PhD Thesis

Statistical analysis of ELF signals' temporal segments measured in high sensibility station

Almería, April 2023

Manuel Soler-Ortiz

UNIVERSIDAD DE ALMERÍA

DEPARTAMENTO DE INGENIERÍA

DOCTORADO EN TECNOLOGÍA DE INVERNADEROS E INGENIERÍA INDUSTRIAL Y
AMBIENTAL

Statistical analysis of ELF signals' temporal segments measured in high sensibility station.

Análisis estadístico de segmentos temporales de señales ELF obtenidas en estación de alta sensibilidad.

Director:

Author:

Jose Antonio Gázquez Parra

Manuel Soler-Ortiz

Co-director:

Manuel Fernández Ros

Resumen

El espectro ELF es el rango de frecuencias contenidas entre 3 y 300 Hz. En este rango se encuentra la resonancias de Schumann; un fenómeno electromagnético presente en la parte baja de la atmósfera terrestre, concretamente en el espacio entre la baja ionosfera y la corteza terrestre que actúa como guía de onda. El impulso electromagnético que principalmente alimenta la señal resonante son los rayos producidos por las tormentas alrededor de todo el globo terrestre, aunque son muchos los factores que adicionalmente contribuyen a la señal ó influencian el estado de la cavidad resonante.

Desde que W.O. Shumann predijo teóricamente el fenómeno en 1952, las resonancias de Schumann han sido objeto de estudio debido al papel central que tienen en el circuito eléctrico global, así como por sus vínculos con el clima terrestre. No obstante, pocos son los estudios que se centran en analizar la señal en el tiempo, ya que la investigación de la resonancia se ha basado principalmente en la descomposición espectral en frecuencia de la misma.

Existen en el campo de las telecomunicaciones metodologías que estudian los aspectos temporales de señales electromagnéticas. Dichas metodologías surgen de la *"Mathematical Theory of Communication"* de Claude Shannon, que trabajan con la información de una señal tratándola como una variable estadística. En esta tesis se recurren a métodos estadísticos para analizar cinco años de datos de las Resonancias de Schumman capturadas en el observatorio de señales ELF (Extremely Low Frequency) que tiene el grupo de investigación TIC-019 de la Universidad de Almería en la Sierra de los Filabres.

De este trabajo surge una nueva metodología para trabajar con fenómenos electromagnéticos naturales, permitiendo segmentar los registros de las resonancias de Schumann y clasificar cada segmento bajo la distribución estadística que mejor se ajusta a los datos de dicho segmento. Esta metodología presenta un modo adicional de extraer información de los registros ELF que complementa los estudios en frecuencia comúnmente utilizados.

El análisis de los resultados de aplicar esta metodología muestra como cada distribución está relacionada con estados identificables en la señal, consiguiendo de forma efectiva una clasificación automática que permite discernir los eventos contenidos en cada segmento analizado.

Adicionalmente se ha desarrollado una técnica para cuantificar la variabilidad de la señal en el tiempo basada en la distancia de Kolmogorov-Smirnov. Ésta ofrece una forma de controlar la estabilidad de la señal, permitiendo seleccionar segmentos específicos para estudios que requieran alta precisión. También revela evidencia de una nueva forma de extraer información sobre la actividad electroatmosférica a partir de la señal.

Abstract

The Extremely Low Frequency (ELF) spectrum is the range of frequencies contained between $3\,\text{Hz}$ and $300\,\text{Hz}$. In this range, the Schumann resonance (SR) can be measured; an electromagnetic phenomenon that manifests in Earth's lower atmosphere, with the space between the ionosphere and Earth's crust acting as a waveguide. The main source of the resonant signal is lightning activity distributed all over the globe, although many other factors contribute to the signal or influence the resonant cavity.

Since W. O. Schumann predicted the phenomenon theoretically in 1952, the Schumann resonance has been studied for its central role in the global electric circuit and its links with Earth's climate. Nonetheless, the number of modern studies that analyze the temporal component of the signal are scarce, relying primarily on spectral decomposition to carry on the research.

In the field of telecommunications, a number of methodologies that study the temporal aspect of electromagnetic signals are applied. These methodologies span from Shannon's *"Mathematical Theory of Communication"*, that deal with the information content of the signal as a statistical variable. On this thesis, statistical methods have been applied to analyze over 5 years of Schumann resonance data captured by the Extremely Low Frequency observatory that research group TIC-019 of Universidad de Almería has installed in Sierra de los Filabres.

A new methodology to work with natural electromagnetic signals emerges from this work, allowing to automatically segment the records and classify each segment under the statistical distribution that best fits the chosen data segment. This methodology represents an

additional way to extract information from the ELF records, complementing the commonly used frequency based approaches.

Analyzing the methodology results shows that each distribution is related to identifiable states of the signal, effectively performing an automatic classification that allows to discern the events within the segments.

A technique to quantify the signal variation over time based on Kolmogorov-Smirnov distance was also tested, showing a way of measuring the signal stability to select specific segments in high precision studies and hinting at a different, automated way to measure lightning activity changes from the signal.

Agradecimientos

En primer lugar quiero agradecer a los profesores Manuel Fernández Ros, Rosa María García Salvador, Jose Antonio Gázquez Parra, y Nuria Novas Castellano su labor como mentores en mi formación, y por brindarme la oportunidad de abrirme paso en el mundo de la investigación. No menos importante han sido mis compañeros en este camino que ha sido nuestra etapa como doctorandos, Carlos Cano Domingo y Francisco Portillo Rodríguez. Por estar siempre dispuestos a escuchar, por compartir sus ideas y puntos de vista, y sobre todo por llevarme la contraria siempre que era necesario, gracias.

La motivación y fuerza para llegar hasta aquí me la han dado mi mujer y mi hijo. Carmen Ruiz Peña me lleva acompañando durante más de la mitad de mi vida y, como siempre, ha estado ahí para apoyarme y hacerme reír. Manolo Soler Ruiz vino al mundo durante esta etapa de mi vida, haciendo de mi mundo un lugar mejor. Verlo aprender y crecer cada día ha sido una motivación para intentar yo hacer lo mismo.

Si he llegado hasta aquí ha sido gracias a la energía que me han aportado mis amigos. Las barbacoas épicas, los juegos de mesa, las conversaciones interesantes, y los chistes malos me permitían desconectar para volver con las pilas cargadas. Lourdes, Adriana, Carlos, Jose A, Pepe, Abel, Isra, Taka, Gon, y sobre todo Carmen... Gracias por todos los buenos ratos.

No menos importante es mencionar a mi padre, Manuel Soler Cruz, y a mis hermanos, Jose y David. En el contexto de mi doctorado, ellos son más que mi familia; son las personas que me han hecho lo que soy hoy día, y es gracias a ellos y a nuestro afán común por comprender el mundo y discutir sobre él que he podido llegar hasta aquí.

Dejando lo mejor para el final, un agradecimiento muy especial para mi madre, María Teresa Ortiz Vázquez. Aún recuerdo su pasión por saber y su interés por transmitirlo. Ella me enseñó que se pueden formar cristales en el agua, cómo hacer un motor con una pila, una pajita, horquillas y cobre, los efectos de la electricidad estática, la electrólisis... En definitiva, me mostró lo fascinante que era el mundo, y que siempre hay algo nuevo que aprender. Defender esta tesis me hará Doctor, pero la persona que me hizo Investigador, fuiste tú. Gracias mamá.

Contents

List of Figures

List of Tables

List of acronyms

AC   *Alternate Current.*

ACF   *Autocorrelation Function.*

ADC   *Analog to Digital Converter.*

AIC   *Akaike Information Criterion.*

BIC   *Bayesian Information Criterion.*

BPF   *Band Pass Filter.*

CDF   *Cumulative Distribution Function.*

DC   *Direct Current.*

EW   *East-West.*

ELF   *Extreme Low Frequency.*

FFT   *Fast Fourier Transform.*

HPF   *High Pass Filter.*

KS   *Kolmogorov-Smirnov.*

MA   *Moving Average.*

MLE   *Maximum Likelihood Estimates.*

NA   *Not Available.*

NS   *North-South.*

PDF   *Probability Density Function.*

QoF   *Quality of Fit.*

RMS   *Root Mean Square.*

SR   *Schumann Resonance.*

TLE   *Transient Luminous Event.*

VLF   *Very Low Frequency.*

Chapter 1

Introduction and justification

Throughout this chapter, the premises of this thesis are presented. First, the state of the art of the *Schumann Resonance* (SR) is reviewed, starting with a short summary of the resonant signal and its basics in Section 1.1, followed by a chronological inspection of SR's most important works (Section 1.2). In the last two sections of this introduction, the exposition of this thesis' theoretical setting to be used as starting point (Section 1.3) and a brief presentation of the mathematical tools that serve as its framework (Section 1.4).

1.1. Summary of Schumann resonance's research

The SR is an electromagnetic resonant phenomenon inside the Earth-ionosphere cavity, which is the portion of the atmosphere contained between these two layers (Figure 1.1, extracted from Perotoni (2018) [1]).

As an electromagnetic signal, it belongs to the *Extreme Low Frequency* (ELF) range of the spectrum [7], with a fundamental mode of 7.8 Hz, and measurable modes up to the sixth, whose central frequencies are around 14 Hz, 20 Hz, 26 Hz, 32 Hz and 40 Hz [3]. Difficulties for clearly measuring modes beyond the sixth come from their low intensity in relation to the rest of electromagnetic signals in that frequency band, especially against the European electric grid signal, centered on 50 Hz. When measuring the SR magnetic components, an
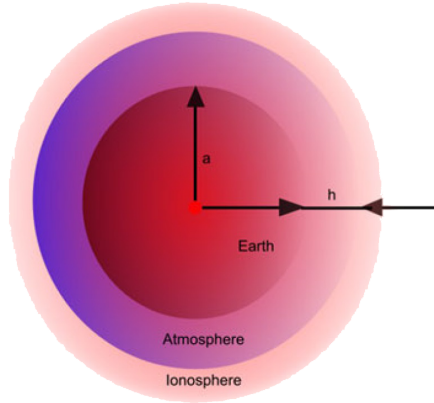
Figure 1.1: Simplified diagram of the Earth-ionosphere cavity with *a* being Earth's radius and *h* the ionosphere's height, from Perotoni (2018) [1]

average peak-to-peak value of 8 pT is registered [8] (commonly called the ELF background noise), although higher peak values appear due to electromagnetic activity inside the cavity.

The concept of conducting concentric spheres as a resonant cavity was under study at the beginning of the 20th century. One of the first mentions to the problem of electromagnetic oscillations in a cavity between two concentric spheres is from Joseph J. Thompson, who dealt with the problem in a book published in 1893 [9]. Winfried Otto Schumann (Figure 1.2, from Besser (2007) [2]) was the first to apply the problem to the Earth-ionosphere cavity and mathematically predict the phenomenon, reason why the signal bears his name.

He published his findings in 1952, on which he calculated the eigenfrequencies using a simplified resonant system model (e.g. he considered Earth and the ionosphere perfect, lossless conductors) considering lightning as the system's main excitation source [10]. His well-known equation, considering the cavity an ideal resonator and further simplified using the fact that Earth's radius is greater than the ionosphere height can be seen in Equation 1.1.

$$f_n = \frac{c}{2\pi a}\sqrt{n(n+1)} \tag{1.1}$$

As for the rest of variables, c is light speed in the vacuum (299,792 $\frac{\text{km}}{\text{s}}$) and n is the resonance mode number. The frequency values predicted by this theoretical model,

Figure 1.2: Photograph of W.O. Schumann from Besser (2007) [2].

being 10.6 Hz, 18.3 Hz and 26.0 Hz for the first three modes, differ considerably from the experimental values mentioned above. This is due to the properties of the resonator (mostly the ionosphere characteristics, being far from a perfect conductor, and its anisotropic conductivity profile [11]) that the ideal model does not take into account. In order to produce values in agreement with the experimental records, the crucial factor for the model to consider is the variation of the conductance of the ionosphere as a function of height [12].

As stated before, the signal originates from global lightning activity, being an integral part of the global electric circuit as it sets its non-continuous component [13]. The equations that describe each component of the resulting electromagnetic field are presented below. Namely the vertical electric field equation (Equation 1.2), and the *North-South* (NS) and *East-West* (EW) components of the horizontal magnetic field (Equation 1.3 and Equation 1.4 respectively. The present formulation of these equations were published in Nickolaenko (1997) [14].

$$E(t) = \sum_{k=-\infty}^{\infty} A_k \cdot g(t - t_k) \qquad (1.2)$$

$$H_{NS}(t) = \sum_{k=-\infty}^{\infty} A_k \cdot \hat{g}(t - t_k) \cdot \sin(B_k) \tag{1.3}$$

$$H_{EW}(t) = \sum_{k=-\infty}^{\infty} A_k \cdot \hat{g}(t - t_k) \cdot \cos(B_k) \tag{1.4}$$

In the equations above, $g(t)$ is the waveform of the electric component and $\hat{g}(t)$ its magnetic counterpart. A_k is the amplitude of the k-th lightning, and B_k is its source bearing in relation to the observer, meaning the wave arrival azimuth measured clockwise. Therefore, as lightning strikes around the globe, the resonant cavity produces a signal whose response is dependent on the summation of every pulse's amplitude and, in the magnetic component, dependent on source-observer distance and orientation [14]. Due to the atmosphere's low attenuation in the ELF band (around $0.5\,\mathrm{dB}$ per $1000\,\mathrm{km}$) [15], it is measurable all around the globe.

The measurements are carried out at stations isolated from human and industrial activity to minimize interference. The electric antenna's design of reference for ELF measurements is the so-called ball antenna. A detailed design was proposed in 1966 by Ogawa *et. al.* [5, 16] and is still used for more modern experiments [17]. Magnetic antennas are solenoids with many turns winded on a ferromagnetic core of high permeability [18], although there are working designs with air cores [19].

Despite being a resonant phenomenon, they are not stationary by nature. The high variability of all different media involved in the resonant cavity (from ionospheric conductance and height to atmospheric composition and lightning activity variations) changes the nature of the SR in many different ways. There are many patterns involved and identified in the electromagnetic signal, such as the daily ionospheric changes [20], the seasonal variations introduced by the changes in sunlight's angle of incidence and power [21] and, as published just recently, those introduced by the variations of the solar cycle [22]. These factors influence in turn the thunderstorms that happen around the globe, introducing patterns in global lightning activity as well [23]. Other patterns directly related with recurrent climatic con-

ditions of Earth have been measured, such as El Niño y La Niña [24] or the Madden-Julian oscillation [25, 26].

Nevertheless, the random nature of lightning activity - separate lightning events being usually modelled using the Poisson distribution [27, 28] - gives the signal a stochastic component that is difficult to resolve when analyzing the fine structure of the spectrum. All the variables involved in the resonant process are still under study; considering the number of variables involved and their individual complexity, make the production of a reliable model a daunting task.

Since their ascertainment and experimental verification, their scientific interest has been steadily growing. This is mostly because they occupy a major role on Earth's electrical circuit [29] and, by extension, many climate and atmospheric processes [30, 31, 32, 33]. That being said, the different breakthroughs achieved in the field are testimony of the untapped potential of this research.

1.2. Historical background

This section is focused on the most significant works from the discovery of the signal onwards. A detailed presentation of the works and scientists that were paramount on this discovery along with the most important scientific advances that stemmed from it can be found in Besser (2007) [2].

1.2.1. Preliminary studies on Schumann resonance

The idea of conducting layers on the atmosphere springs at the end of the 19th century and is brought up sporadically all along the century. There are records of many different scientists considering the idea, names such as Carl Friedrich Gauss, William Thomson (Lord Kelvin), Joseph J. Thomson, and Nikola Tesla [2]. It is at the beginning of the 20th century when concrete evidence is found by George Neville Watson, in the form of a disagreement between its calculations of diffraction of electric waves around Earth and experimental results [34].

He expresses in the same article how the existence of a "perfect reflector" around Earth may justify the obtained difference.

For a couple of decades, evidence about the existence of the ionosphere keeps piling up, and eventually, the resonant characteristics of the Earth-ionosphere cavity were discussed. Despite that, W. O. Schumann was the first to deeply study the theoretical aspects of the problem, publishing his findings in 1952 [10, 35, 36]. These articles and some others (see [2]) published between 1952 and 1954 can be summarized on three main topics.

1. A spherical cavity is able to propagate electromagnetic waves, its resonant frequencies depending on the geometry of the cavity.

2. In the case of our planet, the cavity is delimited by Earth's crust and the lowest layer of the ionosphere, with the atmosphere in between acting as the medium of transmission, and the whole system acting as a waveguide.

3. The electromagnetic impulses that would source the signal would be lightning discharges all over the globe.

Along with his doctoral student Herbert L. König, Schumann also started some experimental work [37]. Given Schumann's theoretical background, he eventually left the experimental measurements to his student who published his findings in 1959 [38]. Nonetheless, their attempts to measure the signal they predicted theoretically were unsuccessful.

The topic also attracted the attention of other scientists, but in the same fashion, studies performed at that time were not conclusive. It is in 1960 that Balser and Wagner develop the instrumentation (a vertical whip electric antenna with cascaded low-pass filtering down to $200\,\mathrm{Hz}$) and techniques to be able to measure the ELF spectrum. Their first attempt yielded no results since they focused on the frequency band between $50\,\mathrm{Hz}$ - $100\,\mathrm{Hz}$ [39]. Nonetheless, in the same year, they publish another work where a clear spectral representation of the first five modes of the resonance can be seen [3]. The latter paper is generally credited as the first scientific work to experimentally confirm the Earth-ionosphere resonance. The obtained spectrum can be seen in Figure 1.3 extracted from Balser and Wagner (1960) [3].
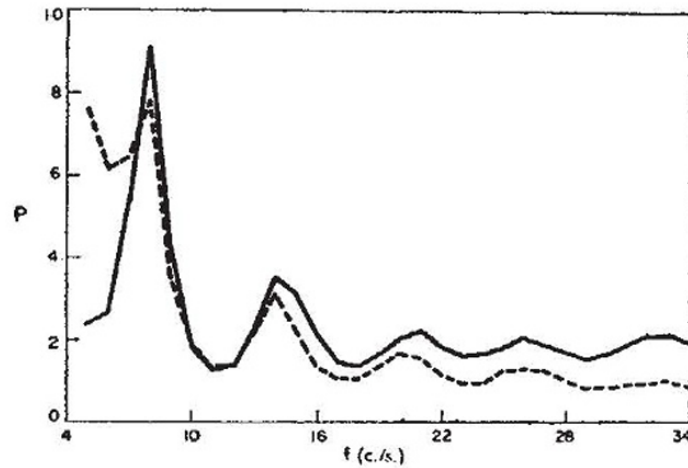
Figure 1.3: Spectrum of the measurements made by Balser and Wagner in 1960, from [3]. Linear vertical scale, The full line presenting the day-time record, and the broken line the night-time record.

1.2.2.   Signal propagation and ionospheric conductivity

One of the first interests of the scientific community at the time is modeling, up to some extent, the signal and its dependence on lightning activity and ionospheric parameters. At this stage of knowledge, most models are focused on extracting the resonant frequencies of the different modes by means of different assumptions and experimental measurements. Given the low attenuation of the ionosphere, the assumption of both Earth and ionosphere working as perfect conductors holds in most cases, but as stated in Section 1.1, the conductivity profile of the ionosphere must be accounted for. The different conductivity profiles established by other studies did not agree completely, achieving different degrees of complexity, but also different results, especially for the lower part of the atmosphere [40, 41, 42]. Studies that account for ELF attenuation rates served to refine the available ionosphere conductivity profiles [43] and so in turn the profiles obtained by different researchers were tested as well.

These early studies of the SR are thoroughly reviewed in Madden and Thompson (1965), an article that gives perspective to the theoretical works performed to date and evaluates them by comparing their results with the available observations. A special interest is dedicated to take into account the ionosphere asymmetries, incorporating them to the theoretical models [44]. The theoretical framework developed in this review is still a reference in the field, with

7

many researchers recurring to Madden and Thompson models to use them as cornerstone for their work.

Another remarkable example of this line of work is Jones (1967), in which the author uses different ionospheric profiles extracted from other papers to model the resonant frequencies of the cavity. His results, considering a single source of lightning activity, show how the typical two-layered profile for the ionosphere is in agreement with experimental data, without the requirement of more complicated models for the ionosphere [45]. Another important finding on this topic that appears in different works of the time is the low importance of the conductivity of the atmosphere on its lowest layer, from the ground to a height of 40 km.

By understanding the relationship between ionosphere and the SR, a general picture of the resonant problem mechanics could be constructed. From this point onwards, more works start to present extensive measurements in order to characterize the short-period variations observed in the resonant modes, such as the one published by Røed Larsen and Egeland (1968) [46]. As time goes by more detailed works about the subject are published, exploring the variations from different geographical locations, and submitting evidence for the causes of these variations. An example is the work of Galejs (1970), on which the SR's frequency and intensity variations are related to the propagation characteristics of the ionosphere [47], detecting differences between day and night propagation and linking the intensity variations of the fundamental mode to the thunderstorm centers' activity.

In the topic of modelling of the ionospheric conductivity profile, a very influential paper is the one presented by Mechtly, Bowhill, and Smith (1972). The SR are not the focus of its study, but their systematic investigation of the lower ionosphere and the subsequent data serve as the basis for many SR-ionosphere publications as time goes by. They measured the electron concentration in the ionosphere along the Sunspot cycle No. 20 by means of Langmuir probes attached to rockets. A total of ten rockets were launched; five during the Sun's minimum magnetic activity period (1964-1965) and the other five during its maximum magnetic activity period (1968-1969). The resulting measurements provided experimental evidence to work with accurate models of the ionospheric conductivity profile [48].

In Ogawa and Murakami (1973), experimental observations of the electrical component of the SR served as a baseline for one of the first models to consider multiple sources of lightning. Again, the model is calculated using different conductivity profiles (three, in this case), with all of them providing similar results and close to the averaged experimental values obtained [11].

Further models are developed and tested against experimental data, each considering different hypothesis and with different advantages. Tran and Polk publish in 1979 two papers in which the effects of ionosphere conductivity on the SR are exhaustively evaluated.

First, Tran and Polk (1979a) is focused entirely on the ionospheric conductivity profile. A three layer stratified ionosphere model is chosen, assembling three profiles from data extracted from the available bibliography. They compare the three profiles by calculating the losses at each of the layers and the resonance frequency under each profile. Also, the effects of conductivity for each of the three layers on the profile with the best results are detailed [49].

Following the methodology of the previous article, Tran and Polk (1979b) present a method for producing conductivity profiles from the SR records by means of the location of active thunderstorms relative to the observatory. After the methodology is explained, conductivity profiles are calculated from a set of records in three different days. Both the calculation of source-observer distance and the resulting ionospheric conductivity profiles are compared with experimental data and discussed in length. The article highlights the independence between the average conductivity in the 0 km to 40 km and the profile between 60 km to 75 km in the applied methodology's results. In consequence, it is required to determine the conductivity in one of these two regions to remove the ambiguity and obtain an unique conductivity profile [50].

1.2.3. Extremely low frequency sensors

Atmospheric noise measurements carried before Schumann theoretical formulation were mostly focused on the *Very Low Frequency* (VLF) part of the spectrum. The studies were

focused on characterizing noise sources in this frequency band to improve the performance of the wireless communication systems of the time. These recordings were performed using whip antennas, much like the ones present in the transmitters and receivers for the mentioned systems [51].

As the interest to take electroatmospherics measurements shifts to lower frequencies, the instrumentation is adapted for the measurements but there are no changes in the design. For the set of measurements carried by Balser and Wagner, a 8 m vertical whip was used at first [39]. In their second and successful attempt, they relied on a metal tower with a height of 40 m [3].

With SR studies drawing the attention of more scientists, certain innovations and novel instrumentation are proposed to perform the measurements. One of the most influential designs comes from Ogawa *et al.* (1966), on which their electric field measurement using a ball antenna is shown, with the main advantage of filtering most of the natural noise [5]. Due to its electromagnetic sensing properties, this design became an standard of sorts in ELF measurements. A working model of this antenna can be seen in Figure 1.4 in Mitspe Ramon Observatory (Israel), obtained from from Nickolaenko and Hayakawa (2014) [4] and ceded to the authors by Dr. Colin Price.

It is also in this work where the patterns showed by ELF signal are classified as its base state (background noise), N-bursts, and Q-bursts (sudden spike in amplitude, being N or Q depending if they are followed by VLF noise or not) and flashes (signal saturation due to high discharge in the sensor proximity). The original image where these three kinds of signals were identified can be seen in Figure 1.5, extracted from Ogawa *et. al.* (1966) [5]. This classification is still used nowadays, with Q-bursts being a major topic of interest.

No trace of magnetic measurements of natural sources has been found in the available bibliography prior to the SR formulation. Applications of magnetic sensing were found in different fields, mostly in aircraft instrumentation, along with theoretical analyses to understand the underlying principles and behavior. An example of the latter is the paper from Rumsey (1953), which analyzes traveling wave slot antennas, proposing approximate

Figure 1.4: Ball antenna for electric field measurements in the ELF range in Mitspe Ramon Observatory (Israel), extracted from Nickolaenko and Hayakawa (2014) [4].

formulas for field configuration that is characteristic for these technologies, as well as for the complex propagation constant [52].

Regarding magnetic antennas' applications, Bolljahn and Reese (1953) bring up a number of analyses along with experimental procedures to test small antennas, such as the ones installed in aircrafts. The problem stems from the antennas size; being smaller than the wavelengths they are supposed to measure, their performance must be evaluated, and the authors recommend techniques to carry such evaluations [53]. In Hemphill (1955) an improvement for magnetic radio compasses in high speed aircrafts is proposed, eliminating the precipitation static problem (which worsens with speed) [54].

