.28	18.05	55.04	24.85	2.59	32.05	38.86
HPF and notch	1.56	11.87	99.82	5.09	99.08	5.62	19.80	75.86	83.23
BPF	2.06	31.44	99.96	10.03	99.82	1.99	24.25	54.48	88.71

value being 68.87% for 10 minute segments under BPF. Despite having high ratio of positives under 50Hz notch filter on 10 and 5 minute segments, and medium to low on all raw data analysis, it decreases dramatically in the rest of filters.

5. Logistic distribution segments are present in all filtering processes, its ratio of positives being specially high under HPF and notch and BPF processes for 10 and 5 minutes segments. There is a slight decrease in both as segment size decreases, mostly in the latter process. Its acceptance ratio generally also decreases with duration, but it's high overall under HPF and notch and BPF.

The segmentation criteria can be corroborated through the experimental standard deviation of the segments' mean. Since it is an estimate of the uncertainty caused by the disturbances of the signal, it is also an indication of the impact transitory events has on the measurement. These results are presented in Table 2.

Table 2: Experimental standard deviation of segment's mean, by segment duration (columns) and applied filter (rows).

	10 minutes [fT]	5 minutes [fT]	1 minute [fT]	20 seconds [fT]
Raw	82.65	115.56	260.51	516.17
HPF	0.42	0.88	4.85	13.88
50Hz notch	82.65	114.56	260.52	516.16
HPF and notch	0.41	0.87	4.80	13.70
BPF	0.17	0.38	1.98	6.13

The way dispersion increases as segment duration decreases illustrates how random error can be compensated by increasing the number of samples per segment. It also indicates how background noise dominates in the longest segments, with a dispersion around 8 times smaller than the shortest. Nevertheless, higher dispersion on shorter segments shows their sensibility to transitory events, a fact reinforced by the general decrease in acceptance ratio shown in Table 1 as segment length is reduced. Looking at the table by rows, we should note that values of Raw (unfiltered) and HPF segments are quite similar. It shows how low frequency disturbances are more detrimental for the signal than the 50Hz power grid signal, with the latter having next to no effect in the uncertainty compared with the former. This can be seen as well in the small difference between HPF and notch and HPF filtered segments' values. Nonetheless, BPF segments do experience an improvement over the other two. The low standard deviation of its segments' mean indicates a low value of uncertainty, which in turn implies the lowest amount of transitory events among all filters.

3.1 Distribution fitting

On Fig. 1b and Fig. 2b it was shown how high power, low frequency noise masks the power grid signal on its histogram representation. Comparing between Raw
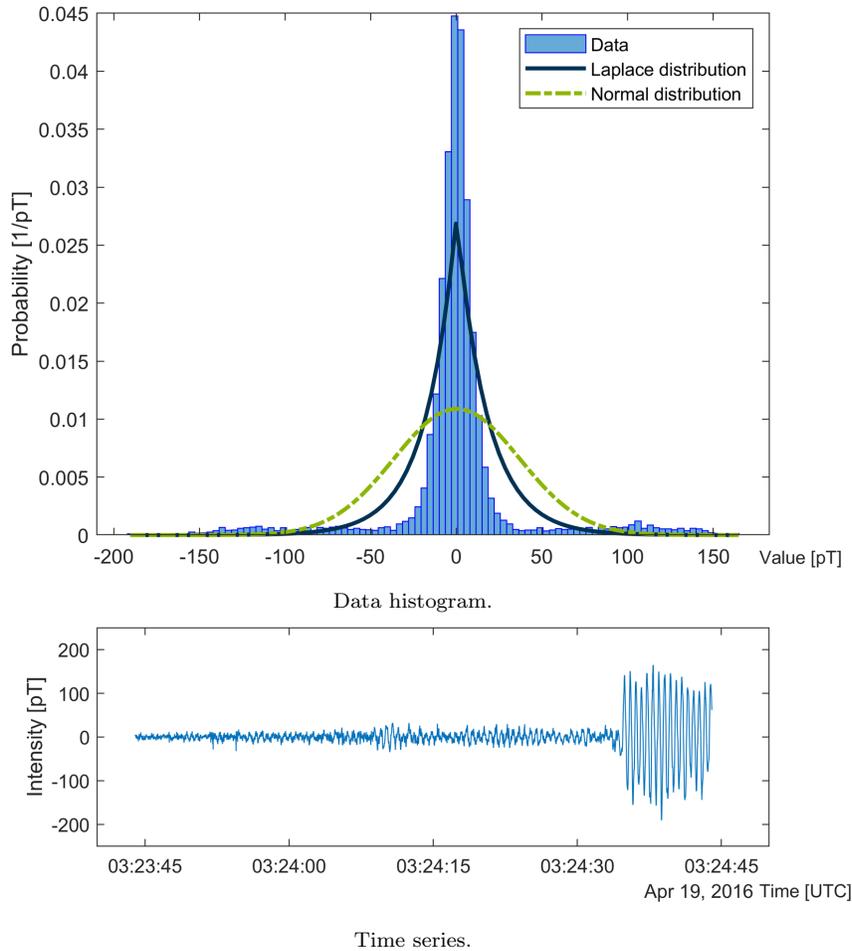
Figure 3: Notch filtered segment from NS, 19-Apr-2016 03:23:44, 60 seconds duration. Low QoF Laplace fit due to high intensity disturbance.

and HPF data, it can be seen how the ratio of Bimodal distributions actually increases from the former to the latter. It shows how the 50Hz signal is more frequent in the segments' histogram after cleaning the low frequency noise. After filtering, there are over 50% more Bimodal segments in 10, 5 and 1 minute segments, whereas in 20 seconds segments the increase is only of 25%. On the other hand, after going through the 50Hz notch filter, all Bimodal distributions disappear in the 10 minute segments, but are increasingly present as segment duration is reduced. This is another consequence of a histogram's scope in time. It shows how segments of short duration are prone to reveal histograms that are either odd (E.g. Bimodal) or hard to classify, as shown by the low number of accepted segments in the 20 second duration.

This line of reasoning might be followed to understand Laplace fits. Their

moderate ratio of positives in raw analysis points out this particular distribution is also related with disturbances, and their high ratio under 50Hz notch filter points at low frequency contributions from wind as their main source. The differences in amplitude between disturbances and ELF background noise contribute to bin dispersion, resulting in histograms with high kurtosis value. Fig. 3 illustrates this case; it shows a segment with low QoF value due to a sudden increase in amplitude near its end.

It can be seen how this is translated into a few high value bins appearing on the histogram, thus expanding the histogram and making the fit difficult.

This interpretation is reinforced by the low values of acceptance ratio displayed by this particular distribution. If we consider segments with Laplace distribution as their best fit as mainly influenced by transitory events, the histogram will usually be presenting two or more distinguishable events. At the very least, the always present ELF background noise and one or more disturbances. Nonetheless, many Laplace segments do get accepted; in those cases it is shown we are facing a noisy segment despite the filters applied. Fig. 4 depicts one of these cases.

Something similar happens with Logistic distribution fits. The verage QoF of Logistic fits value throughout all filters is higher than Laplace's, the former being 99.20 while the latter is 95.03. Nonetheless, due to the high QoF demanded, many samples are not accepted as well. Fig. 5 shows a case where high intensity disturbances introduce extreme values in the histogram but, due to higher segment duration compared with Fig. 3, its impact is lower.

Even when by the nature of its definition it might be connected lightning discharges, the ratio of Rician distribution fits is only meaningful in the 1 minute and 20 second segments. Nonetheless, their ratio of acceptance is insufficient to validate the possibility of modeling thunderstorms through this distribution. This conclusion is reinforced by the fact that Akaike weights' value for Normal and Rician distributions are both significant on Rician fitted segments. Besides, through visual inspection, most Rician fits show that Normal distribution would have been an equally acceptable fit. Fig. 6 is a representative example of most Rician fits, where Normal distribution overlaps with the Rician distribution representation.

Lastly, Normal distribution's ratio of acceptance is very stable regardless of segment duration or filter process, pointing out a very specific and well represented phenomenon being captured. Its ratio is only representative on HPF data on the long duration segments, but they are also present in the shorter segments of BPF data. Looking at the amount of segments whose best fit is a Normal distribution, it is understandable why SRs are usually identified with it. Nonetheless, they never overcome the % of Logistic fits. It has been suggested that the amplitude of radio pulses produced by lightning strikes follows a Normal distribution [14] so there might be a relationship between the time series' signal being represented by a Normal distribution and lightning activity. An initial exploration of this hypothesis will be performed further in the article.

Data histogram.

Time series.
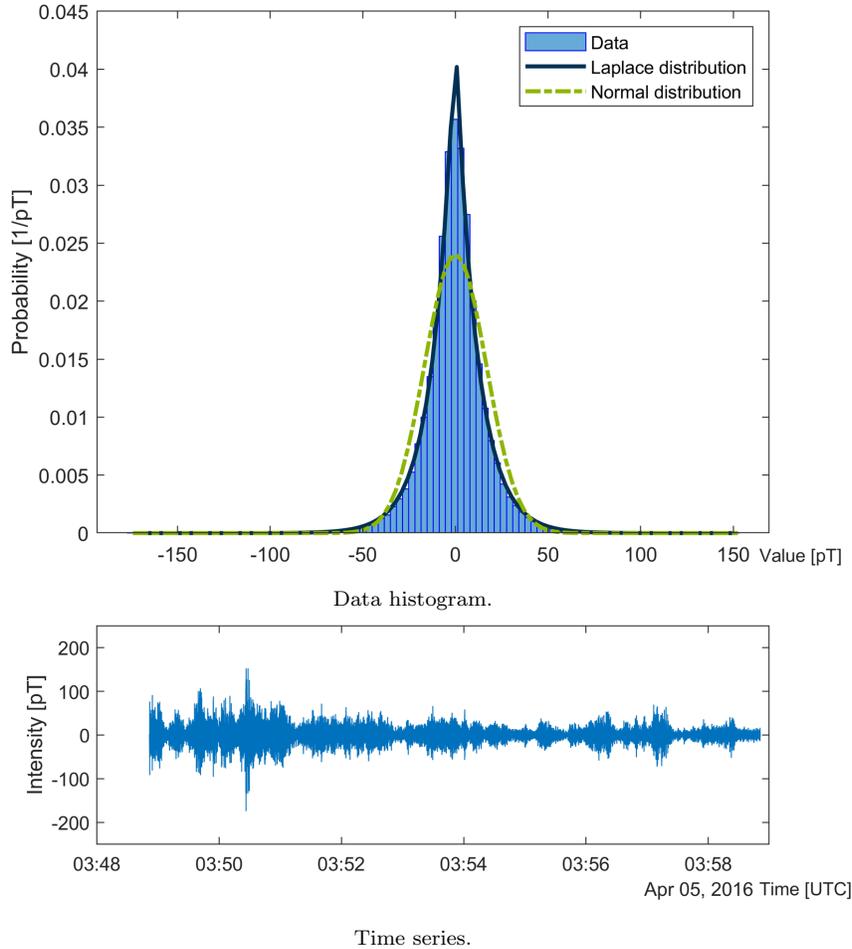
Figure 4: Band pass filtered segment from EW, 05-Apr-2016 03:48:51, 600 seconds duration. Accepted Laplace fitted segment.