It is after the first detection of the SR that measuring systems equipped with magnetic antennas began to be deployed. To properly measure the signal given its low frequency, coils
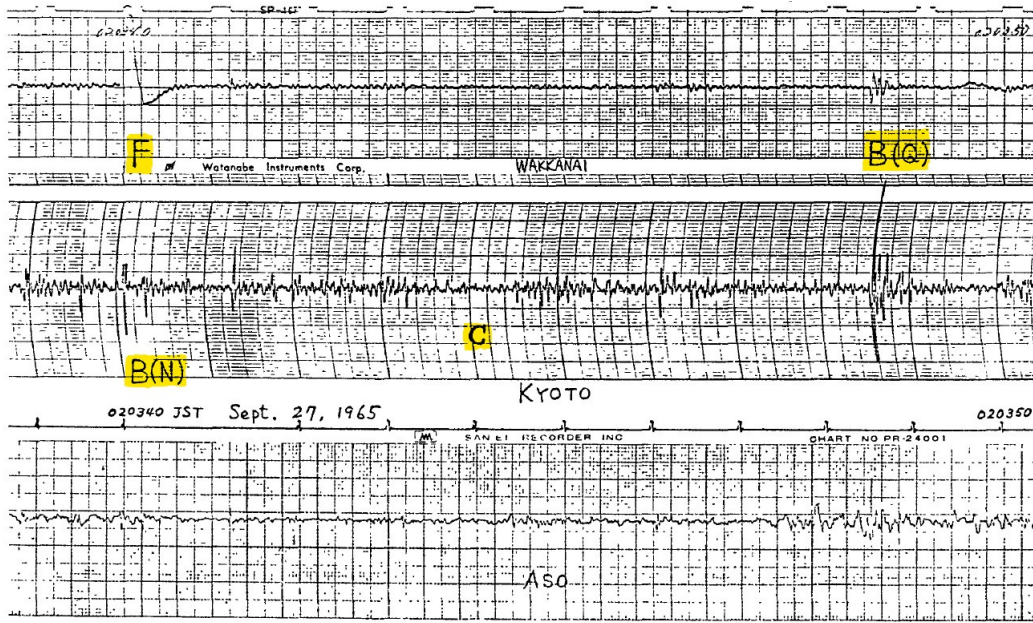
Figure 1.5: Original *Schumann Resonance* records where the three accepted kinds of ELF noise were identified, from Ogawa *et. al.* (1966) [5], with *F* marking an ELF flash, *B(N)* and *B(Q)* identifying the two kinds of ELF bursts, and C showing a section of ELF background noise.

with a high number of turns (usually over 100 000) are required. The magnetic antennas used at the time relied on ferromagnetic cores to increase the magnetic flux captured by the antenna, although varying in shape and size. Polk and Fitchen (1962) use for their measurements two receiving coils each with 249 000 turns of number 38 isomel insulated wire, with a length of merely 6 cm [55]. Gendrin and Stefant (1964) report using five separate coils, each with 2000 turns and a ferrite core [56]. The short length of the coils described in these works strikes as odd, given how the current trend is to build longer coils, over 80 cm long [19, 57]. In the following years the interest for the topic increases, and so does the number of ELF observatories. Most of the works from the 70s onwards include magnetic field measurements, and some of them describe the sensors used [58].

Despite the advances in magnetic materials, magnetic antennas' effective parameters have not changed dramatically. A referent on the field is still a book published by Burrows in 1978 [18], as modern implementations follow the same principles. The experimental measurements' equipment described in the published works from there on offer little or no improvement in the sensing part of the equipment. There are only upgrades in the processing

side of the instrumentation. As technological progress is applied into the experimental sciences, the analog storage is shifted for more efficient digital techniques [59].

As an example of a modern implementation, Figure 1.6 shows the sensor research group TIC-019 developed and installed in the ELF observatory in Sierra de los Filabres (Almería - Spain), in 2012.
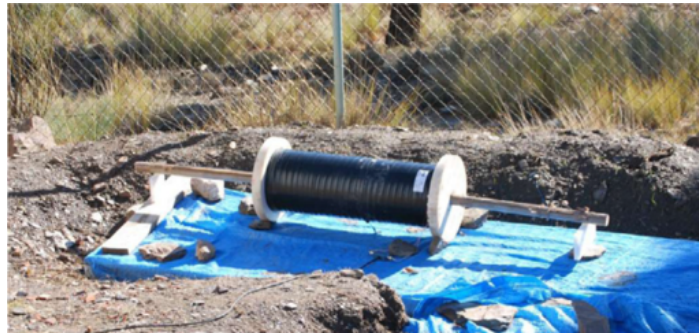
Figure 1.6: Coil sensor from research group TIC-019, University of Almería, currently in service at the observatory in Sierra de los Filabres (Almería - Spain).

1.2.4. ELF-VLF noise amplitude distribution and long-distance communications

The low attenuation at this frequency band invited the possibility of using it as the channel for long distance communications. This possibility attracted the interest of major powers to guarantee long-distance communications with mobile assets, specifically submarines.

Given their focus on radio communications, they deem the sensable electromagnetic activity occurring in the atmosphere as atmospheric noise as a whole, and not paying special attention to the SR. Despite that, it is relevant to bring them up since the articles that perform statistical analysis over the ELF band belongs to this line of research.

Natural radio activity in the atmosphere is an important parameter in wireless communications. When the chosen band is within ELF range the modelling of natural electromagnetic activity in this band is of paramount importance; its continuous presence in the transmission channel gives it the role of noise floor for the signals transmitted. Studies contemporary

to the SR theoretical formulation are mostly focused in characterizing atmospheric noise in the VLF range, such as Horner and Harwood (1956) in which the noise properties between 10 kHz to 35 kHz captured with a narrow band sensor (300 Hz) are studied. The random, high amplitude peaks appearing in the register difficult the standard analysis using *Root Mean Square* (RMS) values of the noise envelope or average voltage values. In consequence, a methodology is presented to define statistically the distribution of peaks amplitude within the envelope and the amplitude of the envelope itself, finding relationships with the lognormal distribution [60].

In Watt and Maxwell (1957), a detailed study about the envelope of the electric field captured in the presence of lightning is presented. By working over a wide range of frequencies (from 1 Hz to 100 kHz) while considering distance between the lightning stroke and the receiver, and the attenuation factor over the range of analyzed frequencies, they calculate the spectra of average lightning discharges under different conditions. They back up their calculations by extracting the lightning spectrum from the amplitude distributions of the measurements at different central frequencies with a bandwidth of 140 Hz. The similarity between calculated and measured spectra is one of the first evidences published of lightning presence in the VLF spectrum [61].

A more formal approach to the statistical distributions of the two phenomena separated by Horner and Harwood (distribution of peak values and distribution of envelope values) is carried at the beginning of the 60s. In Furutsu and Ishida (1961), atmospheric noise is defined as the superposition of independent, randomly occurring Poisson noise. They work into deduce general amplitude distributions and compare them with atmospherics' data [62].

In the same fashion, Beckmann (1963) works further in the amplitude distribution of atmospheric noise. Specifically, the distribution of the peak value of atmospherics and the amplitude distribution of the signal itself is analyzed, concluding that the amplitude distribution in its cumulative form is, approximately, the sum of a vector with lognormally distributed amplitude and a Rayleigh distributed vector [63].

A work that serves as bridge between the two topics explored in this section is the technical report by Evans (1969), in which ELF noise is thoroughly analyzed. It steps up from the previously cited articles and brings them together by characterizing the noise of the wideband spectra (From 1 Hz to 300 Hz) and clearly showing its non-Gaussian nature. Nonetheless, the SR, as happens with the rest of the articles in this section, are only mentioned as a curiosity and part of the noise belonging to the lower part of the spectrum [64].

The previous report is redacted in the context of Project Sanguine, the first attempt of the US government to develop a system able to transmit short messages in the ELF range. The technical advantage of the system is water permeability to signals under 100 Hz, which would allow the Navy to contact submarines while submerged so they could emerge to the surface and receive instructions via communications systems with bigger bandwidth. The project was released in 1972 and received criticism from both the public and the academy [65].

Besides health and environmental related issues, one of the main concerns of the project was the amount of power required to transmit the signal. Another study was commissioned by the U.S. Navy to refine the required instrumentation and try to reduce the power involved [66]. In this case, the instrumentation object of study were the antennas carried by the submarines to improve the receiver's signal to noise ratio. A better receiver implies less transmission power needed, and since 1 dB change of required transmitted power implied saving millions of estimated cost, it was something entirely desirable [66].

According to their observations, they state that ELF noise can be modeled as a Gaussian background process added to an impulsive process generated by a filtered Poisson point process [66]. The former account for regular thunderstorm activity, while the latter is filtered in a way that the resulting effect had the duration and amplitude similar to the bursts measured in the atmospheric noise. Based on the ample number of tests performed, they recommend a system with a non linear processor with a series of filters that clipped the undesirable spikes caused by either high powered lightning strikes or close thunderstorm activity [66]. Despite the technical note being released in 1972, the work does not reach the community until two

years later, where Evans and Griffiths publish their findings and the receiver architecture in a scientific journal [67].

Shortly thereafter, Burrows and Niessen (1972) studies in detail the design of the transmitter, while carefully considering the rest of contributions to the field. On their 1972 article they discuss questions about antenna design, transmitter-receiver performance, and necessary transmission power and ideal transmission frequency, concluding that the best working frequency for this kind of system, power wise, is 75 Hz [68].

The project, initially designed to withstand a nuclear attack with many redundant transmitters and consequently requiring a huge surface, suffers many downscalings over time, finally being implemented as Project ELF in the 80s although criticism do not stop altogether [69].

The tale of Project Sanguine is briefly presented as an example of ELF communications project, quite relevant to the work developed in this thesis because of the noise analysis in the ELF range. The drawn conclusions about the statistical nature of the noise supports the work performed, and its relevance comes from the fact that the number of similar analysis conducted from that moment onwards are scarce.

Regarding long distance communications, there are other examples of engineering projects that rely on the Earth-ionosphere properties, such as the work presented by Ferraro *et. al.* in 1989 to find ways to modulate ionosphere channels by heating it with ground based radiation [70].

1.2.5. Schumann resonance and lightning activity

Another main concern of the time is to better understand the relationship between global lightning activity and SR. In 1969, Charles Polk published an article exploring the mentioned relationship. It is one of the first attempts in the literature to use the resonant phenomenon to locate the focal points of lightning activity since the accused diurnal variations in intensity make it hard to accept the hypothesis of daily lightning activity being close to an average value. Therefore he focuses on the parameters independent from ionosphere conditions,

such as electric to magnetic field ratios, which as can be seen in Eqs. 1.2, 1.4, and 1.3 depend only only on source-observer distance. He also proposes techniques to locate the active thunderstorm cells based on ratios involving power spectrum magnitudes at adjacent resonant peaks [71].

Jones and Kemp (1970) is focused in the high amplitude events appearing from time to time in the registers, such as the Q-Bursts reported by Ogawa *et. al.* [5]. They work on the hypothesis of these pulses being produced by especially powerful lightning discharges. At first they compute the frequency spectra (both electric and magnetic) for a single source (lightning modelled as an instantaneous electrical dipole) at different distances from the measurement station. Then they locate high intensity pulses in the experimental ELF records and work over their spectra, computing their distance by finding the best fit for the experimental spectra among the theoretically computed ones. The authors acknowledge their own limitations pointing out how the model used is very simple and how the background noise limits the possibility of expanding their experimental methodology to other cases. Still, their results are strong enough to be evidence of lightning being the source of the studied ELF pulses, being one of the first works to link these two events together [58].

Following up the previously cited article, in Jones and Kemp (1971) a hundred ELF transient events of high amplitude both in the frequency and time domains are analyzed. They rely on amplitude distribution to select the events more appropriate to be studied under the methodology proposed in [58] and compare the samples with an estimation of the background ELF noise. Then, they work with the chosen ELF events in the time domain, showing by comparison between experimental and modeled sources how the event waveform is characteristic of the distance where the event happened. The reason being the interference between the short path and the long path of the initial impulse towards the observatory and the subsequent pulses as the signal circumvents the globe [72]. The study brings up to a fact the conclusions exposed in the previous article, being able to obtain average peak current moments and average moment destroyed by these great discharges.

Back to a more generalized approach, Clayton and Polk (1976) presents an approach supported by another work published by the latter author in 1969. With almost a year

of continuously recorded data, they extract properly corrected hourly variations of source-receiver distances and on the ionospheric conductivity profile as consequence for changes in solar activity. An estimation of the absolute global lightning intensity in terms of charge moment squared per second is delivered as the article's main result [73]. Although there are some articles on the topic [74, 75], this line of research does not proliferate until 1990.

1.2.6.   Significant studies at the end of the century

At the start of the 80s, books documenting thoroughly all the theoretical aspects of the signal are published, along with recommended instrumentation and experimental procedures used since its discovery to record and analyze the electromagnetic resonance [76, 77].

The signal is still under research and interesting results are produced, but the studies published from the start of the 80s to the early 90s are mostly focused on further characterizing the SR and the elements of the Earth-ionosphere cavity. There is also a shift towards observing the effects of specific electromagnetic events, and how they are factored into the signal. A few works from this time will be briefly introduced, so the reader can grasp the topics treated by the most influential papers.

In Cannon and Rycroft (1982) SR measurements during a period of high magnetic activity were carried away with the intention of understanding the signal's frequency variation during sudden ionospheric disturbances. They identify significant frequency variation during two electromagnetic events, well above the expected diurnal variation, especially in the first and third modes [78].

Sentman (1983) present a theoretical work in which a detailed analysis of variable ionosphere conductivity and SR frequency variations is performed. This is done so by modifying an exponential ionospheric conductivity profile, chosen to resemble the one presented in [48]. The modifications are introduced as additional parameters in the model that allow to introduce perturbations at different heights. Over the disturbed profiles, the author calculates the resulting frequencies and discuss the effects of the different perturbations [32].

Detailed studies about the fine structure of the spectrum are published as well; one of the most cited is Sentman (1987), in which differences in frequency over 0.5 Hz between the magnetic part of the signal captured by EW and NS are discussed. As it is said in the article, the appearance of these differences rule out the possibility of SR being linearly polarized. They describe the mathematical procedure applied to analyze the signal's ellipticity and conclude that the magnetic field displays, for most of the of the day, a considerable elliptical polarization. They also observe the polarization is left-handed at local nighttime, and right-handed during local daytime, shifting dramatically back and forth during sunset and sunrise. The possibility of the nighttime ionosphere magnetization to be responsible of the effect is argued, since it affect differently the phase speeds of propagating waves, depending on whether the propagating front goes either eastward or westward. It is also considered that the anisotropies and inhomogeneities of the cavity could generate enough transverse electric components to distort the magnetic signal due to cross-coupling [59].

More results of this research are published in Sentman (1989), on which Q-bursts are studied through maximum entropy spectral techniques to obtain the instantaneous impulse response of the Earth-ionosphere cavity. Prior to the analysis, they establish a new coordinate system for the hodogram of the conjoined EW and NS magnetic field through a maximum variance transformation. The new coordinate system is chosen so one axis matches the maximum variance value, which should coincide with the central multiplet of the wave, and with the side multiplets appearing in the other axis, as the minimum variance. The frequency differences under the new coordinate system are, on average, 1.6 Hz for the events analyzed; three times greater than the difference found in the previous article. A detailed discussion about the origin of this and other inconsistencies is provided, reasoning which irregularities may be due to the anisotropic nature of the ionosphere, and the importance of mode coupling for both the observed line splitting and the measured elliptical polarization [79].

Sentman and Fraser (1991) compares the measurements of two synchronized observatories; one in Table Mountain, California, and the other in Northwest cape, Australia. The differences between the measurements are accounted for; a corrective function is proposed under the reason that the power received by each station is the product of the global signal

and the effects of local time, represented as a Fourier series over integral multiplies of the terrestrial rotation frequency. According to the global theory of the SR, the same source can be assumed for a specific event simultaneously measured in both observatories. These two assumptions allow the authors to calculate the correction coefficients that, once applied to the measurements, show a remarkably high correlation index. The correlation coefficients of both observatories increase from 0.51 to 0.70 for the measurements taken during September 1989 and from 0.39 to 0.82 for the measurements taken in April 1990. It is argued then how the local time effects are due to the variations in ionospheric height, and it is demonstrated by extracting the local height variation from the coefficients extracted for the corrective function [20].

In the early 90s the interest in SR grew again, mostly because of certain works revealing the untapped potential of the electromagnetic signal. One of the most cited contributions of the field is Williams (1992) where the relationship between the fundamental mode of SR amplitude and tropical temperature is established. This relationship works by means of the dependence between lightning activity (most common in the tropical region) and the convective available potential energy, which in turns depends on ambient temperature [30].

The growing concerns about global warming, the simplicity of the analysis, and the good acceptance by the scientific community put SR back in the spotlight. The topic is still in study with some works acting as follow-up of the theory, adding up evidence of the relationship [80].

Another representative work from this point in time is Boccippio *et al.* (1995), considering the role of *Transient Luminous Event*s (TLEs) in the ELF spectrum while finding a strong correlation between positive Cloud to Ground lightning and Sprite appearance. Sprites is the name given to the mesospheric, striated glows that can be observed over mesoscale convective systems. They are one of the many observable TLEs, phenomena located in the upper atmosphere, characterized for their curious shapes and brief duration [81].

Since they are related to lightning discharges, their nature is also the one of electrostatic charge displacement, but in the case of Sprites, it has placed in the lower part of the ionosphere

[6]. A diagram of Sprite occurrence can be seen in Figure 1.7, obtained from Paras and Rani (2018) [6].
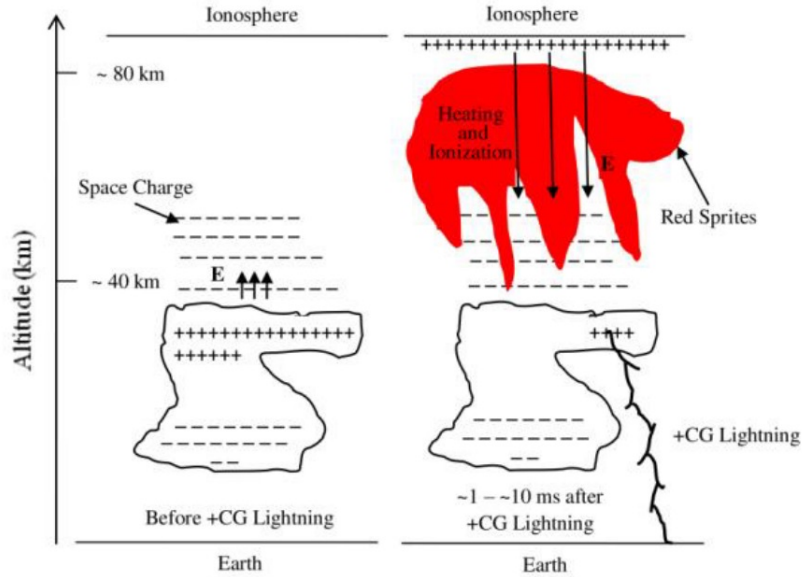
Figure 1.7: Storm cloud charge distribution for sprite generation, from Paras and Rani (2018) [6].

Thus, TLEs, which up to this point were mostly studied visually, were provided with another way of being looked at, and a whole other line of research to focus on [82]. The hard link to an specific cause promoted collaboration between TLEs and ionosphere researchers by showing they had a lot of common ground.

SR' interest as a research topic grows further as the 20th century comes to an end, and has been steadily growing ever since. A direct consequence is the number of ELF observatories established. A detailed list of the active observatories in 2016 can be found in [83]. It can be seen how over a 40% of the listed observatories were established between the years 1994 and 2000 and, since then, the number of active observatories have been on the rise. The new measurement stations set up in the last decade, their specifications, and results still have the interest of the community [84, 85].

On top of the many different research topics with the SR at their core, the main line of detecting and processing the phenomenon is still being worked upon. There are still a few works focused in different and/or better ways of processing and interpreting SR' records

[86, 87, 88]. This thesis falls in this last line of research. Is in this point in time where the research group TIC - 019 of Universidad de Almería, to which the author belongs, sets up their own station [89]. This station has provided all the ELF the data used to carry the present work.

1.3. Recent works on main research lines

In this section, different articles published either in the 21st century or at the end of the 20th are presented. The works selected were those relevant for the present thesis, but also those relevant to define the state of the current framework of SR' studies are included.

The papers have been arbitrarily separated in research lines in order to ease the reading. It is important to highlight this point, since many of the mentioned works could easily fit into many of these categories, since they all share the same focus. The goal of this separation is to orderly present the chosen articles, not the classification itself.

1.3.1. Lightning activity and the inverse problem

Since the beginning, efforts were placed into working mathematical models for lightning activity and SR, with the purpose of inferring lightning activity and ionosphere parameters from the SR records. This is known as the inverse problem of the SR, since the goal is to extract the variables that define the process from the process itself. Although lightning activity is separately researched in the VLF range in which pinpointing their location is possible with a network of enough sensors [90], there are advantages to provide solutions in the ELF range, since the low attenuation of the atmosphere in this band allows global coverage with a lower number of sensors [91].

A bright example of this line of work is presented in Nickolaenko (1997). The article expounds the analytical expressions for Earth's electric and magnetic field and, by accepting a handful of suppositions, he develops the analytical expressions for power spectrum divided by cartesian components in function of source-observer distance. Then, calculations using data from receivers of different characteristics are performed. Due to the intensity variation

experienced in function of distance, it is verified that total electromagnetic field power depends mainly of thunderstorm intensity, and that wide-band measurements (4 Hz to 40 Hz is the band used in the article) of a single component provide also a proportional output, only dependant from the level of global thunderstorm activity. It also shows the utility of the Poynting vector to detect variations in the fundamental frequencies of the ionosphere, since they show no dependence with source-observer distance [14].

Many other works follow the same steps. For example in Heckman, Williams, and Boldi (1998) an inversion method is developed in order to extract lightning activity from the three components of the electromagnetic field. Taking the field power equations from [92] and considering all lightning discharges homogeneous, they reach an average value of 22 flashes per second. Although the estimated value is way below the current estimate of 100 flashes per second, this might be explained by SR being mainly driven by cloud to ground lightning, which are roughly a third of the total flashes accounted [93].

In Nickolaenko *et al.* (1998) the global lightning parameters (localization and effective source width) are deduced from four years' worth of records from Nagycenk observatory. In addition of calculating an effective global thunderstorm diameter between 53° and 105° (roughly from 109 to 428 billions of km^2), they present spatial parameters for each of the global thunderstorm centers and even diurnal variations by the hour for each thunderstorm center. As part of the article is presented a backwards reconstruction that serves to validate the work performed [94].

Nonetheless, one of the main points of the previously cited work is the reveal of annual and semiannual variations which are tuned to the seasons, but which is also similar to the changes in median solar time. Not even a year later another article - Nickolaenko, Hayakawa, and Hobara (1999) - is published where the nature of these variations is explored and confirmed. Old records are explored to locate the mentioned variations and the captured signals from two distant observatories are compared, exhibiting similar tendencies and thusly proving its consistency. There is also an acknowledgement of daily global lightning activity variations being hard to track, exhibiting abrupt variations [95].

Other researchers build over the previous results, while including elements to provide additional accuracy or simply test new methods to solve the problem. Ando *et al.* (2005) address the problem by analyzing the geometry of the resonant cavity using finite difference equations, and taking advantage of the possibility to introduce specific and more detailed profiles to compute the resonances. Nonetheless, some simplifications had to be applied to the model due to the sheer amount of computational power required to solve the 3D problem as was initially planned. Even after considering the simplifications, the method successfully locate the three principal storm centers [96].

A more recent solution to the problem is brought by Prácser *et al.* (2019). They work over the classic equations of Madden and Thompson (1965) [44] and improve them further by adding the possibility of computing different observation zones and different elementary source zones. Each source zone can be associated with a different moment current, and each observation zone can be feed with different data. To test the model, they perform five different tests, each with different datasets with specific conditions. The main advantage of the module is being able to reconstruct lightning activity accurately, with no need of initial information about localization or initial distribution. It also shows fault tolerance, while working best with at least six observation points [91].

Moreover, the study of this relationship has received some specific contributions, studying how different aspects of each phenomena interacts with each other. For example, the contribution from positive and negative cloud-to-ground lightning has been analyzed in Surkov and Hayakawa (2010), concluding the importance of positive cloud to ground lightning over the negative ones. As the article mentions, this is explained by the greater current contribution from positive cloud to ground lightning despite being more scarce [97].

The analytical problem has been extrapolated to other planets. Pechony and Price (2004) is based on the popular knee model for atmospheric conductivity. After comparing the results of the newly developed model against Earth parameters and other works in the literature, they perform the calculations for Venus, Mars, and Titan (other solar system bodies who are candidates to have such a signal). Based on the results, they draw conclusions about the measurability of the signal in each celestial body [98].

Nowadays, the relationship between the electromagnetic signal and lightning activity is well understood, with studies that mathematically infer the latter using the former [99] up to the point that lightning monitoring through the analysis of SR records is more efficient than other lightning activity monitoring techniques [100]. Nonetheless, the topic is still under discussion and improved models are still under development [101].