3.2 SRs statistical characterization.

To properly characterize the SRs, the attention must be shifted to HPF and notch and BPF rows, being the ones that more effectively isolate SRs signal. On these, the best fitting distribution by far is the Logistic distribution, which given its characteristics of heavy tailed, bell shaped distribution, falls in agreement with the literature as a proper candidate [15]. As stated before, its predominance is clear in the ten, five and one minute duration segments, whereas in the twenty second segments the amount of Logistic distribution fits decreases, specially with BPF data, and mostly in benefit of Normal and Rician distributions.

Fig. 7 presents an example of how Logistic distribution fits segments free of disturbances, with Fig. 7a coming from BPF data and Fig. 7b from HPF

13

Data histogram.

Time series.
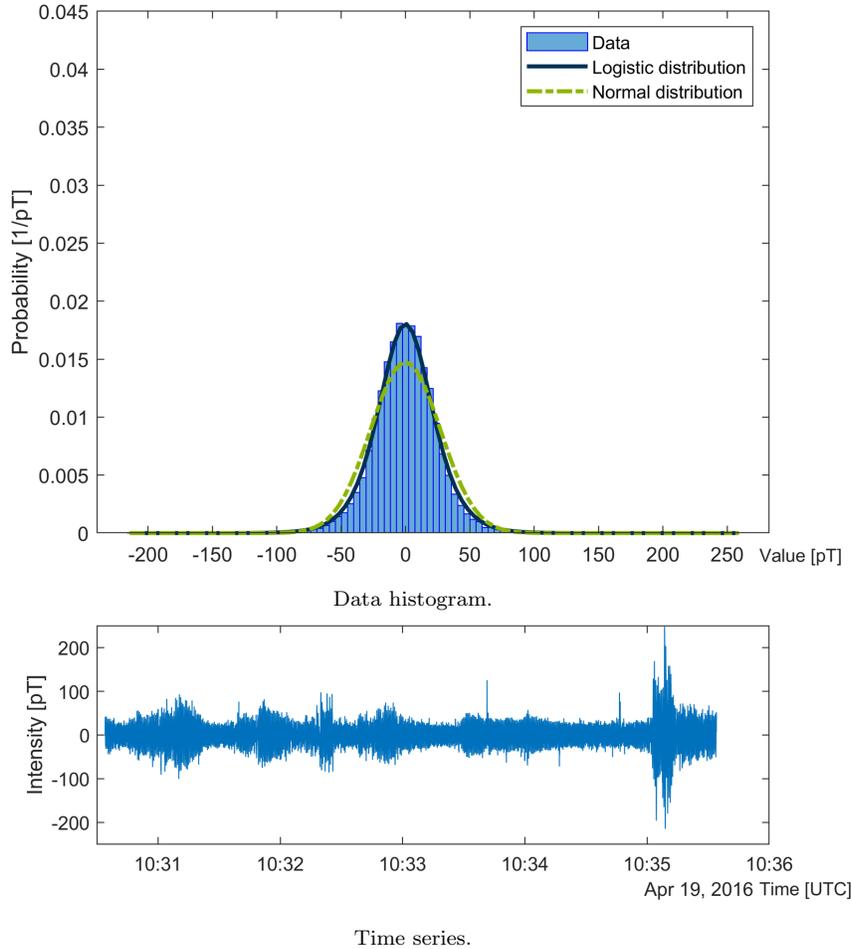
Figure 5: Raw segment from NS, 19-Apr-2016 10:30:34, 300 seconds duration. Rejected Logistic segment with QoF over 95%.

and notch data. Besides that, they represent the same time interval. Other fact worth mentioning is that while both could be considered good fits at a glance, Fig. 7a has a QoF value over 99% whereas Fig. 7b doesn't. This is due to the extreme values displayed in the latter, otherwise filtered in the former. The comparison serves to illustrate how strict the established QoF conditions are; a fact that should be kept in mind while evaluating the results.

To sum up, Fig. 7 serves not only to illustrate how ELF background noise can be represented accurately by a Logistic distribution, but also to see BPF as the most appropriate filter to study SRs.

In order to further elaborate this point, Table 3 presents the analysis results separated by channel of reception, with Table 3a gathering the EW channel and 3b the NS channel. The first item to highlight is how Logistic distribution's

Data histogram.