1.3.2. Schumann resonance and Earth

Given the success of Williams' paper [30], different phenomena were tested against global electromagnetic signals (be lightning discharges, SR, or other electromagnetic signals) and other geophysical phenomena, in search for more relationships. All these studies share two distinct approaches, depending on the phenomenon under research. If it is transient events under study, the chosen events are located in time and the SR records are observed and analyzed in search for anomalies in the moment the event is present. On the other hand, if there is a variable whose relationship should be tested, detailed records are obtained and compared with the SR registers looking for correlations and other evidence of relationship. Among the latter, Price and Asfur (2006) analyzed the variations on troposhperic water vapor in the tropics and compared them with the magnetic counterpart of the SR, obtaining representative correlations between two variables. The highest correlation index being between magnetic field intensity and specific humidity, with $r = 0.90$, and the rest of correlated variables ranging between 0.50 and 0.70 [31].

A most sought target of these studies is to further quantify the variables that may affect the Earth-ionosphere conductivity profile. Some of them works directly towards refining the values of some ionosphere characteristics, such as the propagation parameters or conductivity model. For example, in Mushtak and Williams (2002) the knee model that generates a conductivity profile of the atmosphere by considering two main zones separated by a rapidly changing interface (the "knee") is tested. They conclude the importance of properly modeling the ionosphere region that acts as interface between the ion-dominated conductivity zone (lower ionosphere) and the electron-dominated conductivity zone (upper

ionosphere), considering it indispensable to accurately model the frequency of the signal [102].

In similar lines, in Kudintseva *et. al.* (2016) the conductivity profile as described from the *Direct Current* (DC) global electric circuit models is considered. There is a discrepancy between what the model predicts and current observations. Applying analytical techniques to SR' measurements they verify how the effect of the *Alternate Current* (AC) global electric circuit outweights the DC one. They propose a hybrid DC-AC profile, extracted from the SR records [29].

There are also studies regarding solar activity, such as Kulak *et. al.* (2003) on which the attenuation rate of the resonant cavity of the planet was computed from ELF records of six years. It is also shown the strong dependency of SR' fundamental frequency with the solar cycle, which points out the influence of the sun on the ionosphere conductivity [103].

A recent example of transient events' study is Pazos *et. al.* (2019), in which the effect of geomagnetic disturbances in the ionosphere are analyzed through the SR. To ascertain the effect of the different disturbances, they consider five scenarios where different geomagnetic disturbances are present. They conclude all geomagnetic disturbances have an effect in the first modes of the resonances' intensity. Nonetheless, the intensity of the disturbance does not have to be related with the intensity increase [104].

Another popular topic to study are gamma rays' effect on the ionosphere. In Nickolaenko *et. al.* (2012) a gamma ray burst is located and studied in Schumann resonance records, where a sudden decrease in intensity on all Schumann resonance modes is perceived. The event is captured from all three electromagnetic components, and also evaluated from the perspective of ionosphere conductivity. They compare the signal between other stations, validating the global nature of the effect. It is also pointed out how the sudden nature of the effect makes its undetectable by the usual spectral treatment, requiring direct observation of the time domain records [105].

Kudintseva *et al.* (2018) delves deeper in the topic, modelling the effect of solar flares and estimating the modifications caused in the ionosphere. As solar flares affect the ionospheric

conductivity and they decrease the lower ionosphere height, and these parameters in turn modify Schumann resonance parameters, the model directly connect both parameters. Along the extent of this work, it can be seen how solar flares modify the top half of the resonant cavity by reducing the low ionosphere height. They successfully build a model that relate solar disturbances with deviations on the ionosphere parameters [106].

Although studied from the very start of SR research, the work of Boccippio *et al.* [81] increased the interest of researchers for Q-bursts. Ogawa and Komatsu (2007) analyze over 350 Q-bursts with an electric, wide band sensor (1 Hz to 11 kHz, way beyond the cutoff frequency of the Earth-ionosphere waveguide) to further refine the initial classification made around 40 years ago. They focus their analysis on seven samples, to highlight the main observed characteristics. Despite having only electric field measurements, they calculate lightning flash position aided by the time intervals between direct and antipodal time of arrival. This technique allows them to estimate the wave's travelling velocity, using two methods and showing its results [107].

1.3.3.   Frequency and intensity variations

Another topic that stems from the initial studies and is still active nowadays is the study of the variations experienced by the fundamental parameters of the SR, namely central frequency and intensity for each mode. In the end, they are the most comprehensive expression of the phenomenon, leading to the fact that further understanding over these variations may enhance the knowledge over the entire global electrical circuit. An example of this line of work is Nickolaenko and Rabinowicz (1995), on which the cross-correlation between the frequency variations of two yearly records eleventh years apart evidences the relationship between frequency variations and global thunderstorm activity [108].

In Ouyang *et al.* frequency and intensity variations for the first three modes are presented and analyzed. The work emphasizes the terminator effect along daily variations, and a solar flare event is also accounted for in the records [109]. A more practical work is performed by Kudintseva *et al.* (2018) where frequency variations are studied under the effect of solar flares and other solar events. Many different variables are taken in consideration as well,

such as middle atmosphere conductivity, and characteristic heights of the full conductivity profile under solar event conditions [106]. More recently, Cano *et al.* (2021) display a detailed analysis of both frequency and intensity of the sixth first modes. A comparison is made between other works in the literature, and a novel approximation in order to identify specific phenomena associated with specific times of the day is presented and tested [21].

Although intensity variations are agreed throughout the literature to be related with thunderstorm centers' activity [109], frequency variations are not so easily comparable due to source-observer distance. This increases uncertainty when interpreting the results, and gives way to analytical procedures to compare them. Nowadays, the most common tool to fit SR spectra in order to ascertain peak frequency of each modes are Lorentzian sums, although different approximations to the problem are taken either to the application of the function or the first approximation calculation.

1.3.4. ELF anomalies induced by seismic activities

Earthquake induced anomalies in the ELF spectrum is one of the most active research lines (along with finding a solution to the inverse problem), due to the contribution the research might provide in the topic of earthquake prediction. Although research on the topic begins in 1983 [110], the publications on the subject dating from the 90's are scarce, and mostly published in local journals [111, 112].

It is from the year 2000 onwards that the topic rise in popularity among the global community, finding evidence for earthquake precursors in different electromagnetic media [113]. The first convincing evidence of the effect appearing in the literature is Hayakawa *et al.* (2005). Through a detailed study, they examine the SR records of Nakatsugawa station for 1999, when the great earthquake in Chi-Chi, Taiwan, happened. The observatory is chosen due to its proximity to the epicenter. A comparison is made between the mentioned records and those from other station (located in Lekhta, Russia) used as reference for being a long distance away. The intensity of different modes is evaluated, and the spectra of both magnetic components extracted. An anomaly in the amplitude of the 4th mode is detected on Nakatsugawa records, along with a significant frequency shift. In the discussion, they

propose the hypothesis of the disturbances being a product of a scattering effect caused by the disturbed ionospherical region of Chi-Chi [33].

The work's conclusion are further reinforced by an extensive survey of other earthquakes [114], on which both the intensity enhancement and the frequency shift was detected for inland earthquakes with magnitude over 5. Additional efforts are poured to analyze the subject, modelling a disturbed conductivity profile for the ionosphere and analyzing in depth the ELF scattering hypothesis [115]. Over time, interest over the subject has only risen, with many different groups providing experimental evidence during the course of the last decade, with many events recorded by different teams in different parts of the world. The use of other analysis techniques is explored, compounded with additional data to provide ideas that reinforce the link between the ionospheric and electromagnetic anomalies detected and earthquakes, always with the mind of developing enhanced methods to predict seismic activity [116].

Among the recent findings on this line of research, Christofilakis *et al.* (2019) show that, prior to an earthquake which epicenter is closer than 500 km to a ELF station, the third mode of the resonances is enhanced while the first is suppressed [117]. Other interesting work is Galuk *et al.* (2020), where a mathematical model based on the absolute conductivity of the ionosphere is introduced. According to the authors, this model allows to estimate the magnitude of significant earthquakes from the disturbances observed in the SR records [118]. Although the results are good, authors point out the need of validating the model through case studies. This is done so in the paper by Hayakawa *et al.* (2020), where two cases of study are analyzed under the described model. On this work, previous results are satisfactorily reproduced while different considerations on the presented model are calculated [119]. In the work by Cano *et. al.* (2022), Deep learning algorithms are introduced to the problem of earthquake detection. A system as complex to accurately model as the Earth-ionosphere cavity seems like a proper candidate for the application of these automated techniques. The paper shows promising results, with possibilities of refining the algorithm's training with further data and expert knowledge [120].

1.3.5. Virtual models

Other example of current development in SR' research are virtualization and computerized modelling. Many researchers have used virtual tools in order to contribute to the research, either through neural networks [121], simulated models [122, 123, 124, 125], or other tools. These virtual tools are a step forward in the world of SR studies, since they provide further means to validate theoretical models and to test instrumentation with a digitally generated signal. A recently published article - Füllekrug (2021) - features a frequency model for the Earth-ionosphere cavity that is developed in an analytical fashion by constructing a likelihood function depending on time and surface position in polar coordinates. From there, the equations for each component of the magnetic field are extracted and simulated, then compared with both the SR records and lightning activity records obtained via satellite [126].

Despite the advances produced in this field, most models are based on the statistical nature of the signal and the phenomena that contributes to it. As a consequence, they are able to simulate the steady state of the SR, but not to take account the different responses of the Earth-ionosphere cavity due to temporal disturbances or certain configurations.

A line of research that relies heavily on virtual tools is the study of SR in other celestial bodies, with different models proposed for different planets of the solar system or even their moons [98]. Either with an analytical approach [127] or directly using computer simulations [128], the few experimental records provided by probes and other space exploration equipment is tested by minds and computers to produce additional information. Although electromagnetic propagation in other planets was already a topic of interest, being necessary to retrieve data from the exploration vehicles, its interest increased along with the voyage of the Cassini mission, and spiked in 2005 with the landing of the Huygens probe on Titan. This is so since the ELF electric signals it captured has been considered by some as the moon equivalent of SR [129]. Nonetheless, the topic is still under discussion [130].

1.4. Statistical analysis of Schumann resonances' time series

The state of the art presentation from the previous Section show SR are a recently discovered phenomenon with its own community of active researchers working on different aspects of it. Thanks to SR studies, the knowledge on a few subjects (e.g. the global electric circuit, ionosphere composition and dynamics) has increased and, and there is potential to grow further in these subjects, as well as new ones to be uncovered in case of a new breakthrough (such as earthquake detection, geophysical properties or outer space exploration).

Despite the many different approaches towards SR studies, the vast majority of the published works relies on the frequency domain to extract its conclusions. This is to be expected, since a great deal of information can be extracted by the frequency components of any given signal. On the other hand, that doesn't mean that time domain holds no additional information at all.

1.4.1. Time domain registers of Schumann resonance

Evidence of time-domain records containing additional information can be extracted from the studies available in the bibliography who have relied on them. Special emphasis has been put on these kind of articles along the introduction, to highlight their presence whenever they could be found. To summarize the works that rely on the temporal aspect of the SR records for a different purpose than obtaining their spectral information, a classification will be presented:

- **ELF noise analysis:** The studies performed during the 60s and the 70s to characterize ELF noise to improve the performance of low frequency communications relied considerably on the temporal records, to characterize the noise statistically [18, 63, 64, 66, 68, 131].

- **Transient event analysis:** The common procedure to obtain the SR frequency spectra consists in separately obtain the frequency characteristics of different segments via *Fast Fourier Transform* (FFT) and average the results into the desired spectrum. This can be done using Welch's averaging method of spectral density estimation. If using 10 s realizations, the resonant peaks of the signals will be visible on the averaged spectrum as long as the number of averaged realizations goes over 16 [4]. Nonetheless, there is no consensus on the optimal number of segments to be averaged, although the subject has been addressed in the literature [88]. Transient events such as Q-bursts do not last long enough to extract their spectra through the method described above, which is why the information has to be treated using the temporal records. Although some authors either develop or apply different techniques to extract the approximate spectra of Q-bursts and other transient events [58, 72], most of these works rely greatly on time domain analysis techniques to carry on their research [8, 81, 82, 87, 105, 107, 132].

- **Signal propagation:** Among the articles featuring studies about the fine structure of the signal, some of them resort to time domain analysis techniques. Topics such as signal polarization [79, 133, 134], wave propagation [135, 136], and ionospheric modulation [20] fall in this category.

Among the studies presented above, most of them rely on time domain records to show a certain event or a certain segment relevant to the study. Even when shown they are means to an end, and in too many occasions no information whatsoever is provided about the processes the signal went through before appearing in the article. SR studies with a time-domain based approach are scarce.

By consulting the literature, a few articles focused on analyzing time domain registers are worthy of being brought up at this point. One example is the previously mentioned work - Nickolaenko (1997) - on which he studies the relationship between thunderstorm activity and the SR using mostly time-based methods [14]. There's another work from Nickolaenko, Price, and Iudin (2000) where the Hurst exponent of the signal is derived and its anti-persistent nature analytically verified [131].

Special attention should be placed in Greenberg and Price (2007). On it, a histogram representation of a SR' data segment is displayed, showing an excellent fit for the Gaussian distribution. Nonetheless, their analysis is focused on the distribution of intense transient events (amplitude over 15 pT, five times the standard deviation of the background noise amplitude) which are found to be following a Poisson distribution. This highlights the close relationship between lightning activity and the SR since lightning discharges follow the same distribution [8].

In Cano *et. al.* (2022) a methodology to analyze and classify transient events on the records of a narrow band ELF sensor is developed. Each transient event is modelled using the Heidler function, and thus the main parameters of each transient events are extracted. Similarities and differences are then analyzed between Q-bursts and the rest of transient events detected [137].

The cited articles above emphasize two points; that there is valuable information to be extracted from the time domain aspect of the SR, and the absence of extensive analyses performed over this data.

1.4.2. Data imputation in time series

When performing experimental analysis, it is recommended to extend the amount of data the analysis is applied to. It increases confidence in the results and gives way for generalizations to be reinforced and anomalies to be spotted. That being said, to validate an analysis that rely on sheer quantities of data, the way data imputation is addressed must be exposed and information about data integrity presented.

To appropriately present the results in these cases, one must account for the existence of *Not Available* (NA) values on the records due to data loss in a way that minimizes their impact on the analysis results. In some cases, the analysis can be performed using complete segments of the appropriate size while discarding the rest. Still, when extracting patterns from time series, ideally the whole register should be complete since each segment belongs to a different time frame and the information they bring is different [138], so NA values

must be estimated. This makes data imputation in time series a subject of recent studies, [139, 140, 141] especially for real-time analysis.

Imputation methods development is mostly applied in research where the studied events have a time frame similar to the instrument's capture frequency [142]. There are also studies where the importance of data imputation on aggregated data is evaluated [143]. Data aggregation comes with the setback of obscuring effects only observable under the aggregation threshold [144]. When the purpose of an analysis is to study, over years of data, time trends - yearly, monthly and daily patterns - that can be considered long concerning the number of data points, data can be aggregated in periods that do not obscure the phenomena under study while mitigating the impact of NA values.

1.4.3.   Statistical analysis and the Schumann resonance

In the disciplines related with wireless telecommunication, statistical analysis is fairly common, used to characterize both noise and the signal distribution to improve signal reception [145].

In the framework of the SR, statistical analysis is not uncommon, but only some researchers have performed statistical analysis on the signal itself [131]. Then again, these analysis are rarely thorough, providing only anecdotal results which are used as a stepping stone in order to prove further points. For example, a few works declare that the SR' distribution should be typically Gaussian, either based on the assumption that as a natural phenomenon they should follow that rule of thumb [85] or by analyzing an insufficient amount of data [8].

Given all of the above, the lack of studies focused on analyzing the time domain records of the ELF spectrum present both an untapped research line and a gap in the literature. This thesis aims to explore what information might be obtained through the temporal analysis of ELF records, with its focus on the SR. Although they have not been exploited in the SR studies, there are methods with solid theoretical foundations that study electromagnetic signals from the time domain perspective.

1.4.4. Preliminary investigation on techniques of statistical inference applied to Schumann Resonances' data

The discipline of statistics has been, is, and will be key to scientific development in every discipline. Statistical methods represent the tools of the trade to test experimental results, evaluate the quality of a given model, and treat raw records to obtain manageable data. Nonetheless, it is also an extensive field, and choosing the right tools is not always easy, but fundamental for the purpose selected.

The framework of this thesis was developed while searching for a statistical tool; the goal during the early stages was to find the most appropriate tool to classify segments of a continuous register of our SR records as the most likely distribution to explain the segment. The first tool to be evaluated was goodness of fit tests.

Goodness of fit tests are a subset of statistical hypothesis tests. Statistical hypothesis tests, widely used in many fields of science but especially in Biology and Psychology, allow scientists to extract conclusions from the sample population with a reduced number of samples. Loosely speaking, the test statistic is a representation of how different the two compared samples are.

On these tests, the statistic is used to obtain information about how similar two samples (or a sample and a statistical distribution) are. Among these tests, one of the most widely used is the *Kolmogorov-Smirnov* (KS) test, called simply Smirnov tests on its two sample variant. The concept behind both of them is the same, measuring the point where the vertical distance between the cumulative functions of two samples is maximum. That being said, KS tests a sample against a theoretical distribution whereas Smirnov tests two samples between each other, to see if they follow the same distribution [146].

The usefulness of KS comes from its lack of strong assumptions; while other tests assume data will be normally distributed, or even linear, KS only requires specifying the parameters of the theoretical distribution in advance [147], while Smirnov test has no hard requirements to work.

This quantity measures the maximum absolute difference between the *Cumulative Distribution Function*s (CDFs) of both empirical distributions (Equation 1.5, with $F(x)$ and $G(x)$ representing the CDFs of the two compared time segments) which fulfills all the previous conditions while effectively quantifying the difference between two statistical distributions, as it has been applied in other works for the same purpose [148].

$$KS_d = sup_x(|F(x) - G(x)|) \qquad (1.5)$$

This and other tests were applied to SR data to ascertain their theoretical distribution. Nevertheless, the large number of samples per segment involved (33 660 samples just for a minute-long segment) made the tests hypersensitive. Minimal differences between the sample population and the target distribution were enough to determine a poor fit, while the ones classified were all deemed excellent fit. The lack of gradient between positive and negative results made this approach undesirable. Up to this point, the solution was to recalibrate the tests manually (either relying on the statistic's value instead of the p-value, which would require establishing a metric to be able to compare the different results, or to redefine the statistic for the test, practically creating a different test) or find another tool.

Thanks to the background laid out by Claude Shannon in his *"Mathematical theory of communication (1948)"* [149], additional methods of statistical nature are available for the analysis of electromagnetic signals in the time domain. Despite of the time that passed since the article was published, the techniques outlined have penetrated slowly in the fields where statistics are required, with researchers relying still on outdated significance tests. The mentioned tests are without question a familiar tool (reason why they were the first we tried), useful for many cases, but experiences such as ours points out the importance of looking for statistical tools more appropriate for the case of study [150].

A suitable alternative was found in likelihood statistics, by means of the method of *Maximum Likelihood Estimates* (MLE). The concept behind the likelihood function is fairly simple to grasp, albeit growing complex when treated in a rigorous mathematical fashion. Nonetheless, in the framework of this thesis the application of these concepts are just means

to an end, for which no interest is deemed in delving further in the mathematical aspects of the tool. That being said, a light explanation will be provided to inform the reader about the purpose of the function and its role in the world of modelling.

When a model or statistical distribution is known, the probability of a realization under the model constraints is given by its *Probability Density Function* (PDF) (Equation 1.6), with x being the realization of the random variable - the interval of all possible results the PDF may yield - and θ the parameters that define the model.

$$x \rightarrow f(x|\theta) \tag{1.6}$$

The probability of x "happening" depends then on the values of the parameters of the model θ. When the interest is placed in the model instead of the realization as it is usually the case in experimental science, where a realization (sample) is available and the unknown lies in the model parameters that may yield such realization, the expression is inverted. In this case, the goal is to ascertain the likelihood of a certain set of model parameters to produce the realization. This is expressed as depicted in Equation 1.7, with $p_\theta(x)$ being the likelihood function.

$$\theta \rightarrow f(\theta|x) = p_\theta(x) \tag{1.7}$$

By selecting the PDF for a specific model, Equation 1.7 can be derived to calculate the values for θ that maximizes the likelihood of a specific realization. For computational convenience this is done by finding the maxima in the log-likelihood function [151]. Among the possible values, the MLEs of the model will be the ones that satisfies both Equation 1.8 and Equation 1.9, with $\theta_i \in \{\theta_1, \theta_2, ..., \theta_n\} = \theta$, each of the parameters on which the model relies.

$$\frac{\partial \log p_\theta(x)}{\partial \theta_i} = 0 \tag{1.8}$$

$$\frac{\partial^2 \log p_\theta(x)}{\partial \theta_i^2} < 0 \qquad (1.9)$$

This formal description is akin to the concept of the same name used by everybody everyday. A blunt example that hopefully put the concept in perspective follows: If I am expecting a visit from a friend who lives in another continent, and my friend performs the trip in less than a day (the realization), I would come to think he travelled by plane. Among the means of transportation (parameters) my friend could have used, I infer the most likely is plane, given the time invested in the journey. An inference similar to the intuitive process described in the example above is what MLE produces, but in a strict mathematical sense.

A step by step description of MLE along with examples and comparisons with other classic modelling techniques can be found in Myung (2003) [151].

Formal instances of the maximum likelihood concept appear as early as 1713, when the first solutions to the model estimation problem are being proposed. The idea can be perceived in the works of Lagrange, whereas Laplace and Gauss worked on the concept by means of Bayesian statistics. The case of the latter is especially interesting, since it develops the concept further to develop the Least Squares Method, a method that gained popularity because of its simplicity of application, despite presenting certain constraints that are usually ignored. Despite that, it was never properly formulated until Fisher proposes in 1912 (still an undergraduate) the method as a criterion to obtain the relative probabilities of different set of values (parameters) given the probability distribution is known. Nonetheless, it is not until 1922 that he revises the method and introduces the word likelihood [152].

Despite the intuitiveness of the concept and the many times it resurfaced, the likelihood function and MLEs were shunned in favor of other techniques. It was the enhancement of computer power what has brought likelihood estimation towards the front line, along with its versatility. [153]

To sum up, MLEs provides a way to choose the parameters that, for a given model or probability distribution, are most likely to accommodate the data from which the parameters are estimated. It must be taken in consideration that this does not evaluate the quality of the

model; if the model is not adequate, the parameters obtained through MLE will not provide an adequate description of the data. Since the purpose of the work supporting this thesis was to classify segments of SR temporal records under the most appropriate distribution, estimate the best parameters for a specific model was not enough; another tool is necessary to evaluate the performance of each model and choose among them the one that best describes the data. Statistical goodness of fit tests were out of the question for the reasons explained above, but the use of likelihood allowed for the application of the *Akaike Information Criterion* (AIC).

Developed under the frame of information theory, AIC is essentially, a model comparing mathematical tool, its expression shown in Equation 1.10 with k being the number of parameters used in the model and $\hat{p}_\theta$ the maximum value of the likelihood function for the model.

$$AIC = 2k - 2\log(\hat{p}_\theta) \tag{1.10}$$

AIC returns an index for each comparison between raw data and a suggested model for the data that implies how much information would be lost if the model was used to represent such data [154]. Beforehand, the results of many different comparisons (different AIC indexes) are not comparable between them, but the literature offers a way to normalize these results, called Akaike weights. These transform AIC results into real numbers between 0 and 1 which are comparable between them [155].

Given the goal of the preliminary investigation, which was to determine which distributions fit the data best for each processed segment, AIC was deemed as the appropriate tool. It simplifies the process of determining the distribution that fits the segment best, with no issues regarding the number of samples. This is very reason why is widely used in radio communications [156, 157]. Despite being traditionally associated to information, the theory and its associated methods have been applied in many different fields, such as in epidemiology studies [158].

There is another widely used information criterion which stems from Bayesian statistics, called *Bayesian Information Criterion* (BIC). Its formulation is similar (both methods will

be presented in Chapter 3) with the difference that BIC introduces a higher penalty for models with high variables [159]. Both criteria were compared under the framework of this thesis, while testing SR data's quality of fit against distributions defined by more than two parameters. Regardless of the criteria used, distributions defined by three parameters were selected even when visual inspection or q-q plot graphics showed significant deviation. When tested against distributions with only two parameters, AIC showed a higher tolerance for discrimination than BIC, displaying moderate values in those segments whose fit was not appropriate to any distribution. This was to be expected; as can be found in the literature, BIC is most suited when the goal is to find the true model the data follows, whereas AIC is more suited to find the best candidate among a different set of models [160]. Even when a segment of SR data might be precisely described by a certain distribution with specific parameter values, there is only certainty that the model will be valid only for that specific segment. Under this knowledge, AIC was selected as the most appropriate tool to discern the best fit among the candidate distributions [161].

Likelihood statistics application through AIC make the core methodology of this thesis both solid and modern, with many more advantages than the few exposed here (see [162]). Therefore it becomes the cornerstone of this work and through the chapters presented in this thesis, novel techniques have been applied to produce an actual document.

Chapter 2

Hypothesis and Objectives

In this chapter, the starting point for this thesis will be expanded and its objectives described.

This thesis' purpose is to explore the possibilities that statistical methods offer to advance in the research of SR. To do this, the aforementioned methods will be applied to time domain records of the ELF spectrum, following planned methodologies under previously contemplated hypotheses. Through the statistical results, both new and old observations will be performed to ascertain the usefulness of the procedures.

Specifically, two different techniques have been developed under the research that gives way to this thesis.

The first analysis' method is built around AIC and likelihood estimation to automatically classify SR temporal segments under the distribution that fit the the data best. The second relies on KS distance to quantify the variations experienced from one temporal segment to another, considering the SR as a stochastic process.

The motivation for both of them will be explained below, and a detailed description of both methodologies can be found in Chapter 3.

2.1.   Statistical classification

The technique proposed here is, as previously stated, already in application in other electromagnetic signals [157], and well developed in the frame of information theory. Given a data segment of a certain duration, likelihood estimation provide a tool to find the parameters of a given distribution that fit the data best, while AIC is a way of choosing the most representative distribution from a set of best fits.

The purpose of this technique in the fields where it is applied is to characterize statistically a signal under specific circumstances to improve techniques of data recovery. Nonetheless, it will be applied to the SR with an altogether different goal. The working hypothesis could be resumed as follows:

1. Given the nature of SR and the many variables involved in the phenomenon, its impossible to accurately model it.

2. Through histogram representation, any given amount of data can be statistically compared to a suitable distribution that represents it, but it incurs a loss of information proportional to the amount of data processed this way since it will mask any individual variations into the whole.

3. It is possible to divide a time domain record in different segments and apply the statistical tools to each of them, effectively characterizing each time interval separately.

4. Taking item 3 for granted, through further analysis and observation of the results, relationships may be uncovered between the parameters of each segment and the events that happened on that specific time interval.

Although the proposal introduced in item 3 above is somewhat intuitive, the extent of the analysis' usefulness must be evaluated and tested. In order to fulfill this first stage of the thesis, the following objectives are proposed:

- Create a software tool capable of process via AIC large amounts of data.

- The results must be stored and tagged to the data segment they belong, to increase the analysis' reusability.

- Test the tool in a dataset and evaluate the results, trying to find meaningful relationships between one or more SR' facts and the statistical analysis of the time domain records.

Once the method has been validated successfully by fulfilling the last of the bullet points above, further tests and experiments will be carried, either by correlating them with data of some other relevant phenomena or by further exploring the SR themselves.

2.2. Kolmogorov-Smirnov distance as variability estimator

In this case, KS distance will be used as a metric to evaluate the differences between temporal segments of the SR interpreted as histograms.

Nowadays, it is clear that the SR reflects the state of Earth and the atmosphere [2]. Consequently, SR frequency and intensity values do not depend on the resonant cavity dimension alone, but also from the state of Earth's crust (lower layer of the resonant cavity), ionosphere (upper layer of the resonant cavity), and atmosphere (the media in between). Despite the stability and periodicity of all these layers, they only define the boundaries of the signal. Lightning activity, SR's main source of excitation, can be classified as a stochastic process [27]. Therefore, the intensity variations in the resonant signal can be studied from the perspective of a stochastic process as well [97], even more so when considering the other factors that may influence SR. Although this is not especially relevant when working on monthly or even daily timescales, on which their patterns and structures emerge, it has a noticeable impact when working with data segments.

Following this train of thought, quantifying the signal's variance over time can provide additional information about lightning activity and the state of SR. The previous facts and the following hypothesis can be broken down as follows:

1. Despite all the trends identified in the signal, the SR is generated by an inherently stochastic process (lightning activity) and influenced by other factors that might influence its short-term behavior.

2. Through histogram representation of the SR temporal segments, a measure of the differences between two segments can be obtained by means of a test statistic, such as KS distance.

3. The quantified variability obtained points out the degree of change in the resonance between both segments.

4. Automating this technique in order to apply it to many segments at once, the variance over time of the SR can be quantified.

An easy method to quantify SR variability over time can be helpful as a complementary measurement; for example, it can provide information about the state of the signal in a way researchers can decide which temporal segments to choose for their experiments. It may be crucial to identify oddities in the segments by calculating the differences between many segments on a given period of time, then locate which has the greatest average difference.

Furthermore, providing SR' variability along with the data enhances confidence in the results and helps other researchers reproduce the results by performing the analysis on samples with the appropriate characteristics. Experimentally obtained criteria could be enhanced by having an independent metric to quantify the signal state. An example is the appropriate segment size and the number of averaged segments required for the resonances to manifest in the records [4]. While averaging a sufficient number of segments to ensure the resonances can be visible is effective (as proven by all the meaningful results produced under it), advantages may arise from quantifying the stability of the signal, such as choosing the most stable segment to carry sensitive analyses or controlling the temporal resolution in frequency.

To explore the potential of such a method, the objectives detailed below were proposed:

- Devise a methodology to automate the variability quantification analysis over a defined length of data, providing different parameters for configuration.

- Test the methodology results by analyzing different temporal records of different seasons, varying the parameters introduced above.

- Try to find a practical application to show the impact of signal variability.

Chapter 3

Methodology

This chapter is organized as follows: First, the ELF observatory is shown, describing each part of the measurement system. Secondly, the data registers produced by the sensing station will be explored, along with the main characteristics of the signal. Lastly, the two methodologies developed for this thesis will be presented; the statistical classification in the first place and following, the method to quantify statistical variation.

3.1.   ELF observatory

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. It features a couple of coil sensors, one with NS orientation and EW for the other. The sensors are buried underground, their output fed to two amplification stages each, equipped with analog filters that reduce signal noise, and adapt the signal to be read with a sigma-delta, 24 bits *Analog to Digital Converter* (ADC).

The sensors were developed by our research group. Different sensors were designed, varying their section, number of windings and core material. They were characterized using a network analyzer with a two fold objective; select the one with the best frequency response

for the desired range, and use its characterization to produce a calibration function [19]. The two analog amplification stages were designed, tested and characterized as well, with its main features described in [89].

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 187 Hz (336600 samples per file) and a bandwidth ranging from 1 Hz to 100 Hz.

A photograph of the station where all the instrumentation is stored (with the exception of the sensors, which are underground) can be seen in Figure 3.1.

Figure 3.1: Schumann Resonance observatory in Sierra de los Filabres, Almería, developed by TIC-019.

Its remote location required to provide autonomy for the system in the way of batteries, and the solar panels that provide power to the batteries. They power both the measuring equipment and the security and surveillance equipment as well. Given its remoteness, data is sent from the station to the server via radio-link. A more detailed description of the station and their components can be found in [163].

The station's setup allows for continuous data capture and storage, which made gathering the data for this study possible. That being said, they are also elements prone to failure. Continuous cloudy weather prevents the batteries for being charged via solar power, and atmospheric phenomena may disturb radio-link communication. Maintenance or upgrade on the system may cause longer outages. In conclusion, if rigorousness is to be maximized when analyzing the stored SR records in great quantities, data imputation techniques must be accounted for.

3.2. ELF data registers

The ELF records produced by the observatory described above cannot be directly accessed to retrieve ELF data, since the captured information is convoluted with the response of the sensor. This is why a calibration process of the records is required prior to data analysis. To do so, the whole measurement system (sensors, analog amplifiers, and ADC stage) was properly characterized with a network analyzer, and their characteristic function extracted. From these combined responses a calibration function is obtained, which in turn is applied to the raw registers, producing calibrated data registers from which ELF temporal series can be read.

3.2.1. Preprocessing applied to the records

Depending on the purpose for which the registers are meant to be analyzed, some sort of pre-process might be desirable. It must be taken into account that, in the 1 Hz to 100 Hz band that the station captures, there are a few signals besides the SR, their effect considered and the impact of each independently analyzed. Among these, the European power grid influence with a fundamental frequency of 50 Hz, as well as the 60 Hz power grid from America and Asia are always present, albeit with different levels of magnitude. Other common disturbance appears on the low frequency range (from 1 Hz to 6 Hz) of the sensors' bandwidth, caused by the wind rattling the station fences and the earth itself on which the sensors are buried. The effect of these unwanted signals in the SR' spectra can be recognized in Figure 3.2. It can be appreciated how the power grid signals are present in both Figure 3.2a and 3.2b, and

even if low frequency interference only appears in certain segments, its effect might extend to interfere with the first mode of the SR, as 3.2b shows.

Other undesirable effect produced by low frequency disturbances is the amplitude increase throughout all the spectrum. It can be observed how Figure 3.2a displays a clear spectrum with a smooth gradient, becoming almost flat around 40 Hz, with the presence of modes above the 5th being only hinted at. On the other hand, the spectrum displayed in Figure 3.2b displays higher intensity in the lowest part of the spectrum, with an overall power increase affecting the whole signal. This is because low frequency disturbances have a flicker profile which, given enough amplitude, may mask the signal intended for measurement. Consequently, the sixth mode which was already weak on Fig 3.2a, is not visible on Fig 3.2b.
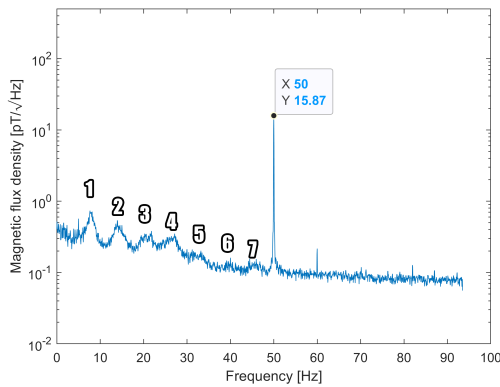
(a) Capture from 15-Apr-2016 02:23:05 to 02:33:05 with no unexpected significant disturbances. The data point shows the value of the European power grid signal.
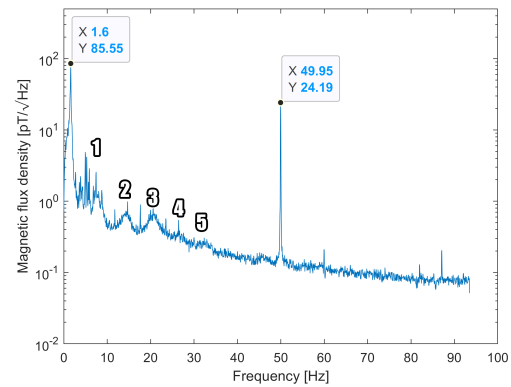
(b) Capture from 15-Apr-2016 11:42:51 to 11:52:51 with low frequency noise and the European power grid signal, both peaks characterized with datapoints.

Figure 3.2: Two 10 minute captures in the frequency domain, with the observable SR' resonance modes numbered.

Figure 3.2 shows how under the frequency representation the power grid signal is only another spectral component, which is only obtrusive to detect SR modes above the 7th - those whose frequency overlaps with the signal's. Nonetheless, the high power displayed by the signal around 50 Hz means it will mask any other signals whose intensity is below it, such as the SR modes. The same goes for low frequency disturbances; if the effect is strong enough (such as in Fig 3.2b), the time series signal will be representing the wind effect

in the measurement system. These masking effects can be clearly seen trough histogram representation. Figure 3.3 display the normalized histograms corresponding to the same segments depicted in Figure 3.2.
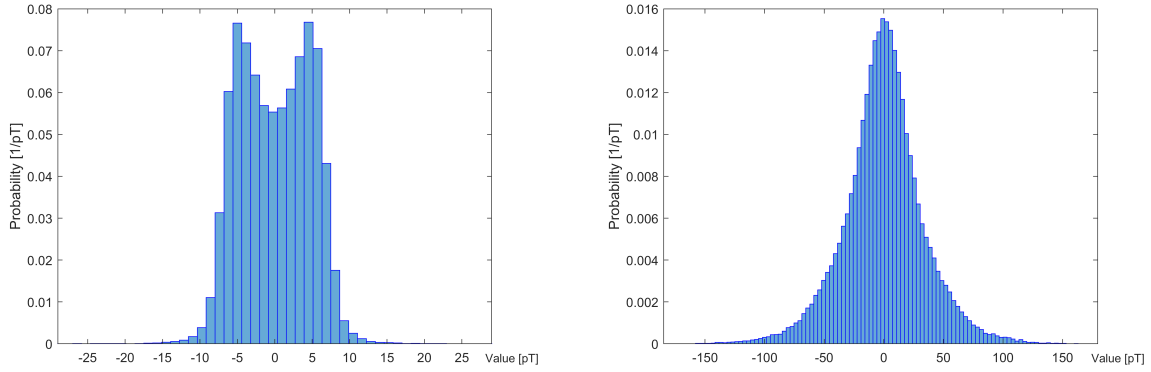
(a) Histogram representation of the sample depicted in Figure 3.2a.

(b) Histogram representation of the sample depicted in Figure 3.2b

Figure 3.3: Normalized histograms of 10-minute segments.

Figure 3.3a shows the characteristic two maxima of sinusoidal signals, implying the dominant signal is the European power grid, with a peak-to-peak value of 10 pT. Despite the bell-shaped curve shown in Figure 3.3b, its frequency representation (Figure 3.2b) tells us that the distribution tails are caused by the wind effect over the measurement system and the station, not the electromagnetic signals in the ELF band.

These artifacts and unwanted signals are present throughout the records, with some other whose spurious nature prevents easy identification. To narrow down which part of the ELF calibrated records it is meant to be analyzed, different filters will be applied to the raw data, and each resulting filtered signal will be analyzed independently. Analyzing raw data, or only partially filtered data may be interesting in order to identify and understand how disturbances contribute to the time series. The filters considered to thoroughly test the impact of all possible interferences are:

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

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

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

4. Both previous filters applied at once. Under this filtering process, SR should be the signal with highest power without removing the contributions to the rest of the band.

5. A *Band Pass Filter* (BPF) with 36 Hz bandwidth (from 6 Hz to 40 Hz) capturing the SR first fourth modes, which represent most of the SR 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 [88], the original records (duration of 30 min) will be split in smaller segments of different durations. Its goal is to characterize different events from the SR records; shorter segments will be deeply influenced by any transient event present in it, whereas longer ones will contain, under standard conditions, mostly the background noise signal. Therefore, shorter segments might be useful to analyze specific transitory events and detect strange occurrences, while longer segments will represent ELF background noise better, as the influence short transitory events may have will be diminished. The possibility to analyze overlapping segments has been introduced in the analysis, giving the possibility to smooth over the transition between one sample and the next by controlling the amount of time they share.

3.2.2. Data imputation implementation

As stated in Section 1.4.2, data loss due to malfunction or maintenance is not uncommon. Our ELF data registers are no exception, and to reliably present any analysis in which a great number of registers are involved, the way data imputation is addressed must be exposed and information about data integrity presented.

To address this problem, NA values' propagation was prevented as long as the amount of NA values goes over a certain threshold. This is done so by reasoning that even in a situation with a noticeable amount of missing values, the aggregation of the remaining values will be

more informative than any other value. If the presence of NA values goes over this threshold for a certain aggregation, then the result of the operation will be NA. Nonetheless, if the aggregation is part of some trend extraction process (such as a *Moving Average* (MA)) the gap will be covered through linear interpolation, since these methods have no tolerance for lost data. It is expected that the less NA values finally displayed, less impact the linear interpolation will have, as long as the interpolated gap is not big enough to interfere with the analyzed trend.

To provide the reader with the data integrity information when necessary, a color coded bar as the one shown in Figure 3.4 will be displayed under the relevant figures, showing the fraction of missing values implied in the calculation of each day's value.
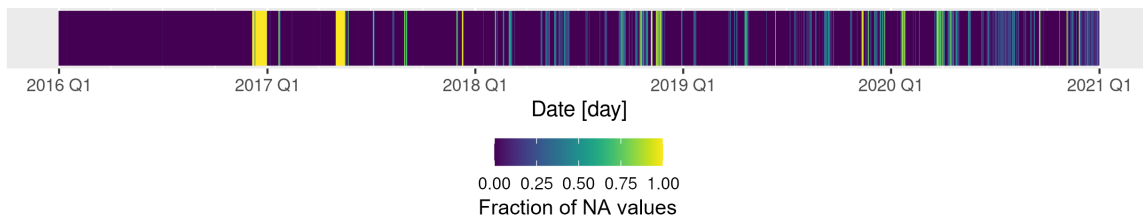
Figure 3.4: Color coded bar showing the fraction of NA values involved in the averaging of each day.

Data imputation techniques are especially relevant for Section 4.2, in which the methodology will be validated by observing the results when applied to five years of data.

3.3.   Statistical classification methodology

As it was discussed in Section 1.3, it is stated in some actual works that SR are represented by the Normal (or Gaussian) distribution. The few works that mentions the previous statement do it without including any references [85], and when studied, only anecdotal analyses are performed [8]. Contrary to the quoted articles, it was also found in modern literature how ELF spectra shows heavier tails than the Normal distribution [131], and there are certain works in the 60s and early 70s [63, 64, 66] reaching the same conclusion. Shed light into this topic is one of the immediate goals of this thesis.

Below, specific pieces of the methodology will be presented in different subsections, to describe in detail the need of each part and explain their implementation. Last, the overall description of the methodology will be provided, showing where each of the described parts belong.

3.3.1. Target distributions

A set of target distributions has been chosen with typical kurtosis value above the one of the Gaussian distribution. The distributions chosen for this matter and the reasons why they were chosen are described below.

- *Gaussian distribution:* Despite the works cited above implying a different statistical distribution for ELF background noise, studies showing a Gaussian response for the SR signal cannot be ignored. Thusly, it is important to test whether the Gaussian distribution can offer a proper fit for the analyzed data. It is, consequently, a clear choice for the analysis' set of target distributions. Its PDF is given by Equation 3.1.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{3.1}$$

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 fading [164]. By definition, it is the probability distribution of the magnitude of a circularly symmetric bivariate normal random variable. 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, justifying its inclusion among the target distributions for this study. Its PDF is presented in Equation 3.2, with s being the distribution's noncentrality parameter, σ its scale parameter, and I_0 is the zero-order modified Bessel function of the first kind.

$$f(x) = \frac{x}{\sigma^2}e^{\frac{-(x^2+s^2)}{2\sigma^2}}I_0\left(\frac{xs}{\sigma^2}\right) \tag{3.2}$$

- *Laplace distribution:* The motivation for choosing this distribution is twofold. First, among the symmetric disributions, it is one with higher kurtosis value. Second, through visual inspection histograms have been observed that display this distribution's distinct peak, especially in those containing high amplitude samples. The relevant parameters of its PDF (Equation 3.3) are the location parameter μ and the scale parameter b.

$$f(x) = \frac{1}{2b} e^{\frac{-|x-\mu|}{b}} \tag{3.3}$$

- *Logistic distribution:* This distribution's main features makes it suitable for this study. It serves as a compromise between the Gaussian and Laplace distributions described above. Logistic distribution, while being bell shaped, it lacks the sharp peak of Laplace distribution, while having greater kurtosis than the Gaussian distribution. Its PDF is featured in Equation 3.4, with μ and s being its location and scale parameters, respectively.

$$f(x) = e^{-\frac{(x-\mu)}{s}} \frac{1}{s\left(1 + e^{-\frac{(x-\mu)}{s}}\right)^2} \tag{3.4}$$

3.3.2. Fitting process

To test the hypothesis presented in this work and ascertain the statistical nature of the SR, a software framework was developed using MATLAB®. The main analysis process is contained in this framework, along with other useful tools to handle and visualize the produced data.

The goal of this main processing stage is to determine which one of the target distributions describes the data best, commonly known as distribution fitting. The mathematical tools applied in this step of the methdology were described in Section 1.4.4. The steps are as follows:

1. The MLE parameters that best describe the analyzed segment for each target distribution are calculated.

2. AIC is applied to the MLEs calculated in the previous step to all distribution candidates.

3. Akaike weights are calculated from the AIC results above, to be able to compare between the distribution candidates.

4. Comparing Akaike weights results, the best candidate among the target distributions (the one that best describes the data) is chosen.

Figure 3.5 presents a diagram of this process.
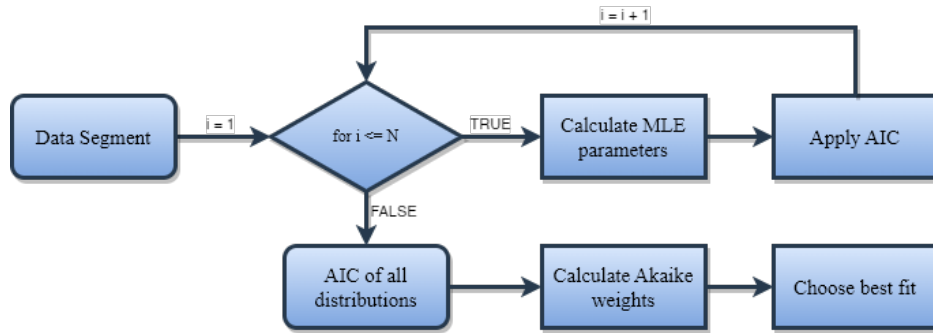
Figure 3.5: Flow chart of data segment fitting to N target distributions through MLE.

Once the fitting process is completed, the MLEs generated for each target distribution are stored as part of the segment information. These are necessary to plot the candidates of each target distribution in case visual inspection of the individual results are necessary. Even more so, they might be useful for future works. The key results of this process are Akaike weights, which are the values that allows us to determine which distribution is the best fit of a segment and are stored as well.

3.3.3. Edge cases control - bimodality

Before undergoing such procedure it is important to identify any segment that, prior to the analysis, can be discarded for not belonging to any of the target distributions. This is the case of any segment that displays two maxima, such as the one depicted in Fig 3.3a. Distributionwise, this shape is coincident with the so called Bimodal distributions, which are defined mathematically as an ensemble of two normal distributions with different mean values. That being said, not all segments whose histograms show bimodality (two distinct maxima) should be discarded, since if the two maxima are close enough, they may produce

an acceptable fit under any of the chosen target distributions. In other words, the goal is excluding Bimodal segments from the classification before it even begins, but keep the ones that could produce an acceptable fit. It was then necessary to detect the segments with an accused degree of bimodality - meaning those whose two maxima are apart enough (indicating a strong sinusoidal interference) from each other to be inappropriately classified. To do so, the method presented in flow chart format in Figure 3.6 was developed.
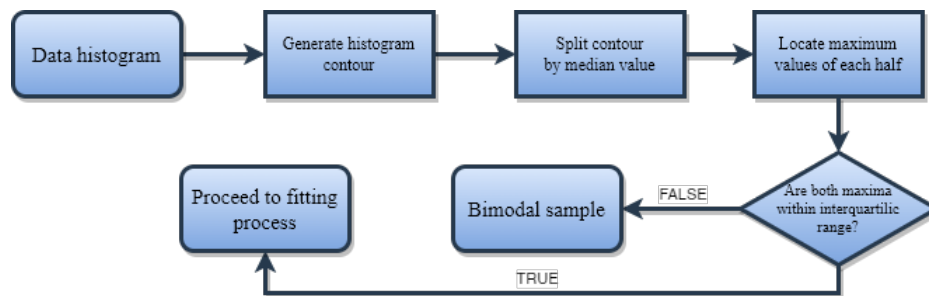
Figure 3.6: Flow chart of bimodality detection algorithm.

As it is, the process would happen as follows:

1. A data point is generated per histogram bin, its abscissa value being the histogram's height and its ordinate value matching the center of the bin. This net of data points serves to evaluate the histogram's contour.

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 located, then analyzed to evaluate if both abscissa values are contained within a predefined interquantilic range, chosen by attending at the system requirements.

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

The range mentioned in item 3 should be carefully chosen to compromise between inadequate segment rejection and saving those segments that could still be classified. Through many tests and visual inspection of the results, the range between the 3rd and the 5th octile has been determined as appropriate.

If the tested segment is deemed as bimodal according to the criteria exposed above, it will be labeled as such (assigning Bimodal as its best fit) and next segment will be selected. On the other hand, if the segment is deemed as non-bimodal it undergoes the distribution fitting process explained in Section 3.3.2.

3.3.4. Edge cases control - Quality of fit

After the distribution fitting process, it is also necessary to evaluate in some way how well the distribution chosen as best fit for a segment actually fits the data. This is since because the best fit distribution calculated through AIC and Akaike weights is only the one that better explains the data from all the chosen target distributions. This, in turn, does not guarantee that the chosen candidate can describe the data faithfully; if all fitting attempts are bad, the algorithm will pick the less bad among all of them. This points out the need for a metric to quantify the *Quality of Fit* (QoF) of each target distribution. The diagram below (Figure 3.7) is a flow chart that shows the steps followed to produce this metric, based on the concept of Q-Q plots.
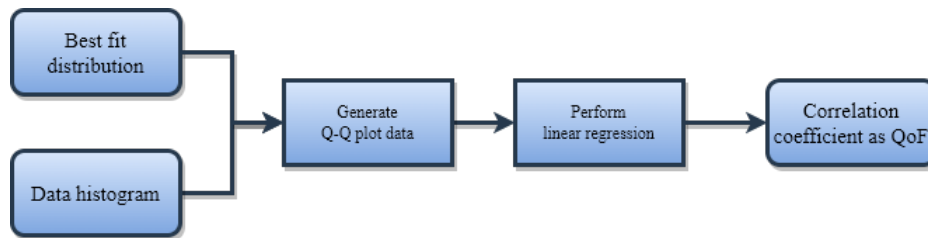
Figure 3.7: Flow chart of quality of fit coefficient calculation.

A Q-Q plot (short for Quantile-Quantile plot) is a graphical method used for this very purpose. On a Q-Q plot, the quantiles of a given dataset (considering the quantile position of each data point) are plotted against those of a chosen statistic distribution (dividing the distribution in the same number of quantiles as data points the tested data set has). Since they are used to evaluate the relationship between the dataset and the distribution, it was decided to rely on them to develop an analytical method that helps ascertaining the QoF of each segment. A perfect fit from the data to the distribution is recognized on a Q-Q plot as a segment coincident with an identity line. Consequently, the QoF coefficient is obtained

by taking the correlation coefficient value after performing a linear regression between the data points presented on a Q-Q plot, taking in consideration the segment and the evaluated distribution. 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.

The QoF coefficient obtained has the additional advantage of giving additional evidence towards the reliability of each model as data descriptor. Therefore, it provides a way to quantitatively evaluate how well each target distribution describes the data and, coupled with Akaike weights, they give further evidence towards the adequacy of the chosen models.