Time series.
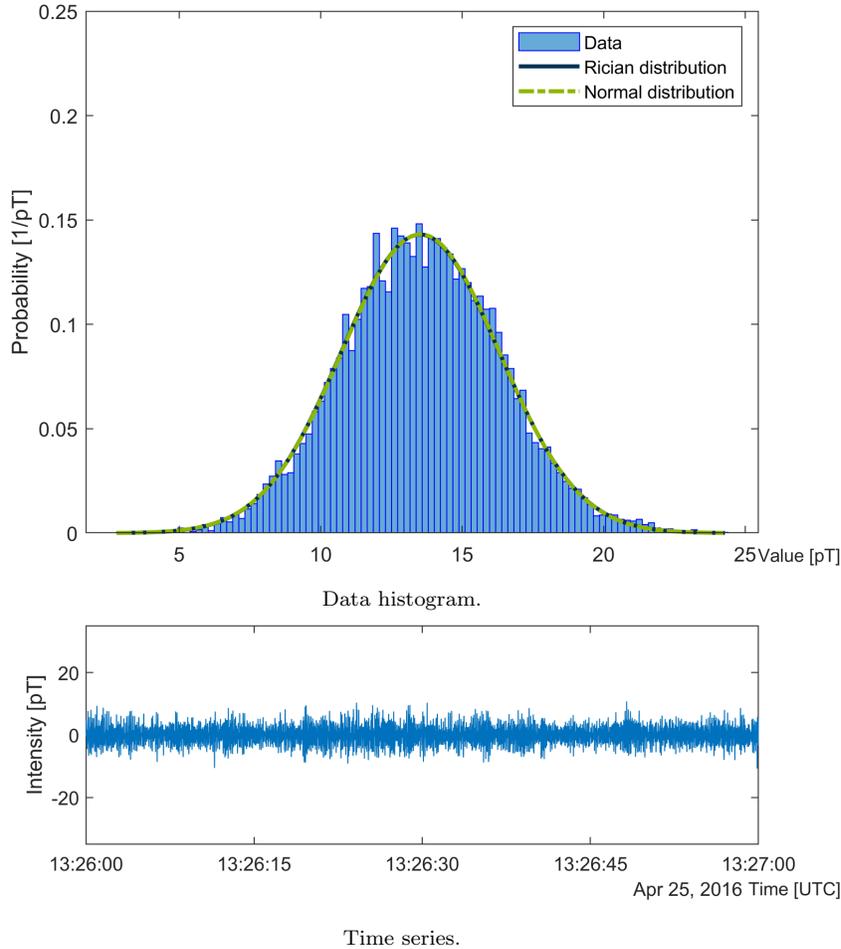
Figure 6: Band pass filtered segment from EW, 25-Apr-2016 13:26:00, 60 seconds duration. Accepted Rician segment.

ratio of positives is high under HPF and notch and BPF on both channels, with remarkably similar values in the 10 and 5 minute segments. Nonetheless, a noticeable difference appears on the BPF process in the 1 minute and 20 seconds duration, where its ratio in the EW channel is 20% and 10% lower, respectively, than in the NS channel. Once again, it can be perceived how the decrease of Logistic ratio happens in favor of Normal distribution's, which are more common in EW than NS. This last fact is of great importance, since it is another stepping stone towards the hypothesis of Normal distribution ratio of positives being linked to lightning activity. Our station's location and proximity to the African thunderstorm center makes the EW channel of our sensor specially sensitive to its activity [28]. The intensity of its influence might account for the difference in Normal distribution ratio between EW and NS channels.

15
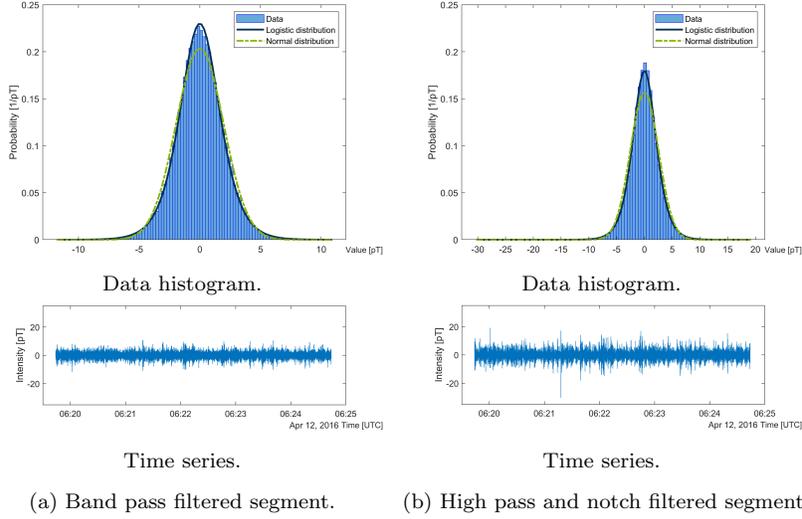
Figure 7: Segment from EW, 12-Apr-2016 06:19:44, 300 seconds duration, under two different filtering processes. Logistic fitted segment containing mainly ELF background noise.