3.3.5. Pattern analysis

An extremely useful tool to detect trends in temporal series is the *Autocorrelation Function* (ACF). The basic concept is the same as a classic correlation, but instead of finding the relationship between two variables, it is focused on finding the relationship with itself while being shifted forwards in time. The ACF that will be applied to the data in this thesis is the discrete form, as displayed in 3.5.

$$R_{yy}(\ell) = \sum_{n \in Z} y(n)\overline{y(n - \ell)} \tag{3.5}$$

This way, ACF is performed by correlating a signal with itself, then shifting forwards one sample, and performing the correlation again until the shifting reaches a certain threshold. The resulting series $R_{yy}(\ell)$ represents then the trends detected in the signal.

3.3.6. Overall description of the methodology

Prior to the analysis, the time series register to be analyzed (total duration of $30\,\text{min}$) is filtered and segmented as described in Section 3.2.

When pre-processing is finished, each of the produced segments is checked for bimodality (Section 3.3.3) and if the criteria for the edge case are not fulfilled, the segment undergoes the fitting process (Section 3.3.2). After that, the QoF of each target distribution to a given segment is calculated (Section 3.3.4).

At a certain point, the need to discard the segments that did not meet the QoF requirements was deemed necessary. Consequently, an optional step was added in which the segment that did not meet the QoF threshold was deemed as Unclassified, while those that did meet the threshold kept their fit. If the analysis is run with the quality evaluation active, all Bimodal segments are considered Unclassified.

Once the fitting and quality evaluation is done, each of the produced segments is analyzed separately, characterizing it statistically through the extraction of the first fourth central moments (namely mean, standard deviation, skewness, and kurtosis). Finally, results are stored in a database.

Figure 3.8 presents an overview of the way data is processed, to offer a step-by-step view of the whole analysis.
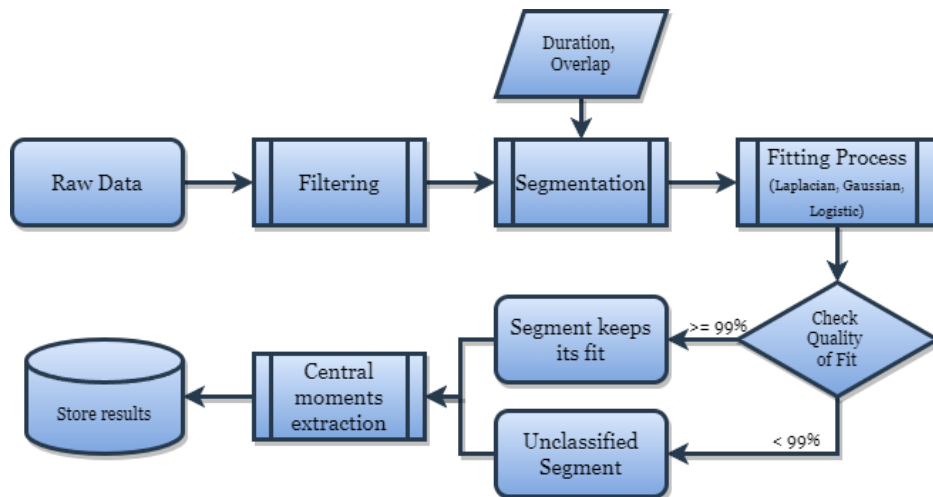
Figure 3.8: Flow chart of data processing for the presented methodology.

3.4.   Schumann resonance's variability estimator

Next, a detailed description of the method developed to quantify the variations experiences on the SR over time will be provided.

Although the quantification technique described in this section differs considerably from the statistical classification methodology explained above, they are both based in the same concept and developed under the same pretense; the possibility of extracting additional information from SR temporal series by the application of appropriate statistical methods.

3.4.1.   Theoretical framework and specifications

It is worth highlighting that calculating the statistical variation of the SR it is not meant to imply that the SR is a purely stochastic process, only that it can be treated as such up to some extent.

Evidences of the many different patterns found in SR are widespread in the literature; namely the consequence of variables that affect the signal and are stable considering the timescale of the resonant phenomenon. Ionospheric height and conductivity play a definitive role in the frequencies and average intensity of the signal [20]. Disregarding anomalies caused by specific events, the ionosphere displays a daily cycle, which hardly influences the signal in the minute timescale. The solar cycle, which in turn influences the ionosphere conductivity, is another example of a stable pattern of a long period - 11 years - with a measurable effect on SR [22]. Lastly, the major storm centers display seasonal and diurnal patterns as well, having a peak of activity between 14:00 and 15:00 at the local time and being more active in the northern hemisphere's summer months [23].

That being said, on the shortest timescales the primary factor is the lightning discharges which are usually modeled as a stochastic process following the Poisson distribution [28]. By statistically comparing temporal segments of the SR, the total contribution of lightning during that time is what determines the resulting statistical distribution. In consequence,

by calculating the differences between several realizations of the time series mainly the stochastic source of the process is being considered.

When discussing the stability of a signal considering it a stochastic process, each realization of the signal is observed as a statistical distribution through its histogram - in a similar way as the one discussed for the methodology for SR statistical classification in Section 4.1. In this particular case, the time series is transformed into a certain amount of histograms by considering two parameters; the segment span - its duration - and the time passing from a segment's starting point to the next - the amount of time segments overlap with each other. The realization's sensitivity to transient events can be controlled using the segment length, whereas segment overlap limits the amount of new information that each segment brings, in relation to its two neighboring segments.

3.4.2. Statistical divergences and distances

There is a wide selection of generalized metrics whose purpose is to quantify the differences between statistical distributions. These measurements are theoretically developed in the field of statistical data processing [165] and are a most relevant tool in information theory [149], for example in the development of speech recognition techniques. One of the best known is the Kullback-Leibler divergence [166], which is central to information theory, and is also related to maximum likelihood estimation [167] and even in optimization problems for power spectral density estimation for stochastic processes [168]. That being said, the Kullback-Leibler divergence and other general metrics widely used or developed in the mentioned contexts have two important drawbacks.

1. Although used as such, they do not qualify as metrics. For example, Kullback-Leibler is not symmetric and does not satisfy the triangle inequality. This is a problem in this context, since the lack of these or other characteristics associated with metrics would make the resulting values harder to understand.

2. They involve computationally intensive calculations. Since the goal of quantifying the SR variations is to provide support or additional information to other analysis, introducing it as a costly algorithm defeats its purpose.

To provide a methodology able to meet the requirements set, something simpler, less computationally intensive, and more straightforward is needed. For these reasons, KS distance was chosen. It is important to highlight that it is the statistic of the KS test that was chosen instead of the p-value. The latter estimates from the statistic the probability of both realizations coming from the same distribution, which could be argued is a measure of similarity, but it has some setbacks such as being dependent on the number of samples.

3.4.3. Analysis' specifics

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

1. The segment's length in time.

2. The analysis' temporal resolution by means of the time each segment shares with its neighbors.

3. The time slot for which variability shall be considered, defined by the time interval from which neighboring segments will be compared against the target segment.

A graphical representation of the concepts involved in the analysis is shown in Figure 3.9.

The scope and reach of a single variability test can be defined through these three variables. Increased time can be thus analyzed either by increasing the time slot or by sequentially repeating the process for all the time slots contained between two selected dates. The test itself is performed by segmenting the chosen interval of the signal after choosing the time slot size, the segment length, and the temporal distance between segments. Then, a specific segment is selected and its KS distance with all segments is calculated. The resulting curve is a representation of the statistical variation under the defined conditions.
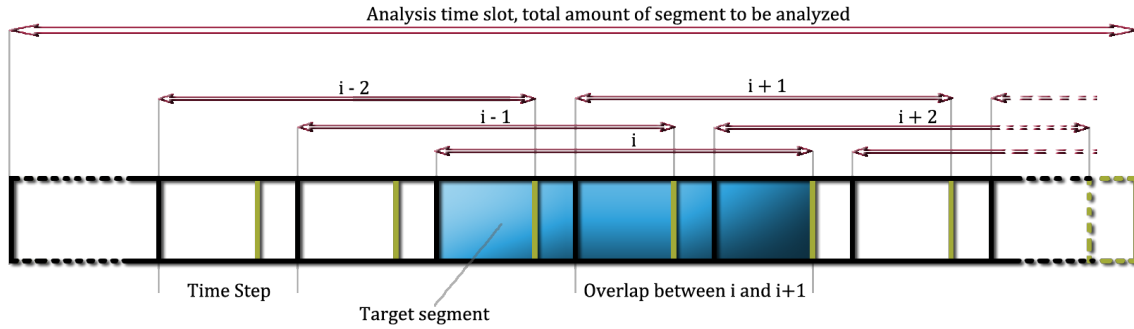
Figure 3.9: Representation of time series slot, already segmented in N parts, each with a specific length and with a specific time between them.

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

Chapter 4

Results and Discussion

In this chapter, the results of the described methodologies will be explained. First, the outcome of the test run of the statistical classification methodology will be presented, exploring the central moments obtained and their possible uses as well as the classification resulting from the fitting process. Second, an extensive study to validate the findings obtained during the methodology's test run is shown, along with a better characterization of the implications of segments' classification. Third and last, the results of measuring statistical differences using Kolmogorov-Smirnov distance among the segments produced by the methodology are discussed.

Part of the contents presented in Section 4.1 and Section 4.2 belong to the results of the article that ratifies this thesis' scientific quality titled *A New Way of Analyzing the Schumann Resonances: A Statistical Approach* [169].

4.1.   Statistical classification of the Schumann Resonance

To test the methodology described in the previous chapter, it was applied to ELF data from April 2016 captured in our research group's observatory located in Sierra de los Filabres (Almería - Spain). To test the methodology thoroughly, different settings are explored, with

four segment lengths (10 min, 5 min, 1 min and 20 s) and all five filtering processes described in Section 4.1. QoF threshold was set at 99 %.

4.1.1. Central moments

The first step in the methodology's evaluation will be an overview of the central moments in different cases, focusing on a few examples to highlight the prospective usefulness of the analysis.

Mean and standard deviation

Figure 4.1 shows the means of all the 10-minute segments analyzed, taken from the filtered data that represents SR best. Given the clear lack of normality displayed by the graph, it can be stated that SR samples are not independent since they don't abide by the Central Limit Theorem. One may even argue that the large amount of samples contained within the first range of standard deviation (approximately 98%) is an indication of the antipersistent nature of the signal, in agreement with [131].
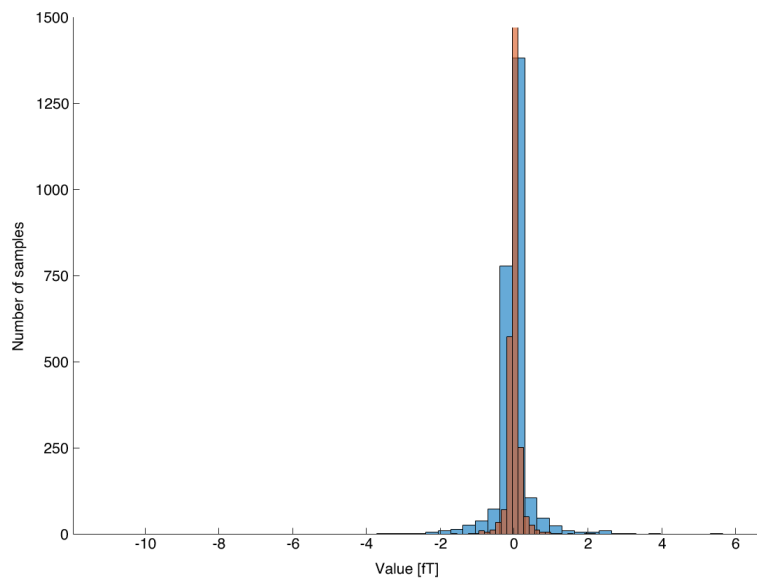
Figure 4.1: Mean histogram of the Schumann resonance with two different filters applied. Blue = HPF and notch filter, red = BPF. 10-minute duration.

Among the displayed means, there are a few extreme values that are worth checking, since they may give additional insight into specific phenomena or just point to an odd segment of data that should be excluded. Figure 4.2 depicts the time signal with the most extreme absolute mean value in Figure 4.1. It can be observed how the time series is populated with great variations in amplitude, accounting for additional disturbances that were not filtered.
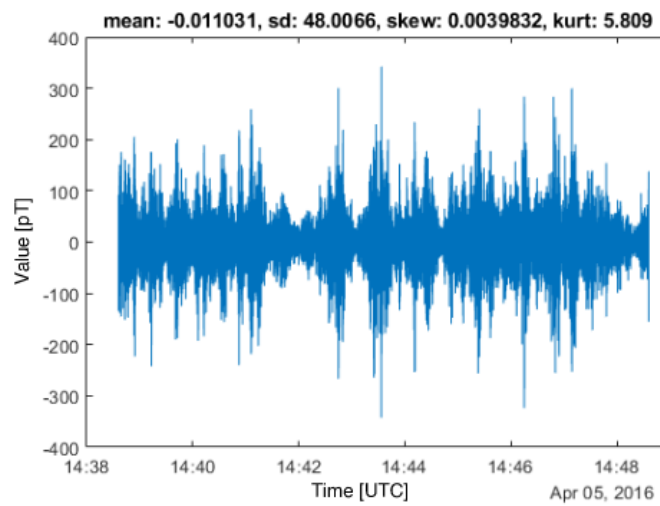
Figure 4.2: Time series of the segment with the lowest mean among the 10-minute High pass and Notch filtered segments.

These kinds of observations may be verified with the value of other central moments. Figure 4.3 shows the standard deviation histogram for the same kind of segments being studied. Since Figure 4.2 shows the segment standard deviation as well, it can also be inferred this sample contains odd data, given the values of amplitude reached. To properly identify the event that triggered this anomaly, additional data would be necessary.

Skewness

It is expected for the skewness of all 20 seconds segments to display a higher standard deviation than the longer ones, being more influenced by transitory effects. Nevertheless one might wonder if, in the case of SR, segment duration affects skewness in some other way - say, for example, increasing the appearance of extreme values.
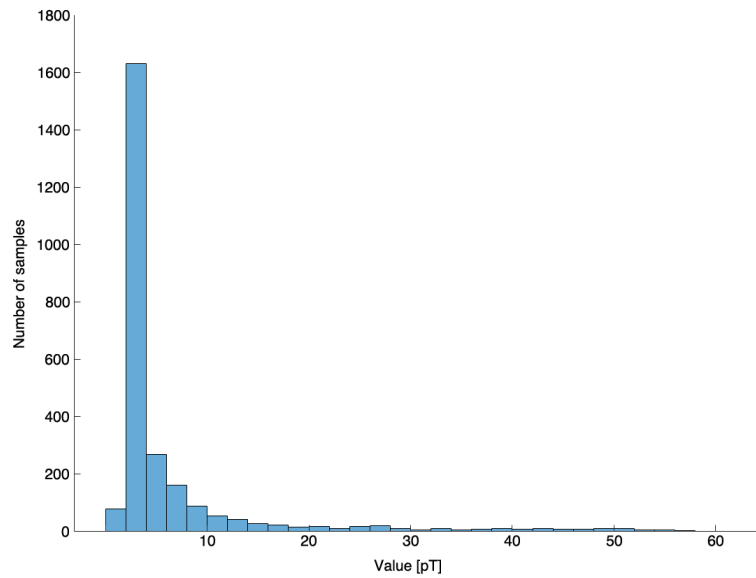
Figure 4.3: Standard deviation histogram of the Schumann resonance, 10 minute segments with high pass and notch filter.

In order to do so, Figure 4.4 displays the band pass filtered segments' skewness of different duration, and the data contained within one standard deviation was calculated. Thusly, the analysis method provides a fast way to evaluate these kinds of effects before doing further testing. In this case, given the small differences between the values obtained for different durations, there is an initial predisposition to consider that segment duration doesn't produce more dramatically skewed data.

Kurtosis

A comparison between the kurtosis of raw data and each of the filtered data is shown in Figure 4.5, in order to grasp the effects of removing certain frequency components on this central moment.

Since there are a few samples with kurtosis greater than a hundred, these graphs have been cropped in order to fully appreciate the differences in the main part of the histogram. Looking at Figure 4.5a we can see how for HPF data's kurtosis values are reduced when compared with raw data. Nonetheless, this filter's data is overly populated with Bimodal
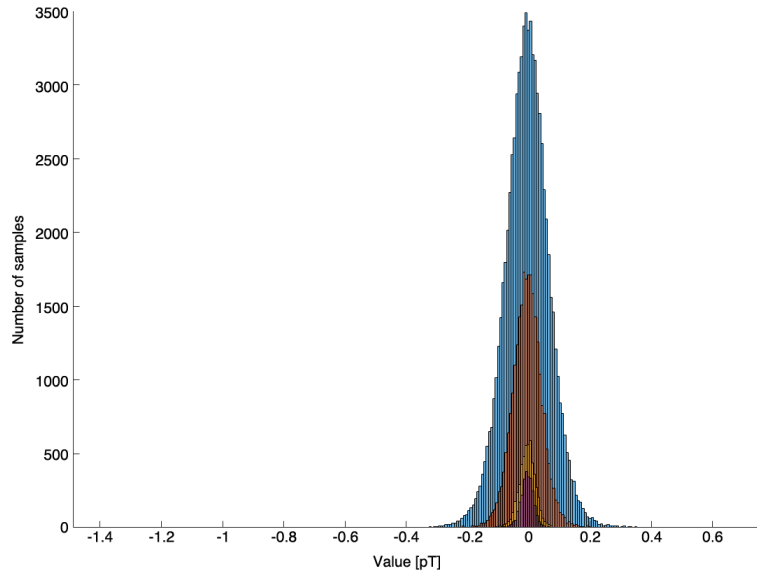
Figure 4.4: Skewness of BPF segments of different duration, with the % of data contained within one standard deviation. Blue: 20 seconds, 71.34%. Red: 1 minute, 72.27%. Yellow: 5 minutes, 72.28%. Purple: 10 minutes, 72.18%.

segments, so this must be taken with a grain of salt. On the other hand, Figure 4.5b shows how the notch filter's effect on kurtosis may be considered negative since the values are way more spread than in the raw histogram, as the bin size indicates. Lastly, Figure 4.5c and Figure 4.5d show how both filtering processes reduce kurtosis dispersion (slimmer bins), but the latter manages to do so better than the former. This can be concluded from how the most prominent histogram bars concentrate on the same values as the tallest histogram bar for the raw data.

Segmentation overview

One of the ways to corroborate the segmentation criteria is to study the standard deviation of the segments' mean. This value is an estimate of the uncertainty caused by the disturbances of the signal so it can be considered as an indication of how important transitory events are on the segments of a certain length. These results are presented in Table 4.1.

The way dispersion increases as segment duration decreases illustrate how random error can be compensated by increasing the number of samples per segment. It is also evidence of

(a) high pass filter.

(b) 50Hz notch filter.

(c) High pass and notch filter.

(d) Band pass filter.
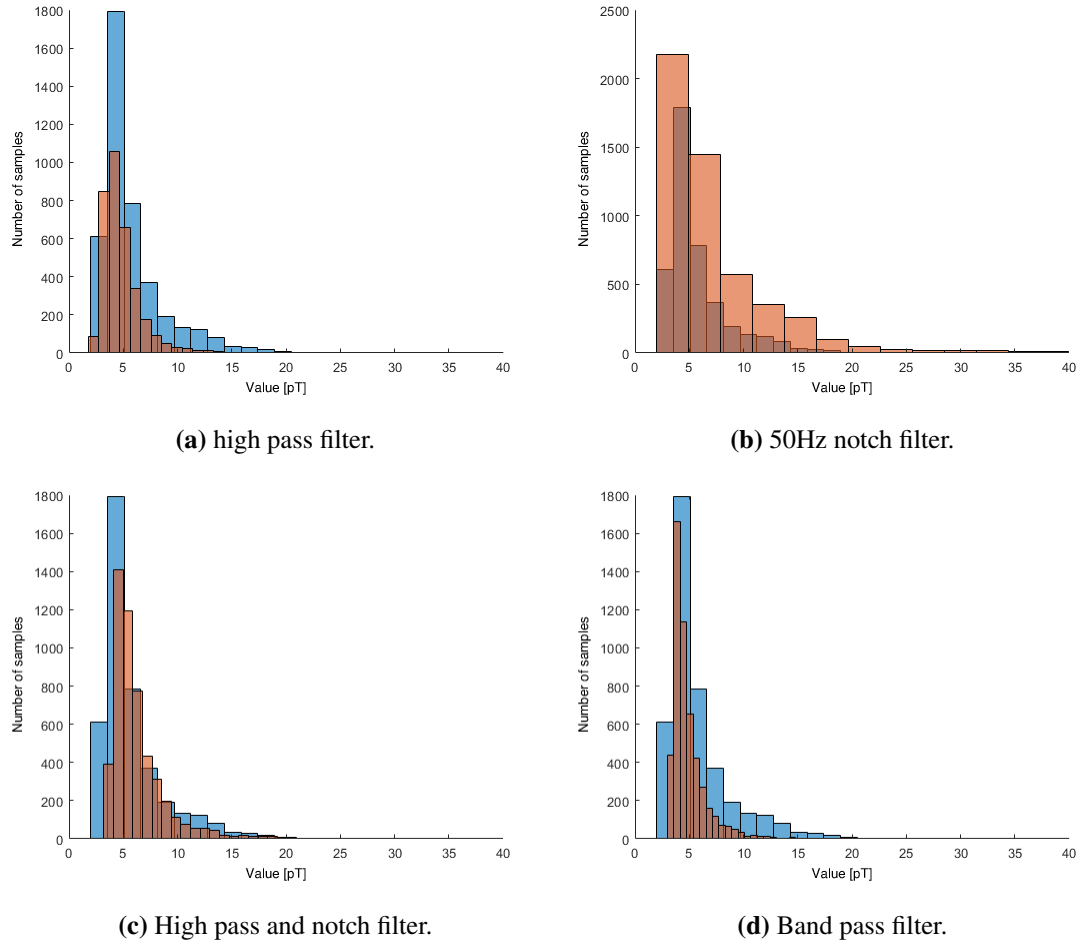
Figure 4.5: Kurtosis comparison between raw data (blue) and data processed under different filters (red). Cropped graphs (see text).

Table 4.1: 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

one of the ideas expressed since the beginning of the thesis; that background noise dominates in the longest segments, producing more stable signals, with a dispersion around 8 times

smaller than the 20 s segments that stems from more samples with lower (close to the mean) amplitude value. Nevertheless, higher dispersion on shorter segments shows their sensibility to transitory events. Although this was to be expected, it is an interesting fact to point out since it reinforces the utility of the method; choosing different segmentation values allows the analysis to classify different phenomena present on ELF records.

Looking at the table by rows, we should note that values of Raw and HPF segments are quite similar. It shows how low frequency disturbances are more detrimental for the signal than the 50 Hz power grid signal, with the latter having next to no effect in the uncertainty compared with the former. This is attributable to the nature of each disturbance. Whereas the power grid is stable for long periods, since the observatory isolation guarantees the measured signal is related to general power consumption, low frequency disturbances are related to environmental effects which have much greater uncertainty.

This detrimental effect can be seen as well in the small difference between HPF and notch and HPF filtered segments' values. In both, meteorological disturbances affect the sensor and hence, produce an increased statistical deviation. 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.

4.1.2. Best fit classification

Table 4.2 consolidates the outcome of the whole analysis process, grouped in sub-tables by segment duration.

Each row on each sub-table presents the number of segments identified as any of the possible distributions (columns) after applying the specified filter to the data. This value is expressed under Ratio, indicating 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 number of that distribution's segments whose QoF is over 99 %

Table 4.2: 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	0		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

to the total amount of segments classified under that distribution. The Bimodal distribution column does not come with a QoF ratio since its classification is not the product of a fitting process.

A glance over these two parameters gives pertinent information about the methodology's performance. Following what was exposed in the previous section, distributions whose acceptance ratio shows a general decrease as segment length is reduced may be associated with the main frequency component of the measured signal. On the other hand, the distributions that experience a slight increase or otherwise remain stable indicate the presence of transitory events.

Regarding QoF ratio, it allows the researcher to consider how well each target distributions represent the phenomena they fit. Hence, it points out how likely is the distribution to be present in ELF records and how accurately certain distribution models the behavior present in the fitted segment.

Each distribution's results are worth being commented on their own, highlighting the most important values and discussing their implications.

Bimodal

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 their ratio reaching 1 % only in 20 seconds segments. Its appearance on Raw segments and on those filtered with HPF, as well as their absence on the rest clearly shows the relationship between Bimodal segments and the presence of sinusoidal signals.

Gaussian

Gaussian 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. The elevated QoF ratio points out that most of the ELF segments fitted under this distribution are appropriately classified. Its ratio of positives is only significant under HPF

in the 10 min and 5 min duration segments. Nonetheless, it experiences an overall increase in the lower duration segments, especially under BPF. This confirms that the Gaussian distribution is an acceptable model for certain SR segments, whose circumstances are still to be discovered. Even more so, their proliferation in lower duration segments is an indication of a relationship with transitory events since their prominence in the histogram increases the shorter the segment is.

Rician

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

Laplace

Laplace distribution displays the lowest acceptance ratio, the maximum value being 68.87 % for 10 minute segments under BPF. This evidences that a considerable amount of segments classified as Laplace's are actually not so, but lacking a better fit among the selection of target distributions, it is still the best fit. Despite having a high ratio of positives under 50 Hz 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. The way this distribution's ratio of acceptance is spread seems to point out that Laplace fitted segments represent mostly the overlapping of different phenomena; either non-sinusoidal disturbances or transitory events along with ELF background noise.