Bimodal and Laplace distributions' ratio are very different between channels as well. Segments with these distributions as best fit have been shown to be affected by power grid interference and low frequency disturbance, respectively. Table 3 shows how each channel is affected by each one of these effects in different proportions. NS channel shows a higher number of Bimodal distributions than EW on raw data, and even higher on HPF data. This points out that the NS channel is coupled with some kind of power station. On the other hand, EW channel seems more sensitive than NS to wind disturbances, given the high ratio of Laplace distributions on raw data and specially under 50Hz notch filtering. Nonetheless, the low ratio of positives of both distributions under HPF and notch and BPF show that these disturbances are removed, with similar ratios to those depicted on Table 1. This illustrates once again the importance of proper filtering before analyzing data.

3.3 Normal distribution ratio and lightning activity

The hypothesis of the Normal distribution ratio being an expression of global lightning activity has been hinted previously in the article. In this section, an inspection of this hypothesis will be performed to gather initial evidence to test its feasibility. For this procedure, we focus on the 1 minute segments, based on how segments of this duration display a higher Normal distribution ratio than the previous while Bimodal, Rician and Laplace distributions ratio stays low (see Table 1). The BPF analysis will be used, since it is the filter that most clearly presents the SRs signal. Results were arranged hourly by grouping the ratio of each distribution throughout the month by UTC hours of the day.

16

Table 3: Distribution fitting results separated by channel.

(a) East-West channel results.

10 MINUTE SEGMENTS, TOTAL 4320 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	2.73	0.86	100	0.09	25	52.38	34.82	43.94	59.85
HPF	5.32	21.99	90	1.76	48.68	4.91	63.68	66.02	80.47
50Hz notch	0	0.49	100	0		70.76	30.98	28.75	73.43
HPF and notch	0	0		0		7.48	60.06	92.52	85.21
BPF	0	6.16	100	0		3.54	67.32	90.30	93.90

5 MINUTE SEGMENTS, TOTAL 8640 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	2.71	2.36	96.57	0.21	88.89	43.94	30.01	50.79	64.15
HPF	5.49	22.49	90.27	2.58	56.95	4.06	61.82	65.38	79.13
50Hz notch	0.06	1.70	100	0.12	100	61.09	26.13	37.04	74.53
HPF and notch	0	0.15	100	0		6.45	61.04	93.40	84.91
BPF	0	10.43	100	0.02	100	3	63.32	86.55	92.81

1 MINUTE SEGMENTS, TOTAL 43200 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	4.78	13.06	95.32	7.67	90.65	27.08	10.01	47.41	59.15
HPF	6.16	25.51	93.82	8.41	71.68	2.64	44.51	57.29	79.42
50Hz notch	2.74	11.07	95.65	7.13	88.64	41.26	7.98	37.79	67.79
HPF and notch	0.19	3.70	100	0.61	100	4.56	38.26	90.94	85.86
BPF	0.36	28.18	100	2.01	100	2	42.61	67.45	91.43

20 SECOND SEGMENTS, TOTAL 129600 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	11.12	14.55	82.64	19.45	65.46	20.94	1.32	33.94	37.10
HPF	7.82	27.60	95.50	17.42	79.85	1.95	24.23	45.23	78.66
50Hz notch	10.44	12.36	78.38	18	61.35	33.99	1.40	25.20	41.21
HPF and notch	1.54	11.14	99.85	4.94	99.47	4.27	20.32	78.11	85.82
BPF	2.17	36.18	99.98	10.83	99.96	1.67	23.07	49.15	89.72

To start exploring this relationship, all the grouped samples were plotted by hours using boxplots. The results for both channels can be seen on Fig. 8.

(a) EW channel.

(b) NS channel.
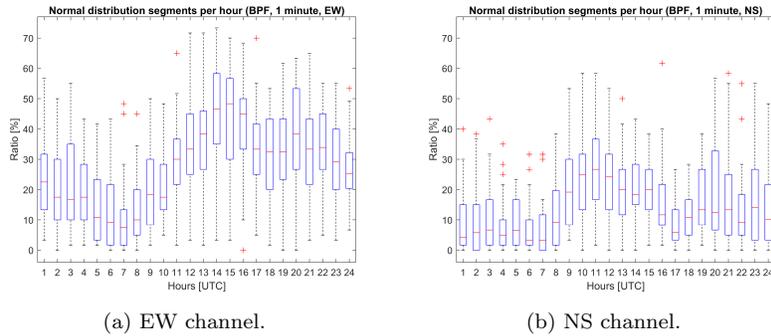
Figure 8: Normal distribution ratio of positives in 1 minute segments each day, organized by hours.

(b) North-South channel results.

10 MINUTE SEGMENTS, TOTAL 4320 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	45.97	7.08	82.03	1.34	53.45	12.52	47.50	33.08	67.67
HPF	70.69	10.42	79.78	4.17	24.44	1.64	59.15	13.08	69.56
50Hz notch	0	1.69	100	0.09	100	50.49	32.92	47.73	47.91
HPF and notch	0	0.07	100	0		8.19	44.63	91.74	71.81
BPF	0	1.25	100	0		3.45	70.47	95.30	92.18

5 MINUTE SEGMENTS, TOTAL 8640 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	45.97	9.94	87.08	2.92	59.92	9.31	45.15	31.86	68.94
HPF	71.33	10.23	79.98	4.31	28.23	1.44	62.90	12.70	70.01
50Hz notch	0.34	3.89	99.11	0.54	100	40.10	28.48	55.13	51.77
HPF and notch	0	0.53	100	0.01	100	7.41	45.47	92.05	73.18
BPF	0	2.60	100	0		3.08	63.53	94.32	91.53

1 MINUTE SEGMENTS, TOTAL 43200 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	46.76	13.23	85.93	17.75	70.75	3.86	25.69	18.41	59.35
HPF	71.10	9.78	83.18	7.70	39.22	0.92	55.78	10.50	67.74
50Hz notch	6.50	10.90	90.82	11.42	72.70	22.12	9.24	49.06	43.51
HPF and notch	0.22	5.28	100	0.63	100	6.87	26.74	86.99	78.33
BPF	0.27	15.18	100	1.16	100	2.39	39.92	81.01	89.94

20 SECOND SEGMENTS, TOTAL 129600 SEGMENTS/ROW

	Bimodal	Normal		Rician		Laplace		Logistic	
		Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%	Ratio [%]	QoF>99%
Raw	51.30	9.32	75.61	27.43	53.78	1.75	5.26	10.21	42.02
HPF	70.26	9.05	86.05	12.56	44.24	0.51	36.82	7.62	66.91
50Hz notch	18.13	9.17	75.78	18.11	48.76	15.70	5.17	38.90	37.34
HPF and notch	1.58	12.60	99.79	5.24	98.72	6.98	19.48	73.61	80.49
BPF	1.95	26.69	99.93	9.24	99.66	2.31	25.15	59.80	87.88

Fig. 8a depicts the Normal distribution ratio rising at 8:00 UTC and peaking at 15:00 UTC, which falls in line with the African thunderstorm's intensity pattern. This is interesting because as it was stated before, the EW channel of our station is strongly influenced by the African thunderstorm center due to its orientation and proximity. In the same way, the noon crest on Fig. 8b is quite similar to the Asian thunderstorm intensity pattern (the most influential thunderstorm center for this channel), although delayed in time by two hours. Nonetheless, the samples for the rest of the hours in both Fig. 8a and Fig. 8b show a pattern unrelated with their respective most influential thunderstorm center.

It must be acknowledged that boxplot whiskers indicate high variance for every hour of both channels. Despite that, there are only a few atypical values and, even when widely distributed, the boxplot for every hour shows coherence between median values and its quantile distribution. We consider this a piece of evidence towards the validity of the hypothesis, albeit small.

Next, a straight comparison between the median values of the Normal distribution ratio and an average estimation of thunderstorm intensity for each thunderstorm center, extracted from [28].

This will allow to visually account the similarities and differences between hourly Normal distribution ratio by days and thunderstorm activity. Fig. 9 depicts the mentioned data, on which NS Normal distribution occurrence has been given a two hours advance in order to enhance its visual comparison with the Asian thunderstorm center.
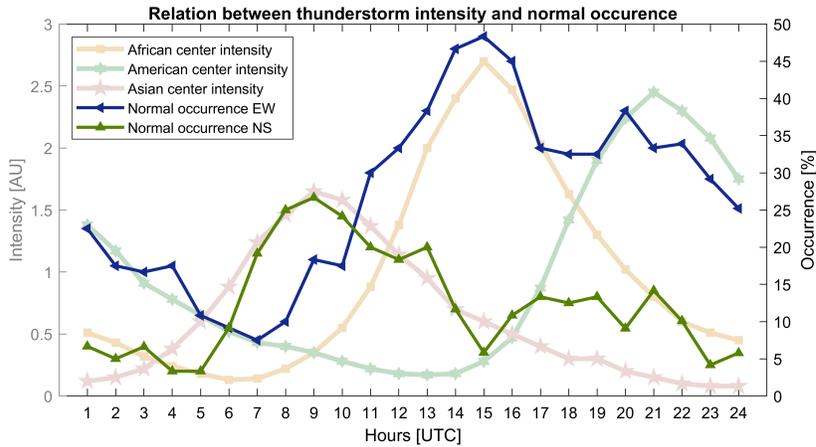
Figure 9: Thunderstorm intensity (faded lines, left axis) and median values of hourly Normal distribution ratio (solid lines, right axis). Normal distribution ratio in NS channel shifted two hours forward.

This shows how EW data fits not only the African thunderstorm activity, but also follows the American one in the hours where its intensity is higher. This offers a possible explanation to the behavior shown on all hours. Looking

at how EW data follows both African and American thunderstorm activity with no apparent attenuation gives evidence for Normal distribution ratio to be related with global thunderstorm activity, with independence of of source-observer distance, though source-observer angle is most probably a factor to be considered.

Lastly, the relation observed in the previous figure for both channels will be tested by means of linear regression. An analysis will be performed for each channel, between the hourly median ratio of positives for the Normal distribution and the most influential thunderstorm intensity center. Along with them, a Student's t-test will be performed to test the correlation, with its null hypothesis being that the relationship between the two variables holds no significance. The results can be seen in Fig. 10.