Logistic

Logistic distribution segments are present in all filtering processes, its ratio of positives being especially 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. Their overwhelming presence under these two filters, along with a high QoF ratio places this distribution in the same situation as Gaussian's. It can be safely stated that ELF records display Logistic behavior, and even more specifically on the SR themselves due to the high occurrence ratio under the processes that best represent the resonances.

4.1.3. Distribution fitting

In Figure 3.2b and Figure 3.3b it was shown how high power, low frequency noise masks the power grid signal on its histogram representation. Comparing Raw 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 50 Hz 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 25 %. On the other hand, after going through the 50 Hz 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's fits. Their moderate ratio of positives in Raw analysis points out that this particular distribution is also related to disturbances and their high ratio under 50 Hz notch filter points at low frequency contributions from the 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. Figure 4.6 illustrates this case; it shows a segment with a 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.

Data histogram.

Time series.
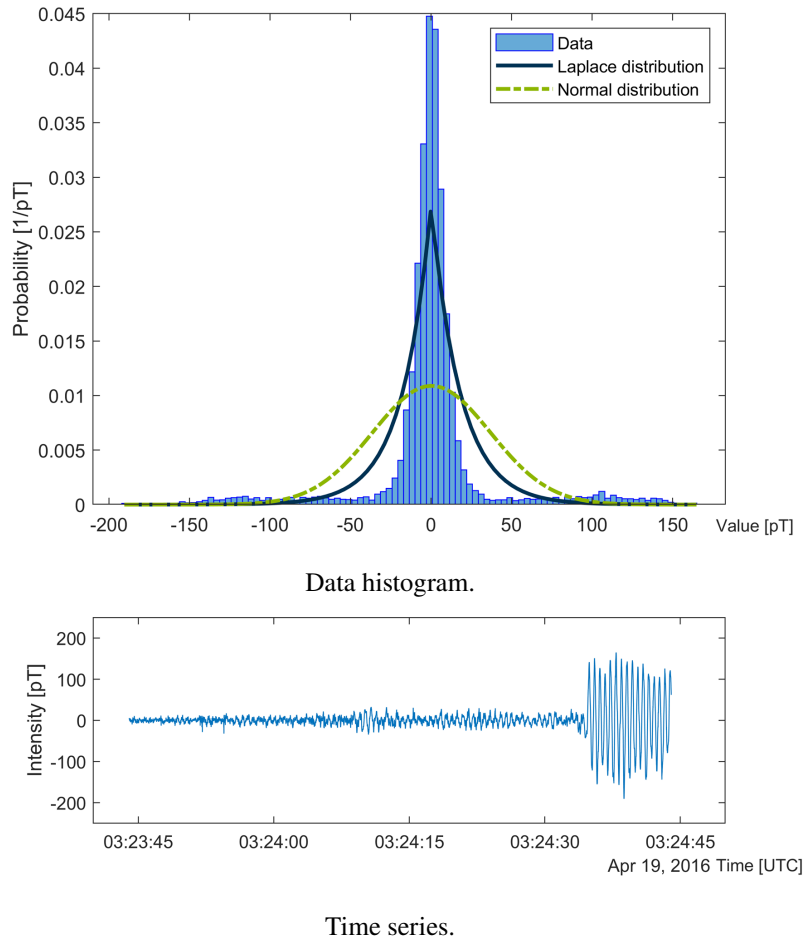
Figure 4.6: Notch filtered segment from NS, 19-Apr-2016 03:23:44, 60 seconds duration. Low QoF Laplace fit due to high intensity disturbance.

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, as it was outlined before. 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. Figure 4.7 depicts one of these cases.

Something similar happens with Logistic distribution fits. The average 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

Data histogram.

Time series.
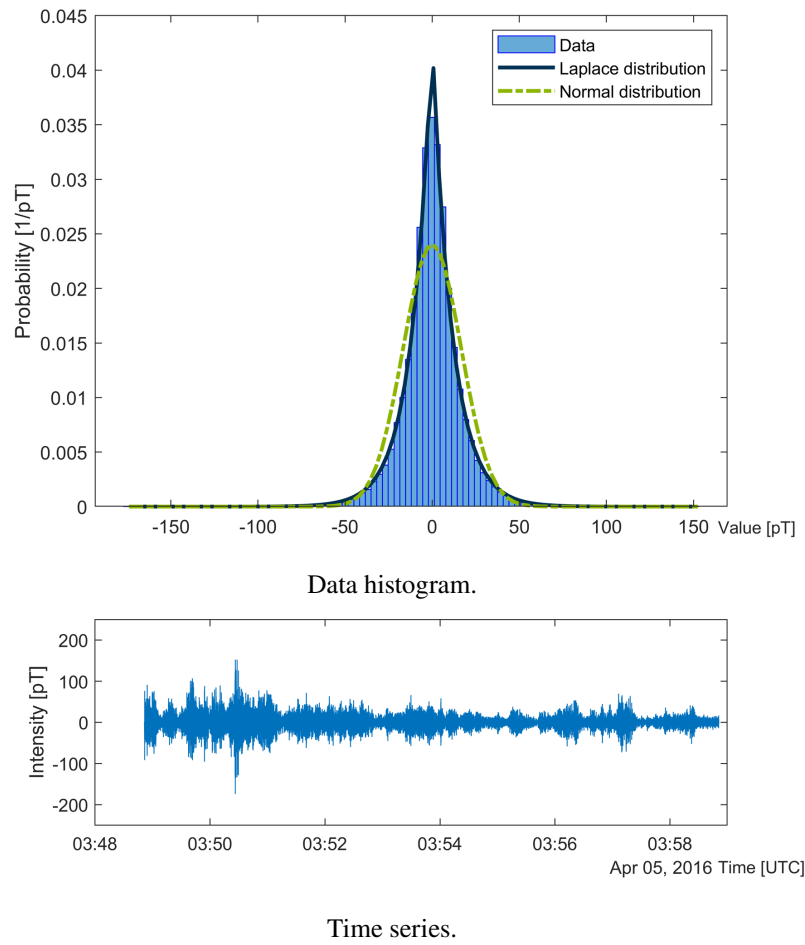
Figure 4.7: Band pass filtered segment from EW, 05-Apr-2016 03:48:51, 600 seconds duration. Accepted Laplace fitted segment.

accepted as well. This raises an issue regarding data interpretation, since certain samples that otherwise would be acceptable fits must be discarded. Nonetheless, the goal of the methodology is to characterize the time segments; this kind of samples should be analyzed further to deepen the classification in a later stage.

Figure 4.8 shows a case where high intensity disturbances introduce extreme values in the histogram but, due to higher segment duration compared with Figure 4.6, its impact is lower.

Even when the Rician distribution was promising, being one of the target distributions for which a prospective conceptual model was introduced in Chapter 3, the ratio of Rician distribution fits is only meaningful in the 1 minute and 20 second segments. Nonetheless,

Data histogram.

Time series.
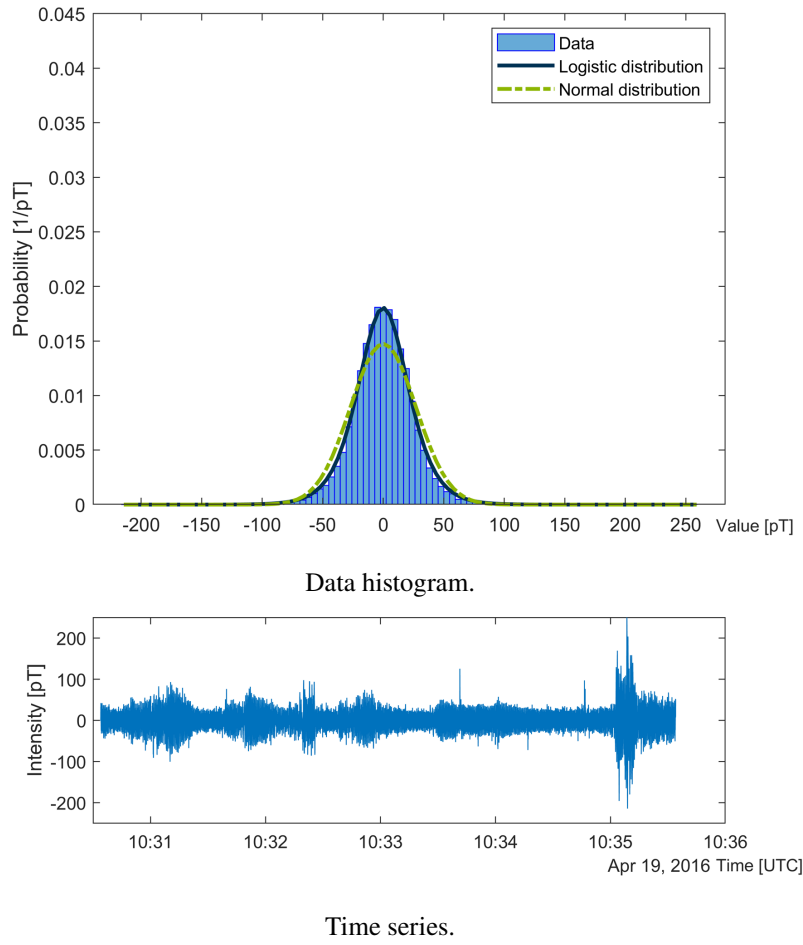
Figure 4.8: Raw segment from NS, 19-Apr-2016 10:30:34, 300 seconds duration. Rejected Logistic segment with QoF over 95%.

their ratio of acceptance is insufficient to validate the possibility of modeling lightning impact in the SR through this distribution. This conclusion is reinforced by the fact that Akaike weights' values for Gaussian and Rician distributions are both significant on Rician fitted segments. This can be double-checked by means of visual inspection; most Rician fits show that Gaussian distribution would have been an equally acceptable fit. Figure 4.9 is a representative example of most Rician fits, where the Gaussian distribution overlaps with the Rician distribution representation.

Lastly, Gaussian 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 of HPF data on the long duration segments,

Data histogram.

Time series.
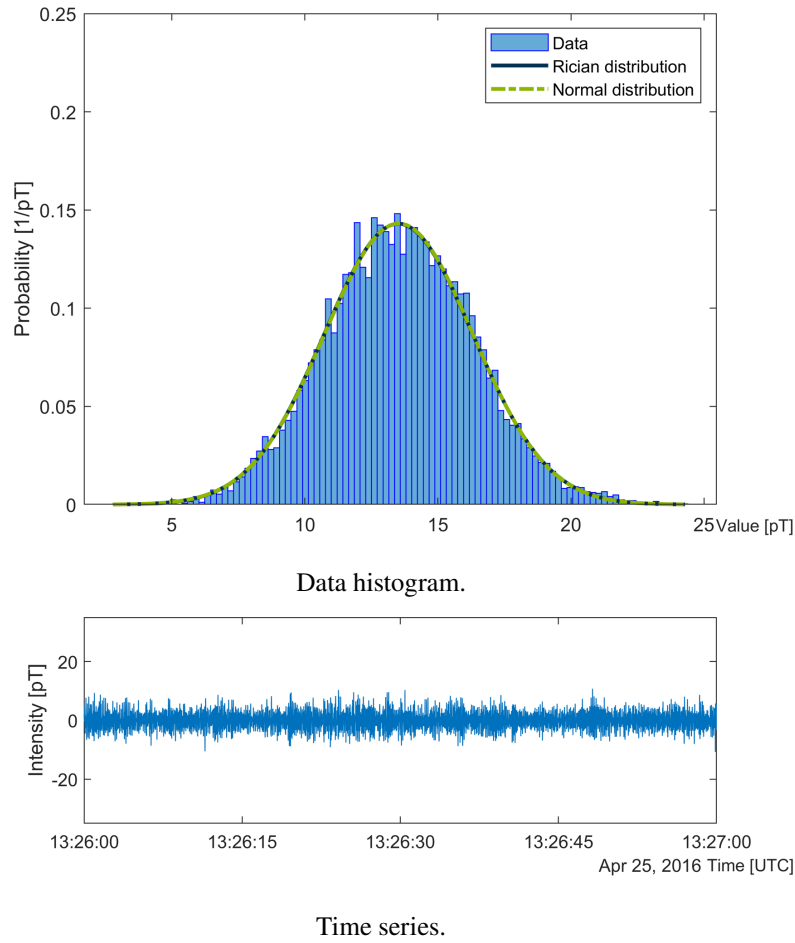
Figure 4.9: Band pass filtered segment from EW, 25-Apr-2016 13:26:00, 60 seconds duration. Accepted Rician segment.

but they are also present in the shorter segments of BPF data. Looking at the number of segments whose best fit is a Gaussian distribution, it is understandable why SR are usually identified with it. Nonetheless, they never overcome the % of Logistic fits.

4.1.4.   Schumann Resonance's statistical characterization.

To properly characterize the SR using the statistical classification described in this thesis, the attention must be shifted to HPF and notch and BPF rows, being the ones that more effectively isolate the SR signal from the rest of ELF noise. 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 [131].

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 Gaussian and Rician distributions.

Figure 4.10 presents an example of how Logistic distribution fits segments free of disturbances, with Figure 4.10a coming from BPF data and Figure 4.10b from HPF 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, Figure 4.10a has a QoF value over 99% whereas Figure 4.10b does not. 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.

Data histogram.

Time series.

(a) Band pass filtered segment.

Data histogram.

Time series.

(b) High pass and notch filtered segment.
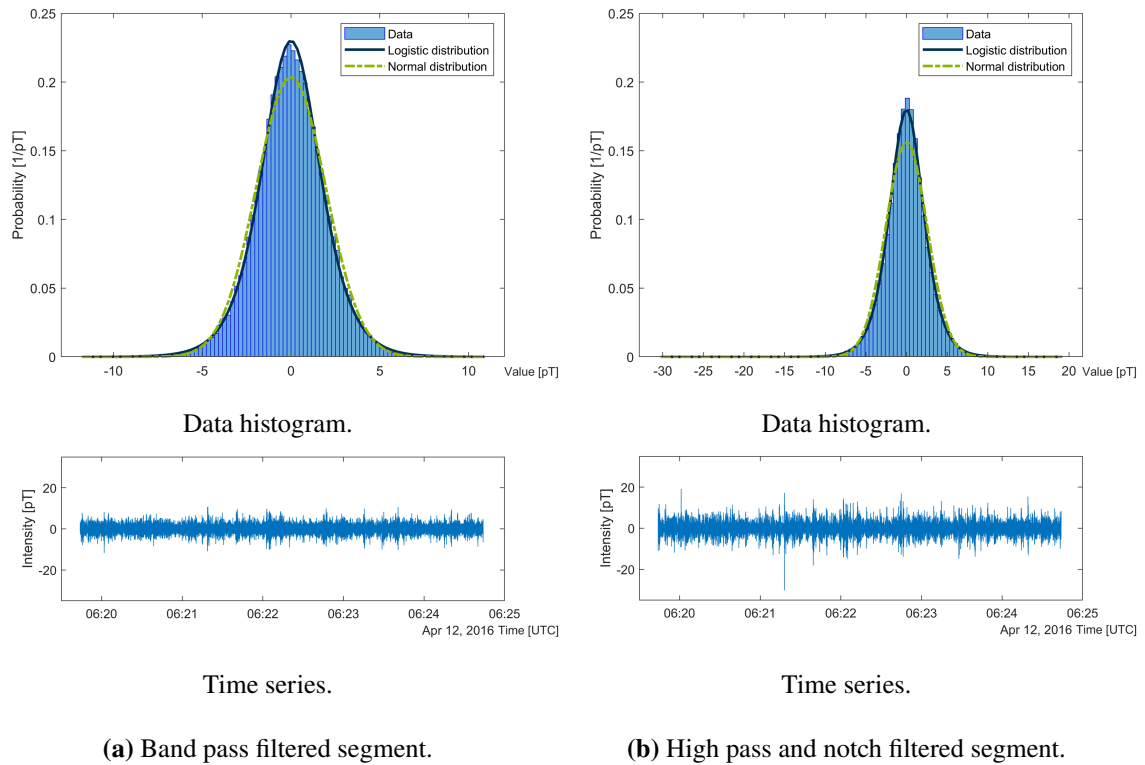
Figure 4.10: Segment from EW, 12-Apr-2016 06:19:44, 300 seconds duration, under two different filtering processes. The Logistic fitted segment contains mainly ELF background noise.

To sum up, Figure 4.10 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 SR.

In order to further elaborate this point, Table 4.3 presents the analysis results separated by channel of reception, with Table 4.3a gathering the EW channel and 4.3b the NS channel. The first item to highlight is how Logistic distribution's 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.

Bimodal and Laplace distributions' ratios 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 4.3 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 in Table 4.2. This illustrates once again the importance of proper filtering before analyzing data.

4.2.   ELF Statistical analysis and atmospheric phenomena

This section will be dedicated to delving deeper into the implications of the previous section's results. First, an initial assessment of the significance of any given segment being

Table 4.3: 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

(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

classified under a specific distribution. Next, the first hints of one of these relationships will be studied, extracted from the results presented in the previous sections.

4.2.1. Distribution fitting and related phenomena

During the previous sections, some significant parallelisms were outlined between the methodology results and what could be specific phenomena measurable in ELF records. The most remarkable outcome is SR background noise being identified with the Logistic distribution. The overwhelming occurrence ratio of this distribution in long duration segments, and their moderate to high presence in shorter segments offer wide support to the previous affirmation.

Also, Laplace distributions were abundant especially under the notch filter processing, but were also present in Raw segments and on HPF and notch. Through visual inspection, it was confirmed that segments classified as Laplace distributions usually contained a great difference in amplitude within, associated mostly with low frequency disturbances under Raw and notch processes. Nonetheless, their presence on HPF and notch filtered segments indicate either a non filtered disturbance or a signal related amplitude increase, such as Q-burst presence. On the other hand, it must be said that these observations lack proper support and up to this point, are based only on individual observations done by the researchers.

One of the most revealing results is the presence of segments whose best fit is the Gaussian distribution on all durations. The ratio of Gaussian distributions under BPF process, especially in the shorter segments, and the high QoF ratio displayed under most processes for all durations (reaching even 100 %), points out to a relationship with an SR related phenomena. It has been suggested that the amplitude of radio pulses produced by lightning strikes follows a Gaussian distribution [123] so there might be a relationship between the time series' signal being represented by a Gaussian distribution and lightning activity.

4.2.2. Gaussian distribution ratio and lightning activity

Further results obtained in the previous analysis strengthen the hypothesis of moments of intense lightning activity being represented by the Gaussian distribution. The first soft evidence favoring this hypothesis comes from the results of Table 4.3. It can be observed how, for all durations, The Gaussian distribution' ratio is higher in the EW channel (Table 4.3a) than in the NS' (Table 4.3b). Our station's location and proximity to the African thunderstorm center makes the EW channel of our sensor especially sensitive to its activity [4]. The intensity of its influence might account for the difference in the Gaussian distribution ratio between EW and NS channels.

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 Gaussian distribution ratio than the previous while Bimodal, Rician, and Laplace distributions ratio stays low (see Table 4.2). The BPF analysis will be used since it is the filter that most clearly presents the SR signal. Results were arranged hourly by grouping the ratio of each distribution throughout the month by UTC hours of the day.

To start exploring this relationship, all the grouped samples were plotted by hours using boxplots. The interest in using this graphical method is the amount of information displayed in a single graph; the red line in the middle of each box being the median, the upper and lower edges of the box showing the interquartilic range (from quartile 1 to quartile 3) meaning the range where 50 % or the data is contained, and the dotted lines (usually called whiskers) being an indication of how dispersed is the data overall. The outliers (data points that are too far away from the main group) are displayed as red crosses. The results for both channels can be seen in Figure 4.11.

Figure 4.11a depicts the Gaussian 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 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, similarities can be traced between the noon crest on Figure 4.11b and the Asian thunderstorm's average intensity pattern, although delayed in time by two hours. Nonetheless, the samples for the rest of the hours in both Figure 4.11a and Figure 4.11b show a pattern unrelated with their respective most influential thunderstorm center.

(a) EW channel.

(b) NS channel.
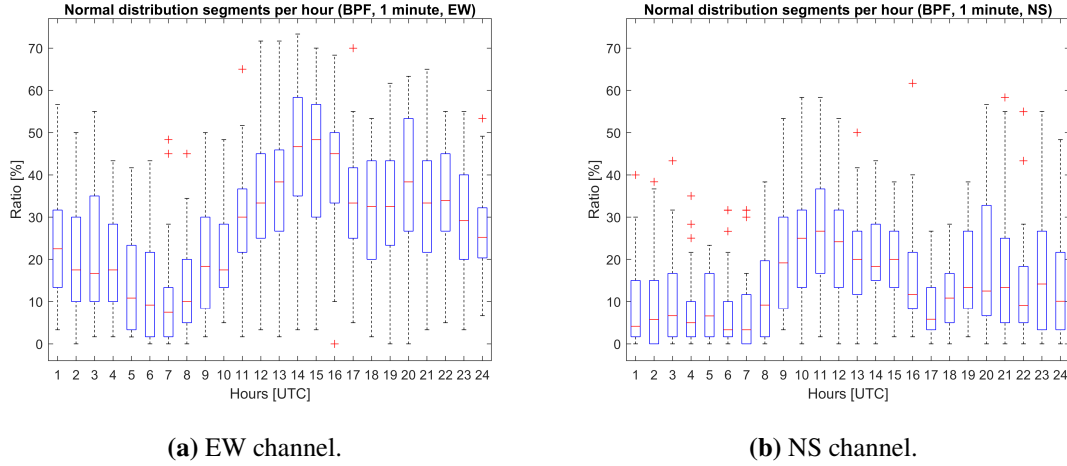
Figure 4.11: Gaussian distribution ratio of positives in 1 minute segments each day, organized by hours.

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 Gaussian distribution ratio and an average estimation of thunderstorm intensity for each thunderstorm center, extracted from [4].

This will allow to visually account for the similarities and differences between hourly Gaussian distribution ratio by days and thunderstorm activity. Figure 4.12 depicts the mentioned data, on which NS Gaussian distribution occurrence has been given a two hours advance in order to enhance its visual comparison with the Asian thunderstorm center.

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 for the behavior shown at all hours. Looking at how EW data follows both African and American thunderstorm activity with no apparent attenuation gives evidence for Gaussian distribution ratio to be related to global thunderstorm activity. Still, the effect of source-observer distance and orientation should be taken into account.
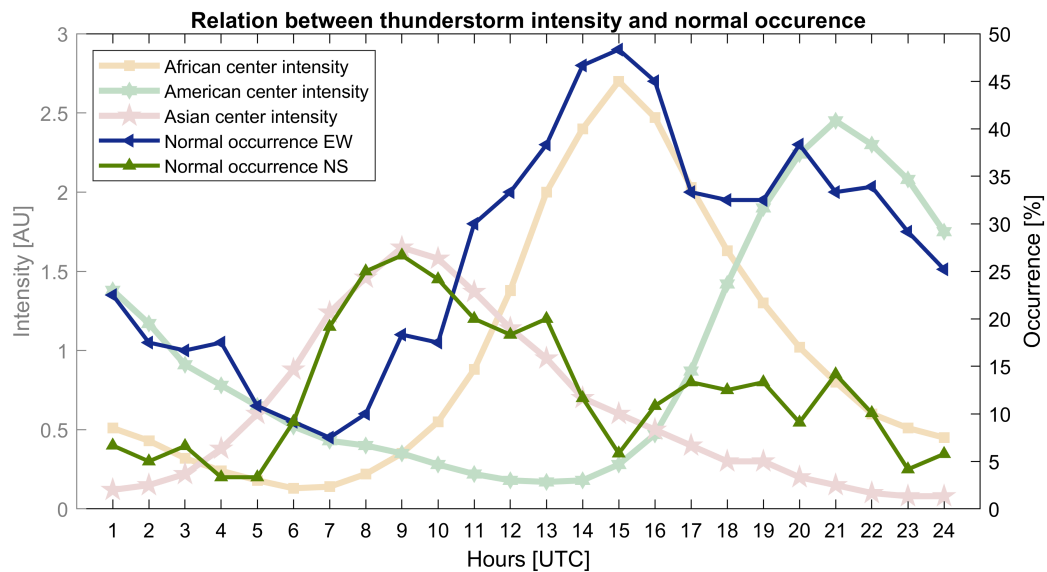
Figure 4.12: Thunderstorm intensity (faded lines, left axis) and median values of hourly Gaussian distribution ratio (solid lines, right axis). Gaussian distribution ratio in NS channel shifted two hours forward.

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 Gaussian 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 Figure 4.13.

Although data deviation is high, the regression coefficient points out a strong relationship between the hourly Gaussian 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 as the previous one. As Figure 4.13b shows, there is a strong positive correlation of statistical significance.

(a) Regression analysis for EW channel.

(b) Regression analysis for NS channel.
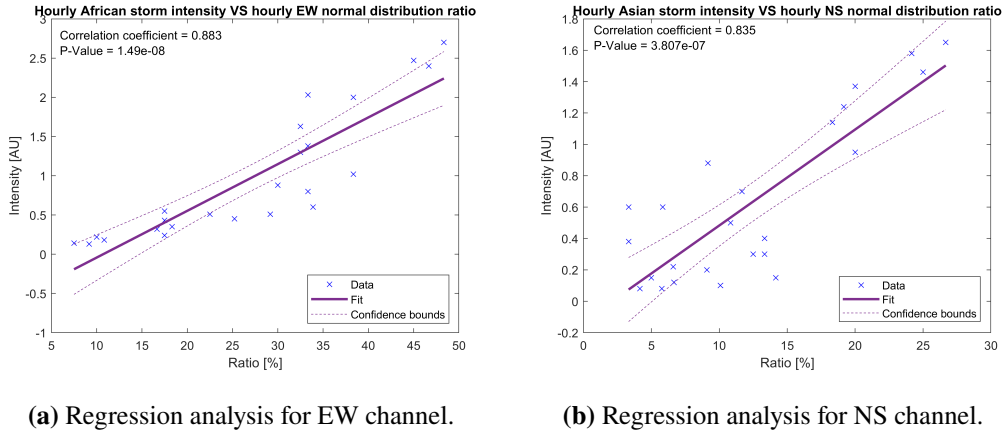
Figure 4.13: Linear regression analysis with Student's t-test results.

The conducted tests show interesting pieces of evidence favoring the hypothesis of the Gaussian distribution ratio being related to global lightning activity. Nonetheless, despite the good correlation and significance of the relationship, the two hours delay in NS channel remains to be explained.

These and other aspects will be dealt with in the next section.

4.3.   Methodology validation - extensive analysis

The relationship between the median of Gaussian (Normal) segments per hour and lightning activity presented above will be explored further in this section. To do so, the analysis technique has been applied to five years of ELF data. With the goal of characterizing the SR in mind, only the data filtered using the 6 Hz to 40 Hz BPF will be used. This frequency band contains the first six modes of the SR [170] and allows to remove low frequency disturbances and the 50 Hz electric grid signal.

The following in-depth analysis was performed using R, an open-source software widely used in the scientific community, focused on statistical analysis and providing tools to deal with the massive amount of data required for this analysis. Through this analysis, further relationships between the classification provided and the known facts and patterns on SR and lightning activity will be established. The main hypothesis to be tested is that SR's temporal segments where noticeable global lightning activity is present display a Gaussian distribution, as a counterpoint to the Logistic distribution shown by SR background noise. That being said, other patterns present in the analysis will be studied as well.

The overall results of the analysis will be presented first, as they are compared with the known general behavior of the studied signal. Next, different data aggregations will be shown to inquire about the implications and usefulness of temporal segments' statistical classification. Lastly, ACFs will be applied to the resulting data for Gaussian occurrence to compare the patterns displayed by the classified segments with the ones uncovered on the SR.

4.3.1. Hypothesis verification

To verify the exposed conclusions in the previous Section about lightning activity and Gaussian segment occurrence, the five years of data have been processed and presented in the same way as in section 4.2.2. The results are shown in Figure 4.14. Figure 4.14a and Figure 4.14b show the hourly average of Gaussian segments per year in the EW and NS channels, respectively. Figure 4.14 presents the hourly average of the five years for both channels with the intensity given to the SR by each thunderstorm center, extracted from [4].

It has remarkable similarities with the averaged lightning activity of the African thunderstorm center, such as peak value around 15:00 UTC for every year and a minimum value around 8:00 UTC. The main difference between years is the pattern depicted after 18 UTC. 2016 and 2020 display a downward slope that settles on a flat value, experimenting little variation from 20:00 to 0:00 UTC. 2017 and 2019 follow the same pattern but the amount of Gaussian segments are 20% higher than on the other pair of years and a decrease of barely
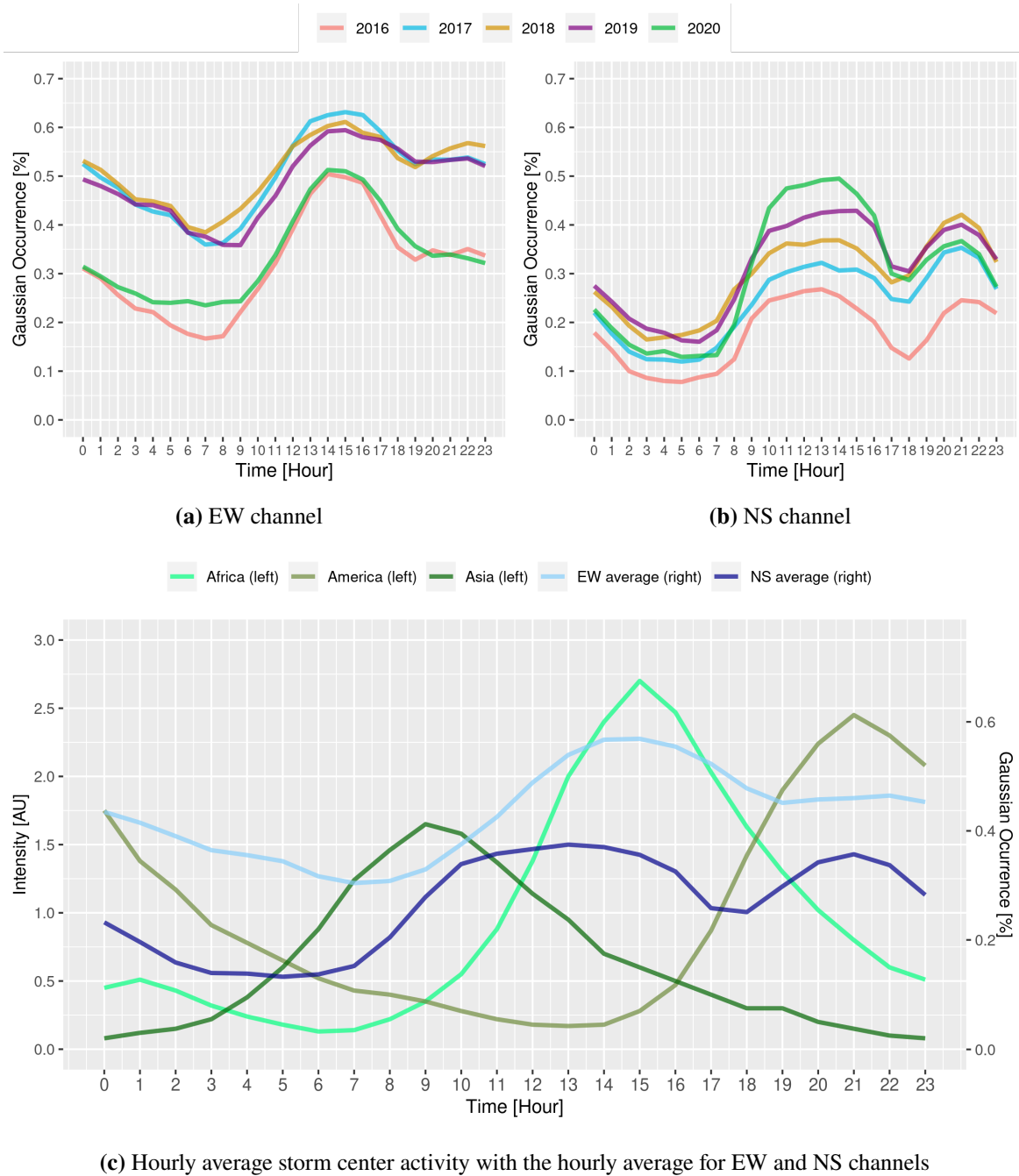
(a) EW channel

(b) NS channel

(c) Hourly average storm center activity with the hourly average for EW and NS channels

Figure 4.14: Hourly averaged fraction of Gaussian classified segments for each year (1 min length).

5% respect their maximum, whereas in 2016 and 2020 the difference is around 15%. 2018 even displays a slight increase at the end of the day.

In Figure 4.14b the pattern is also consistent, but in this case, there are remarkable differences between the hourly averaged values for each year and the one depicted in Section 4.2.2 which presents only a month. The pattern obtained is also consistent between years, the main difference being once again their average value. In the NS data the yearly average value is more dispersed, while in EW there are two groups, each with a similar average value. That being said, all years on the NS data show a plateau ranging from 9 to 16:00 UTC, albeit each showing different characteristics. The peak that starts at 18:00 UTC and has its maximum at 21:00 UTC is present in all years as well. For this part of the plot, the values are closer together, with two observable groups 2017 and 2020 in the middle, 2018 and 2019 with the highest peak value, and 2016 alone with the lowest peak value. There are differences as well between the maxima of the plateau and the peak; in 2016 and 2019 the peak maximum value is slightly lower than the plateau maximum value, whereas in 2017 and 2019 it is the other way around. The oddity in this data set is 2020, where the maximum peak value is noticeably lower than the maximum plateau value.

The consistent pattern displayed in Figure 4.14a by the hourly averaged per year data from the EW channel offers plausible evidence of the relationship between lightning activity and SR temporal segments belonging to a Gaussian distribution. The differences between the values reached each year reinforce the evidence as an indicator of intensity variations since the global lightning activity intensity is not constant throughout the years, although their daily activity patterns are. Now, Figure 4.14c compares the hourly average of both channels for all years with the average intensity that each of the storm's centers contributes to the SR. The latter is extracted from [4]. In the figure, it can be identified how EW Gaussian occurrence rises with the activity of the African thunderstorm center, while the peak during the late hours on the NS channel follows the American thunderstorm center. Nonetheless, other factors of the relationship are unclear, such as the contribution of the Asian thunderstorm center to any of the channels. In Section 4.2.2 two hours shifting was applied to the obtained values to highlight its similarities with the Maritime Continent's thunderstorm center. In light of this

analysis 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 (as was mathematically verified in the previous section). Confronted by this refutation, the issue must be explored further.

It must be explained first 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 [18]. 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 [94]. That being said, the equations that describe the horizontal magnetic field at the ELF range are mentioned in the literature [14] and were highlighted in Section 1.1. These equations are reproduced here for the reader's convenience in Equation 4.1 and Equation 4.2.

$$H_x(t) = H_{NS} = \sum_{k=-\infty}^{\infty} A_k \cdot g(t - t_k) \cdot sin(B_k) \tag{4.1}$$

$$H_y(t) = H_{EW} = \sum_{k=-\infty}^{\infty} A_k \cdot g(t - t_k) \cdot cos(B_k) \tag{4.2}$$

With A_k being the amplitude of the k-th lightning pulse, $g(t)$ the waveform of the magnetic field components, and B_k the k-th source bearing. Given the data depicted in Figure 4.14, The equations above imply that NS and EW channels are affected by total field intensity proportionally to the sine or cosine of source bearing, respectively. Nickolaenko *et al.* published a model for the monthly location of thunderstorm centers [94] from which an average yearly location of each can be calculated. With them, the azimuth of each location with respect to our observatory can be calculated, thus giving us the coefficient by which intensity is modified for each channel. Results are depicted in Table 4.4.

The distance to the thunderstorm center is relevant as well since the intensity measured by the sensors depends on the distance from the thunderstorm center [14]. Looking at the figures depicted in the cited work, a difference of approximately 6 dB in the reception of the first mode can be appreciated between the Asian thunderstorm center and the rest, given

Table 4.4: Geographical coordinates of the thunderstorm centers and resulting coefficients to calculate each channel intensity.

	Coordinates	**Azimuth**	**EW coef.**	**EW coef.**
African	63.66º S, 28º E	11.5 Mm, 166.56°	0.232	0.973
American	66.21º S, 66º W	12.7 Mm, 203.28°	0.395	0.919
Asian	55.70º S, 120º E	15.3 Mm, 134.93°	0.708	0.706

their distance to our sensing station. In consequence, the perceived intensity of the Asian thunderstorm center will be reduced in half. The calculated coefficients are then applied to intensity data, and presented in Figure 4.15.

(a) EW channel

(b) NS channel
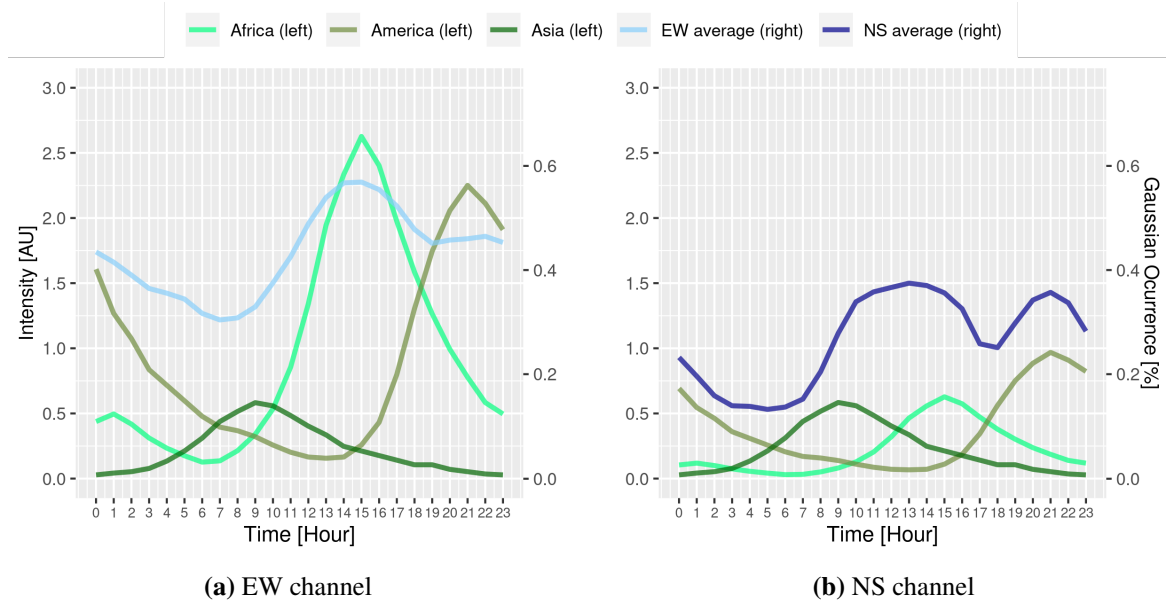
Figure 4.15: Intensity and Gaussian occurrence comparison with localization correction.

After modifying intensity values to match the ones received by each sensor, the relationship between the fraction of segments identified as Gaussian and lightning activity is displayed more clearly. It can be seen in Figure 4.15a how the African thunderstorm center's intensity contributes in full to the signal, explaining the high peak displayed by Gaussian occurrence. American thunderstorm center's contribution is also significant, but its shape is not entirely carried into Gaussian occurrence. This points out how the relationship between thunderstorm intensity and Gaussian occurrence on the SR is not linear, but something

more detailed and worthy of further study. Finally, the Asian thunderstorm center's real contribution to EW intensity is reduced by 66 % due to its geographical position and distance in relation to the measurement station, being almost 4 times lower than the American contribution, thus explaining how its effect is not observable in the Figure.

On the other hand, Figure 4.15b shows a very reduced contribution from Africa, having 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, having one hour difference between the hours where Asian and African thunderstorm centers reach their peaks (9 UTC and 15 UTC, respectively). In this case, the American thunderstorm center's contribution has the maximum value of the three, and its shape is indeed carried into Gaussian occurrence.

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. A rigorous mathematical relationship between both of these phenomena is still to be defined. Nonetheless, this method presents a way to process considerable amounts of data while being able to identify the time segments where lightning activity was measurable on the SR. This allows researchers to automatically classify their records and have a general idea of the phenomena that occurred within them.

4.3.2. Overall Results

Figure 4.16 gives an overview of the analysis performed, showing the results of the data with segments of 1 min duration along with the color-coded bar that indicates data integrity (see Section 3.2).

Figure 4.16a displays the results for the EW channel and Figure 4.16b 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.
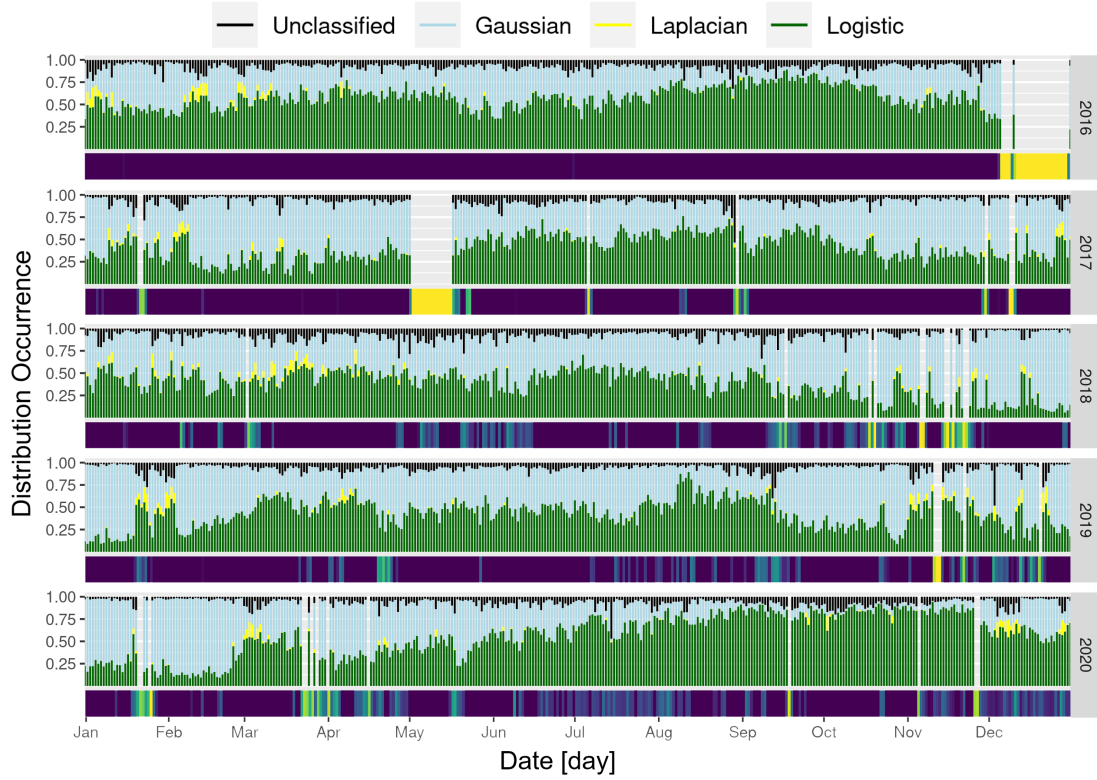
(a) EW channel
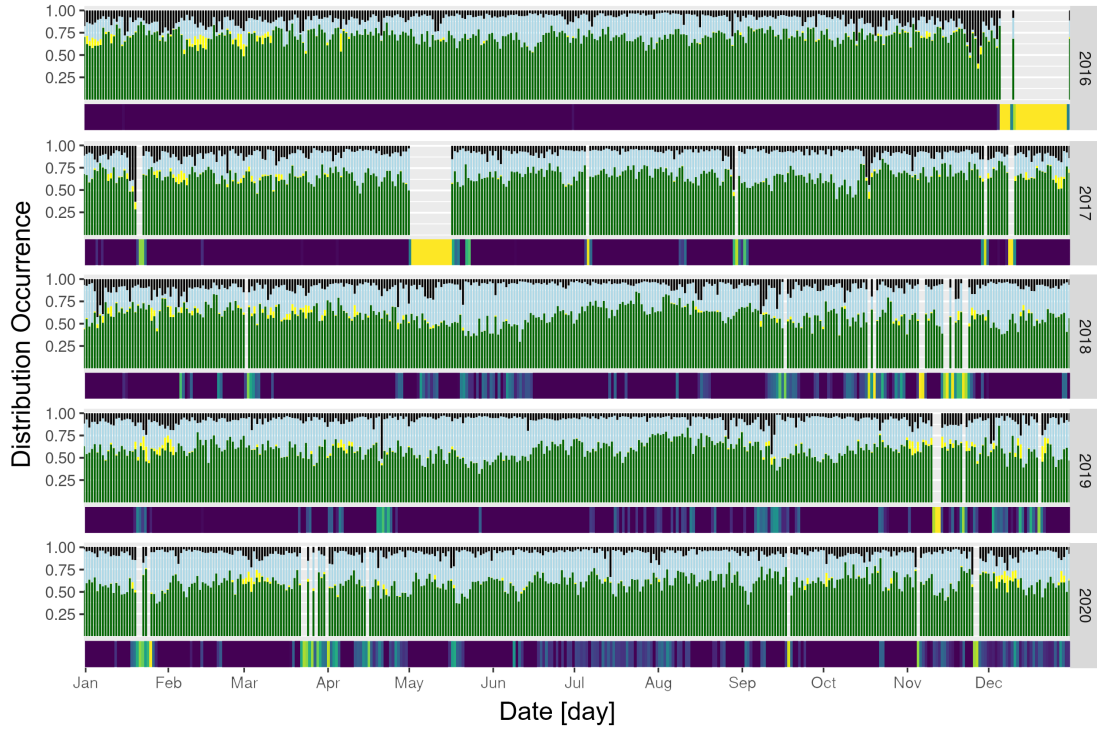
(b) NS channel

Figure 4.16: Daily average of distribution occurrence for all the analyzed data (1 min segments).

Table 4.5: 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 effects of the averaging method explained in Section 3.2 are applied in this Figure. It can be seen how the two big gaps in December 2016 and May 2017 caused by system repairs are not computed due to their length. Despite NA values being scattered along the whole data, there are no other meaningful gaps; save for a few days where most data is missing, the rest of the data at this level of aggregation are continuous.

It can be appreciated throughout all years how Laplacian occurrence is concentrated in the months from November to April. On the days when Laplacian segments appear, the presence of Logistic segments is also higher, for example in the last trimester of 2018. This is especially noticeable in Figure 4.16a, where they reach values similar to the ones displayed in Figure 4.16b.

The preliminary hypothesis regarding Laplacian segments is that they are related to strong lightning discharges closer to the sensor. This hypothesis is driven by how a similar amount of Laplacian segments appear on both channels when there is a concentration of them on the same day. The sensing capacity of the magnetic coils that serve as sensors only overlaps at close range. Inspection of Figure 4.16 confirms that despite presenting different data, Laplacian segments appear on the same days on both channels, and with sizes that imply similar occurrences 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. The results are in Table 4.5.

The high correlation values are interesting evidence in favor of 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 show how all thunderstorm centers affect both sensors in a specific proportion, and how this relationship can be extracted from the statistically analyzed records.

At a glance, one can appreciate how Gaussian occurrence increases around late Autumn/early Winter which in turn implies an increase in Logistic occurrence in the Summer months. Admittedly, this shift is less noticeable in 2017 and even less in 2016. In these years the shift is still present, but Logistic occurrence is more stable than in the rest. It is worth mentioning how in both 2016 and 2020 early Autumn has a high Logistic occurrence that is not present in the other years.

Having a target distribution group of three and ascertaining that Laplacian manifests only occasionally, Logistic and Gaussian can be considered complementary. 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 Figure 4.16a than in Figure 4.16b. This is to be expected if Gaussian occurrence is tied to lightning activity since the overall intensity captured by the EW sensor is higher than what the NS sensor receives as was shown.

The number of data points available allows to explore the data showing the hourly mean of each target distribution. Figure 4.17 displays the analysis results for April of all the analyzed years for both channels (Figure 4.17b for NS and Figure 4.17a for EW). In this case, each segment represents an hour. Data integrity is absent since this level of aggregation 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 of analyzed data. It is most common to lack data for an entire hour, which is why the gaps in the figure suffice to show data integrity.

The asymmetry between both channels is noticeable here as well; in Figure 4.17a the fluctuation between Gaussian and Logistic segments is more noticeable than in Figure 4.17b. Also, Gaussian occurrence is way higher in the former than in the latter. In all years a daily
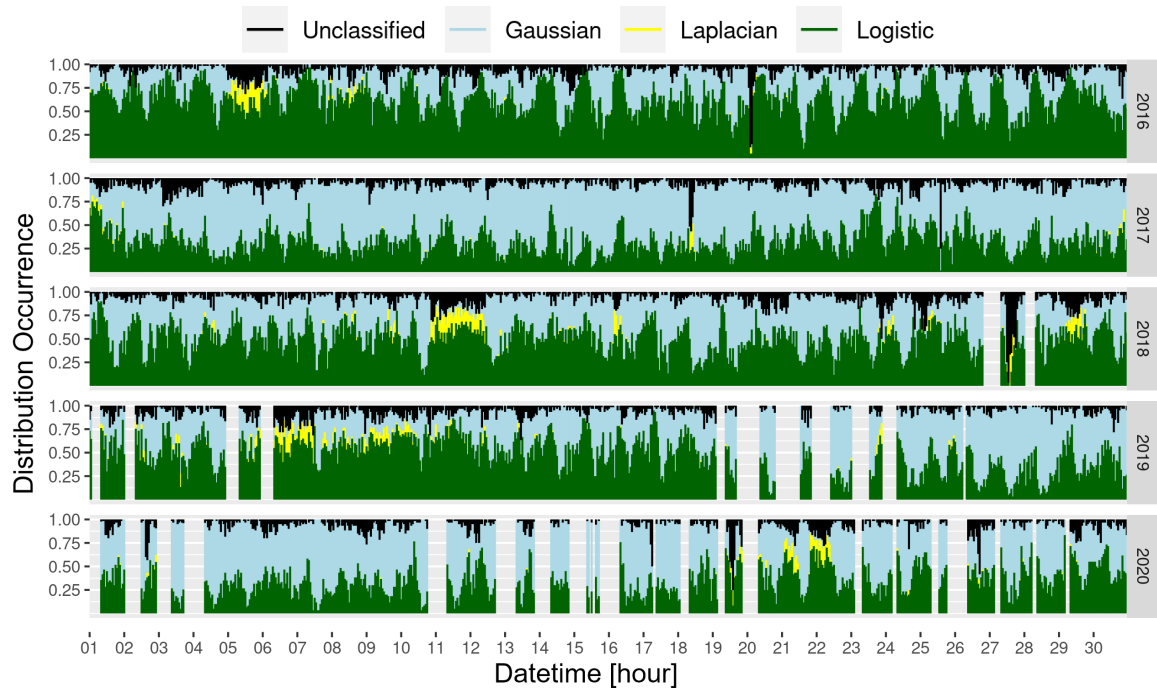
(a) EW channel
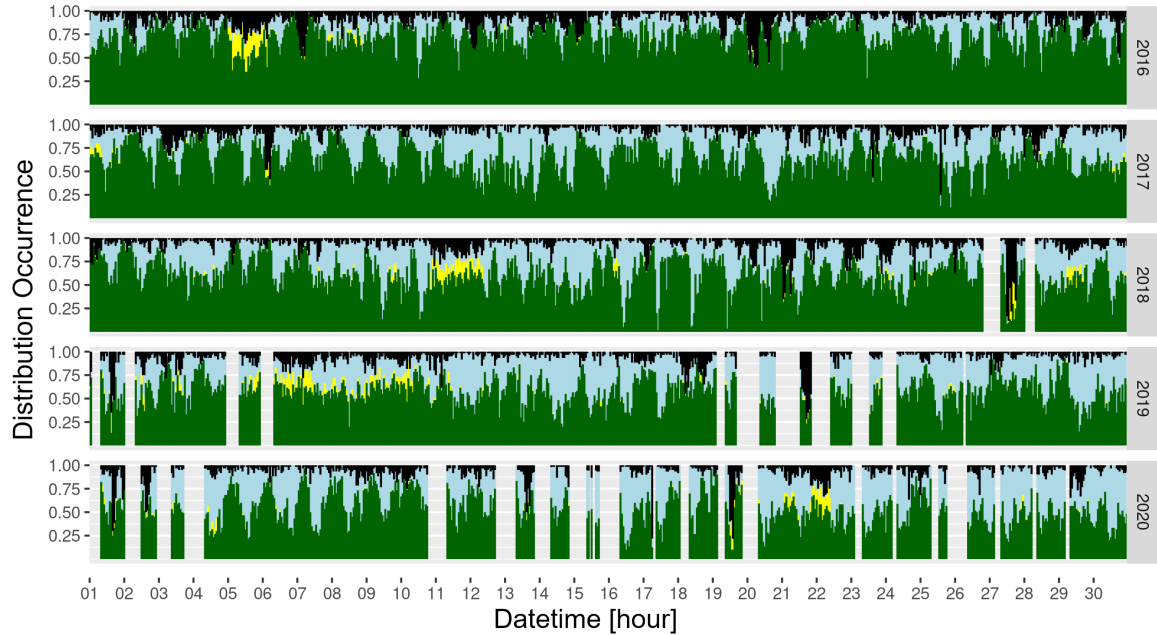
(b) NS channel

Figure 4.17: Hourly average of distribution occurrence in April of each year (1 min segments).

variation can be appreciated. This falls in line with the initial hypothesis since thunderstorm centers have, among others, a daily cycle [91, 171].