(a) Regression analysis for EW channel.   (b) Regression analysis for NS channel.
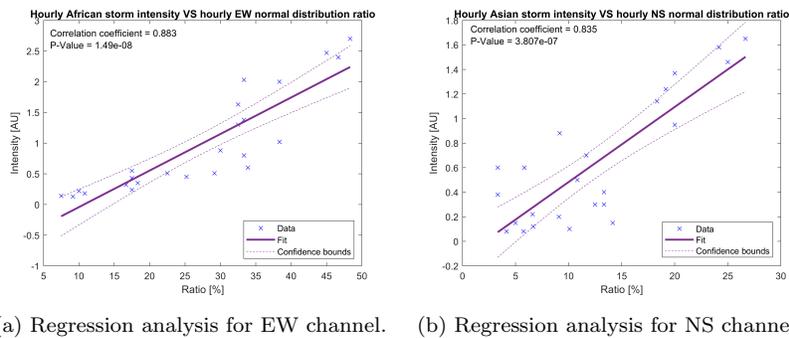
Figure 10: Linear regression analysis with Student's t-test results

Although data deviation is high, The regression coefficient points out to a strong relationship between hourly Normal distribution ratio and the most influential storm center for each channel. Furthermore, the obtained P-value for both correlations corroborates the relationship by strongly rejecting the null hypothesis.

Still, the perceived relationship between shifted NS data and the Asian thunderstorm center intensity can be tested as well. Another linear regression between these two variables was performed in the same terms than the previous one. As Fig. 10b shows, there is a strong positive correlation of statistical significance.

The conducted tests shows interesting pieces of evidence favoring the hypothesis of Normal distribution ratio being related to global lightning activity. Nonetheless, the two hours delay in NS channel remains to be explained, and further research is needed in order to prove the extent of this relationship. That being said, such tests are out of the scope of this paper, since its objective was to present the analysis method and demonstrate its utility.

4   Conclusions

Statistical analysis was applied to characterize the time series of the Schumann Resonances and the bandwidth of ELF spectrum where they are located, by means of maximum likelihood parameters estimation and distribution fitting through Akaike Information Criterion. This is a kind of analysis that, to the best of our knowledge, has never been applied to these signals. The analysis method performs the same analysis over the same data by selecting different segment sizes and different filters, allowing for comparative observations between them. Ways to ensure the Quality of Fit were included in the algorithm, like a graphical method to detect Bimodal segments due to an excess of sinusoidal interference and a Q-Q plot based Quality of Fit value for each classified segment.

In order to test its capabilities, the algorithm was applied to a month worth of data from our ELF station located in Sierra de los Filabres, Almería. The method's results and their utility have been discussed, first by analyzing the results by themselves and last by means of a comparative study. From this discussion four major points should be highlighted.

1. The statistical nature of ELF background noise appears to be closely related with the Logistic distribution.

2. As expected, longer segments represent the ELF background noise better whereas the shorter ones show certain sensibility towards transitory effects. A special mention should be done for 1 minute segments. Among all the tested durations, they were short enough to display transitory behavior while keeping a high acceptance.

3. Each of the different target distributions seems to fit ELF records under different circumstances, serving as a sort of behavioral classification.

4. The connection between global lightning activity and Schumann resonances is a well established fact. With the small comparative study performed in this article, we have uncovered evidence for a relationship between the main thunderstorm center's intensity and the amount of segments best represented by the Normal distribution by hour. Further studying this hypothesis could open new ways of looking at this phenomena.

By means of the analysis method proposed in this paper, the potential for ELF time series study has been shown, with evidence pointing out for another expression of global lightning activity in Schumann Resonances. It shows an untapped way of progressing in other lines of research, since the studies performed from the time series perspective are scarce. It could also contribute to refine the response of time series oriented simulations.

Looking through the results offered by statistical analysis, a different approach of analyzing ELF spectra has been tested by finding key ELF facts in the results. It is worth to keep working in this kind of analysis to see what else it has to offer, specially in relationship with other phenomena; either those who have been tried and tested (global temperature) or those who are currently being investigated (earthquakes).

5 Acknowledgements

References

[1] W. O. Schumann, "Über die strahlungslosen eigenschwingungen einer leitenden kugel, die von einer luftschicht und einer ionosphärenhülle umgeben ist," *Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences*, vol. 7, no. 2, pp. 149–154, 1952.

[2] C. Price, "ELF electromagnetic waves from lightning: The schumann resonances," *Atmosphere*, vol. 7, no. 9, 2016.

[3] E. Prácser, T. Bozóki, G. Sátori, E. Williams, A. Guha, and H. Yu, "Reconstruction of global lightning activity based on schumann resonance measurements: Model description and synthetic tests," *Radio science*, vol. 54, pp. 254–267, 2019.

[4] A. Kulak, J. Kubisz, A. Michalec, S. Zieba, and Z. Nieckarz, "Solar variations in extremely low frequency propagation parameters: 2. Observations of Schumann resonances and computation of the ELF attenuation parameter," *Journal of Geophysical Research: Space Physics*, vol. 108, no. A7, pp. 1–9, 2003.

[5] M. Pazos, B. Mendoza, P. Sierra, E. Andrade, D. Rodríguez, V. Mendoza, and R. Garduño, "Analysis of the effects of geomagnetic storms in the schumann resonance station data in mexico," *Journal of atmospheric and solar-terrestrial physics*, vol. 193, pp. 105 091–, 2019.

[6] A. P. Nickolaenko, I. G. Kudintseva, O. Pechony, M. Hayakawa, Y. Hobara, and Y. T. Tanaka, "The effect of a gamma ray flare on Schumann resonances," *Annales Geophysicae*, vol. 30, no. 9, pp. 1321–1329, 2012.

[7] I. G. Kudintseva, Y. P. Galuk, A. P. Nickolaenko, and M. Hayakawa, "Modifications of middle atmosphere conductivity during sudden ionospheric disturbances deduced from changes of schumann resonance peak frequencies," *Radio science*, vol. 53, pp. 670–682, 2018.

[8] E. R. Williams, "The Schumann Resonance : A Global Tropical Thermometer," *Science*, vol. 256, no. 5060, pp. 1184–1187, 1992.

[Online]. Available: http://search.proquest.com/docview/213547553?pq-origsite=gscholar

[9] V. Christofilakis, G. Tatsis, C. Votis, I. Contopoulos, C. Repapis, and V. Tritakis, "Significant elf perturbations in the schumann resonance band before and during a shallow mid-magnitude seismic activity in the greek area (kalpaki)," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 182, pp. 138 – 146, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364682617305278

[10] J. Gazquez, R. Garcia, N. Castellano, M. Fernandez-Ros, A.-J. Perea-Moreno, and F. Manzano-Agugliaro, "Applied engineering using Schumann Resonance for earthquakes monitoring," *Applied Sciences (Switzerland)*, vol. 7, no. 11, 2017.

[11] T. Bozóki, E. Prácser, G. Sátori, G. Dálya, K. Kapás, and J. Takátsy, "Modeling Schumann resonances with schupy," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 196, no. October, 2019.

[12] M. B. Perotoni, "Eigenmode Prediction of the Schumann Resonances," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 6, pp. 942–945, 2018.

[13] Y. Wang, Y. Zhou, and Q. Cao, "Study of elf propagation parameters based on the simulated schumann resonances," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 63–66, 2014.

[14] I. G. Kudintseva, S. A. Nikolayenko, A. P. Nickolaenko, and M. Hayakawa, "Schumann resonance background signal synthesized in time," *Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika)*, vol. 76, no. 9, pp. 807–825, 2017.

[15] D. I. A. Nickolaenko, P. Colin, "Hurst exponent derived from natural terrestrial radio noise in Schumann resonance band," *Geophysical Research Letters*, vol. 27, no. 19, pp. 3185–3188, 2000.

[16] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[17] A. M. Al-Samman, T. A. Rahman, M. H. Azmi, M. N. Hindia, I. Khan, and E. Hanafi, "Statistical modelling and characterization of experimental mm-wave indoor channels for future 5g wireless communication networks," *PloS one*, vol. 11, no. 9, pp. e0 163 034–, 2016.

[18] X. Chen, "Using akaike information criterion for selecting the field distribution in a reverberation chamber," *IEEE Transactions on Electromagnetic Compatibility*, vol. 55, no. 4, pp. 664–670, 2013.

[19] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 204–207, 2003.

[20] L. Yan, H. Wang, X. Zhang, M.-Y. Li, and J. He, "Impact of meteorological factors on the incidence of bacillary dysentery in beijing, china: A time series analysis (1970-2012)," *PLOS ONE*, vol. 12, no. 8, pp. 1–13, 08 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0182937

[21] A. P. Nickolaenko, "Modern aspects of Schumann resonance studies," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 59, no. 7, pp. 805–816, 1997.

[22] A. Kulak, J. Kubisz, S. Klucjasz, A. Michalec, J. Mlynarczyk, Z. Nieckarz, M. Ostrowski, and S. Zieba, "Extremely low frequency electromagnetic field measurements at the Hylaty station and methodology of signal analysis," *Radio Science*, vol. 49, no. 6, pp. 361–370, 2014.

[23] E. Greenberg and C. Price, "Diurnal variations of ELF transients and background noise in the Schumann resonance band," *Radio Science*, vol. 42, no. 2, 2007.

[24] P. D. Sardeshmukh, G. P. Compo, and C. Penland, "Need for caution in interpreting extreme weather statistics," *Journal of Climate*, vol. 28, no. 23, pp. 9166–9187, 2015.

[25] J. A. Parra, M. F. Ros, N. N. Castellano, and R. M. Salvador, "Techniques for Schumann Resonance Measurements: A Comparison of Four Amplifiers with a Noise Floor Estimate," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 10, pp. 2759–2768, 2015.

[26] M. Fernández-Ros, J. A. Gázquez Parra, R. M. García Salvador, and N. N. Castellano, "Optimization of the periodogram average for the estimation of the power spectral density (PSD) of weak signals in the ELF band," *Measurement*, vol. 78, pp. 207–218, jan 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224115005369

[27] E. J. Wagenmakers and S. Farrell, "AIC model selection using Akaike weights," *Psychonomic Bulletin and Review*, vol. 11, no. 1, pp. 192–196, 2004.

[28] A. Nickolaenko and M. Hayakawa, *Schumann resonance for tyros: Essentials of global electromagnetic resonance in the earth-ionosphere cavity*, 2014.