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 with small or nil Laplacian occurrence can be observed as well. For example, in Figure 4.17a April 2016 and 2018 Logistic segments clearly overcome Gaussian's whereas in 2017 and 2020 it is the opposite. For April 2019 it appears to be split since towards the end an increase in Gaussian segments can be seen. On the other channel (Figure 4.17b), Logistic dominates most years although the fluctuations between Logistic and Gaussian segments can be seen throughout the day as well. On the hourly averaged records, more extreme cases are observed, during which most of the 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, not crossing the quality of fit threshold, fell in this category. Laplacian segments that do not overcome the quality threshold are mostly segments capturing high pulses among the background noise, known as Q-bursts, or the sensor saturation due to ELF flashes [5], such as the segment displayed in Figure 4.18.
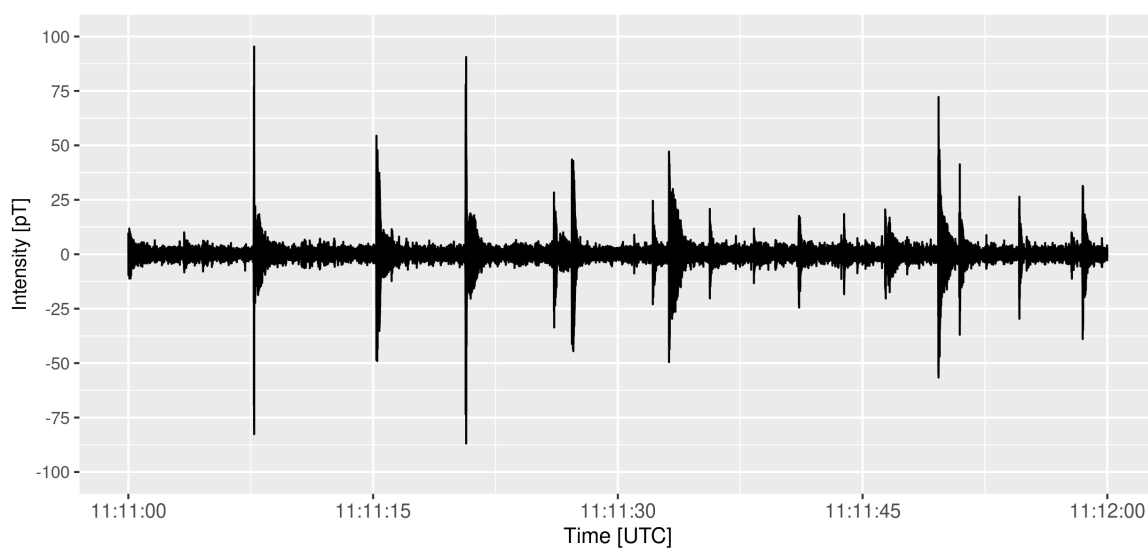

Figure 4.18: Unclassified segment from 2018-04-27, displaying high amplitude pulses.

Since the goal of this analysis is to show how certain distributions may identify specific behaviors on the analyzed signals, those that encompass different phenomena should be left apart for this study.

After checking the general results presented in Figure 4.16 and Figure 4.17, it can be stated with enough confidence that three out of four categories of this analysis can be considered as a state in the SR registers:

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

- Intervals containing several high amplitude peaks, which are the segments classified as Laplacian.

- The segments 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 deviation. Such a filter was the one used to get the segment displayed in Figure 4.18. The same can be said about the rest of the 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.

4.3.3. Gaussian occurrence analysis

To compare how segment length affects the classification results and to delve further into the relationship between the fraction of segments classified as Gaussian and lightning activity, the data of both analyses will be plotted along with the main trends extracted from it, with a figure for each channel.

First, Figure 4.19 presents the daily fraction of Gaussian segments captured by the NS channel. Over them, their monthly and seasonal trends generated from the data are presented. Under the plot, data integrity is displayed. It will be used to support the conclusions, but also so the observer can keep in mind a value of confidence for the data shown, which in turn enhances the researchers' ability to interpret the data.
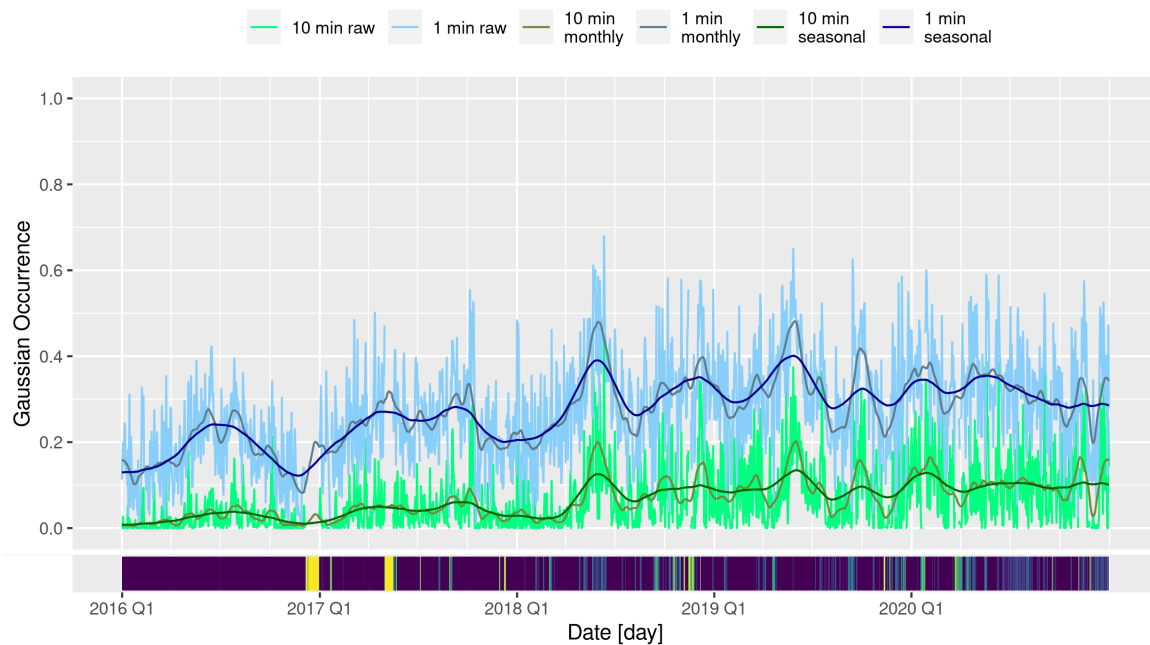
Figure 4.19: NS data with monthly and seasonal trends.

Some similarities and differences between both can be highlighted:

- 1 min segments' average value is higher than in the 10 min ones. It seems 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 segments are quite similar. Nonetheless, the excursions in the 1 min trend are more pronounced than on 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, peaks in the 10 min trend has always a counterpart on the 1 min one. It is not so clear if the relationship works 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.

About the general shape of the trends, it is hard to find common patterns throughout the years at first glance. 2018 and 2019 show a distinct peak at mid-year, but 2019 has another pronounced peak at the start of the third trimester. The latter is present in 2017 and 2020 as well but this year shows no mid-year peak. 2016 and 2020 do not show great excursions on the signal, especially in 2020.

A few NA values are involved in the 2018 mid-year peak calculation but the 2019 data for this period is clear and reaches the same height. It seems clear to assume those values are reconstructed properly. The year that shows more deviant behavior is 2020, with next to no variation in Gaussian occurrence throughout most of the year. This lack of excursions is coincidental with some NA values throughout the whole period. Although the smoothness of the seasonal trend has some similarities to the one displayed in 2016, the former trend monthly trend variations are more pronounced than the latter. This sheds a shadow of doubt over the behavior displayed in 2020.

Next, Figure 4.20 presents the data from the EW channel in the same format used for Figure 4.19. The same observations highlighted for NS data hold in this case, but there are also two additional aspects to be considered:

- 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 raw peak value overlaps with the 1 min peak, sometimes even going over it. That being said, its average value is still lower; 10 min reaching the same value as 1 min happens only for high values of occurrence.

Regarding the trend, a yearly pattern can be seen with a peak in Winter and a trough in Summer, although the exact location of the peak changes by year. Once again both the 10 min and 1 min monthly trends' behavior is similar, although the relative variation of each
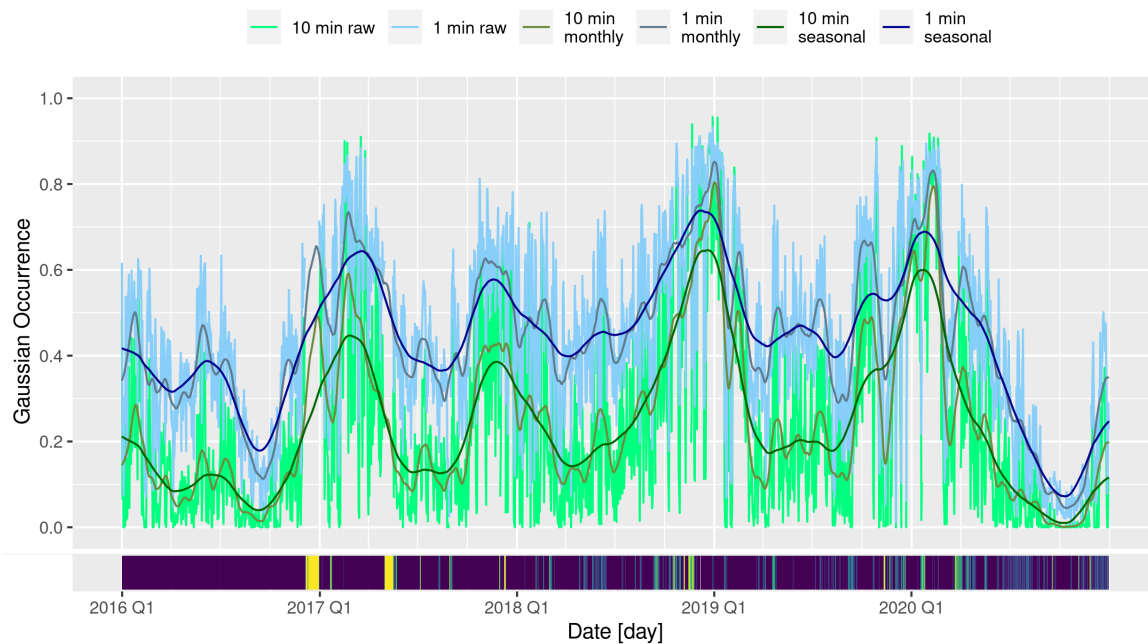
Figure 4.20: EW data with monthly and seasonal trends.

is different. For example, in 2016 both trends show two peaks in Summer. Still, the second peak is lower than the first on 1 min length segments. On the other hand, on 10 min length segments both peaks have almost the same height. Another example is in the first trimester of 2018, on which 10 min first peak is higher than the second, whereas on 1 min it is the other way around.

2020 trends are unique in comparison with the rest of the years. It displays the lowest occurrence value of all records and from the January peak onwards, the downward trend is dominant even in the 1 min records. In the presentation of NS data, the amount of NA values present was highlighted as to how they might be affecting the conclusions extracted. Nonetheless, the lowest value is reached at the start of the fourth trimester and, as the integrity bar displays the data from that period is fully consistent. This gives some confidence in the seasonal trend, despite the amount of NA values present. The same cannot be said regarding monthly trends, since some incremental shifts are perceived in the big gaps, such as the end of 2016 and the start of the second trimester in 2020.

After reviewing the data from both channels, one of the main differences between the two analyzed lengths was their average value and their peak value. The former seems to be

directly related to the analysis features. Generally speaking, trend shifts are more noticeable in the 1 min segments than in the 10 min. The ELF signature of a lightning event can have a duration of a whole second, so for long segments, more events are necessary to affect the statistical distribution's shape. The only situation in which Gaussian occurrence in 10 min can raise over the value displayed in the 1 min segments is when the number of lightning events is enough to affect a considerable amount of 10 min segments. in consequence, 10 min segments have a higher threshold of lightning activity than 1 min ones, as can be appreciated in Figure 4.21. On it, it can be appreciated how in the 1 min analysis Gaussian occurrence rarely reaches 0.0, whereas in the 10 min analysis it is a common value.
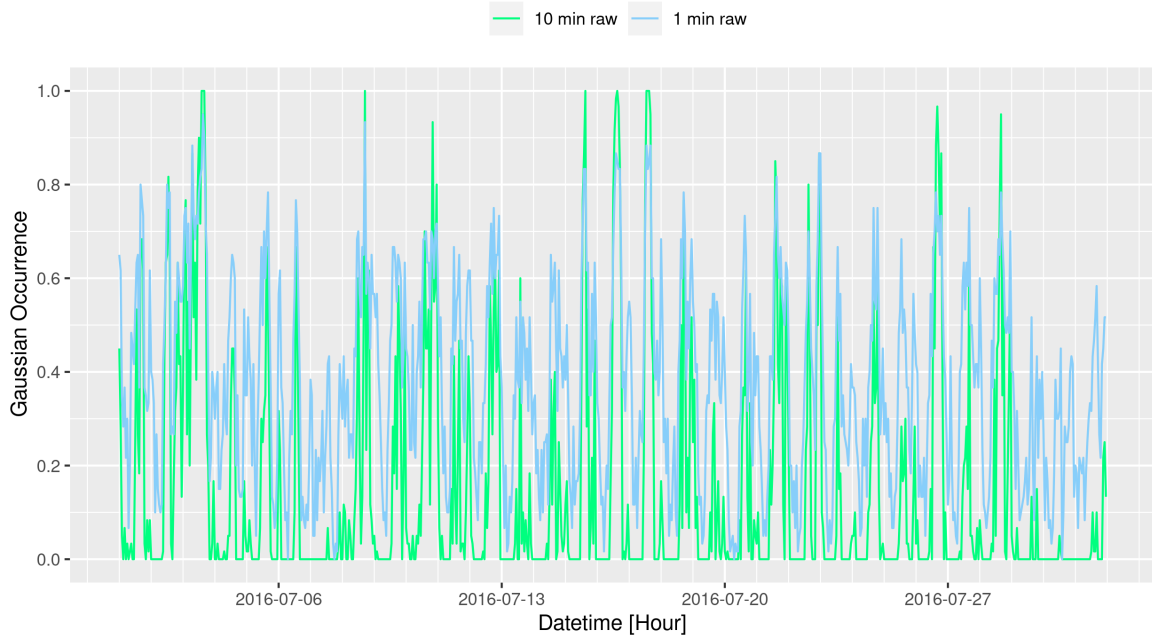
Figure 4.21: Hourly averaged fraction of Gaussian segments for July 2016, EW channel.

On the other hand, during the high occurrence hours, the number of segments deemed as Gaussian is higher for the 10 min than for the 1 min. This consequence stems from the 9 min overlap applied to 10 min segments since a specific event will appear on 10 different segments, whereas in 1 min length segments (no overlap) each event affects only one segment. In conclusion, by controlling segment length and overlap, the researcher can modify the selectivity of the method.

In relation with the significance of 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 is once again coincident with the registered lightning intensity of the storm centers after considering source bearing with respect to the observatory:

- On the NS channel (Figure 4.19) the max value is around 0.7, reached by the 1 min values. 10 min segments stay always at lower values, its maximum being around 0.4.

- EW channel behaves differently on what peak values respect (Figure 4.20). Its maximum occurrence value is close to 0.95, reached by the 10 min segments.

Interpreting the results of each length as discussed above, it can be suggested that Gaussian occurrence is indicative of an effect that is registered on both channels differently.

4.3.4. Pattern analysis

To further analyze the daily trends displayed by Gaussian occurrence, an analysis using ACF (see Section 3.3.5) has been applied to four months of data from the EW channel, for all the analyzed years. The results are displayed in Figure 4.22.

NS channel's ACF (Figure 4.22a) shows an unclear pattern, reinforced by the difference between the 10 min and the 1 min values, whereas from the EW channel (Figure 4.22b) a clear seasonal pattern emerges.

The EW channel, heavily influenced by a single thunderstorm center should display its seasonal pattern, whereas the NS channel would be hard to analyze by this method, given the interaction of two storm centers. This implies that the NS channel would require a more elaborate analysis, so the attention for this section's analyses will be shifted entirely to the

(a) NS channel   **(b)** EW channel
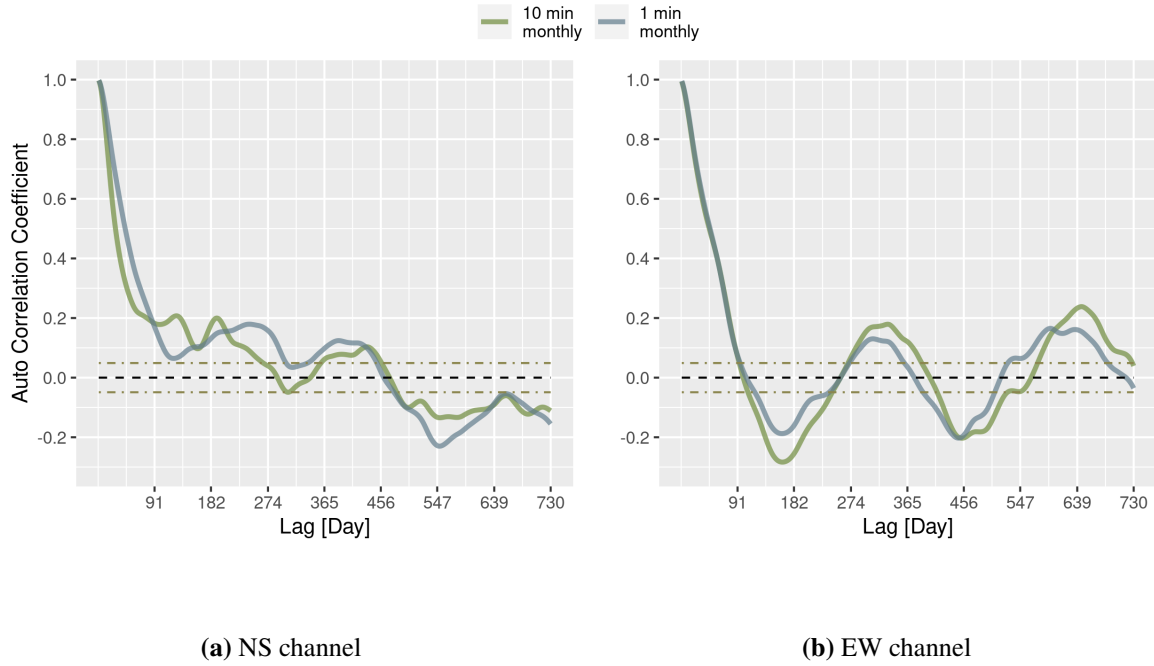
Figure 4.22: ACFs for the monthly trends of the data analyzed with different segment lengths.

EW channel, on which it will be easier to find the sought patterns. It can be seen how the amplification of the extremes caused by the 10 min segments enhance the ACF coefficient, so these registers will be the ones used to further explore how well the Gaussian occurrence trends fit those of the African thunderstorm center or lightning activity in general.

To deeply analyze the continuity of these trends the data has been grouped into three years, and the ACF has been applied to each group (Figure 4.23) for both the seasonal trend (Figure 4.23a) and the monthly trend (Figure 4.23b). The group size of three years allows for a maximum lag of a year and a half on which still 50% of the data is being actively compared.

Figure 4.23a leaves no doubt about a seasonal yearly trend on Gaussian occurrence, once again similar to the African thunderstorm center. Confirmed the trend it is worth observing how in Figure 4.20 the peaks and troughs of the seasonal trend match those of the African thunderstorm center as shown in the literature [172].

Removing the seasonal trend from the monthly trend allows for clear visualization of the monthly patterns. Further patterns have been identified at different time resolutions [23, 26],

(a) Seasonal trend

(b) Monthly trend without the seasonal trend
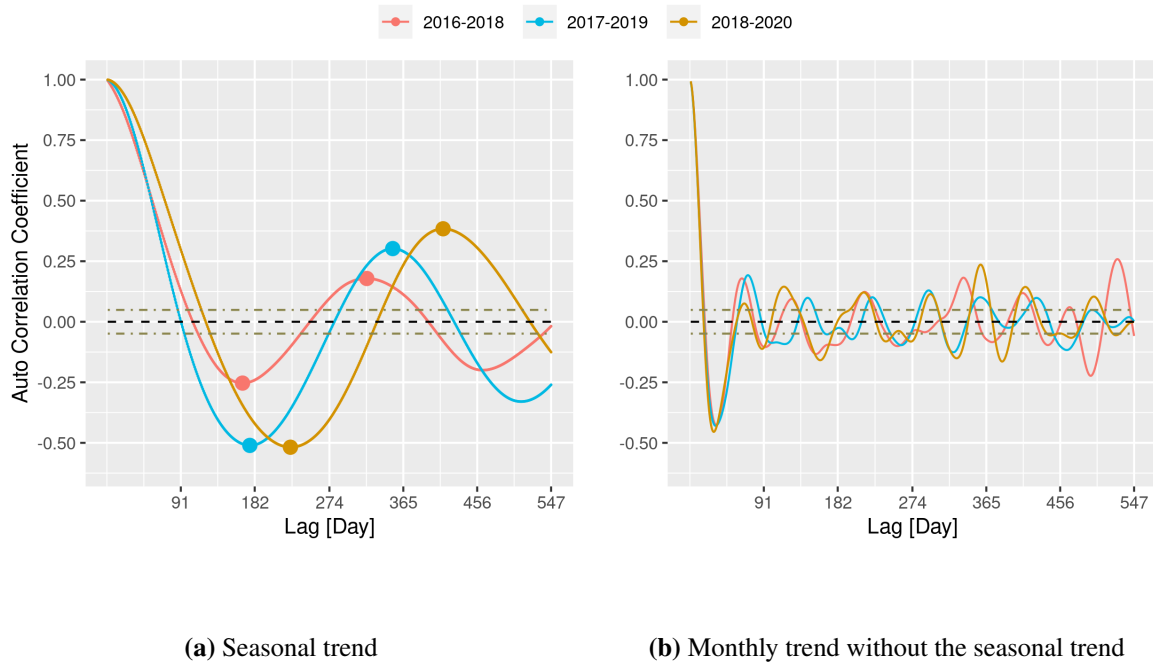
Figure 4.23: Different ACFs with a maximum lag of 18 months, with three groups of three years.

and what can be seen here is most probably an interaction of these different effects modifying the general shape.

To understand how this trend changes daily from the hourly averaged records, Figure 4.24 shows the ACFs of four months, representative of each season, for all of the analyzed years.

The expected daily trend can be observed almost uniformly in every case, although some deviations exist and the possible reasons are worth discussing. First, the trends shown in Figure 4.24a are warped by an underlying trend, and being nonexistent for this month in 2019 and hardly recognizable in 2016. Fig 4.24b is still affected by a different trend but the daily peaks are now clearly observed. The daily trend is observable for all years in Figure 4.24c with only 2019 deviating from the general pattern, while in Figure 4.24d 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

Figure 4.24: ACFs results for the hourly averaged data of a chosen month on all the analyzed years, with 7 days of maximum lag.

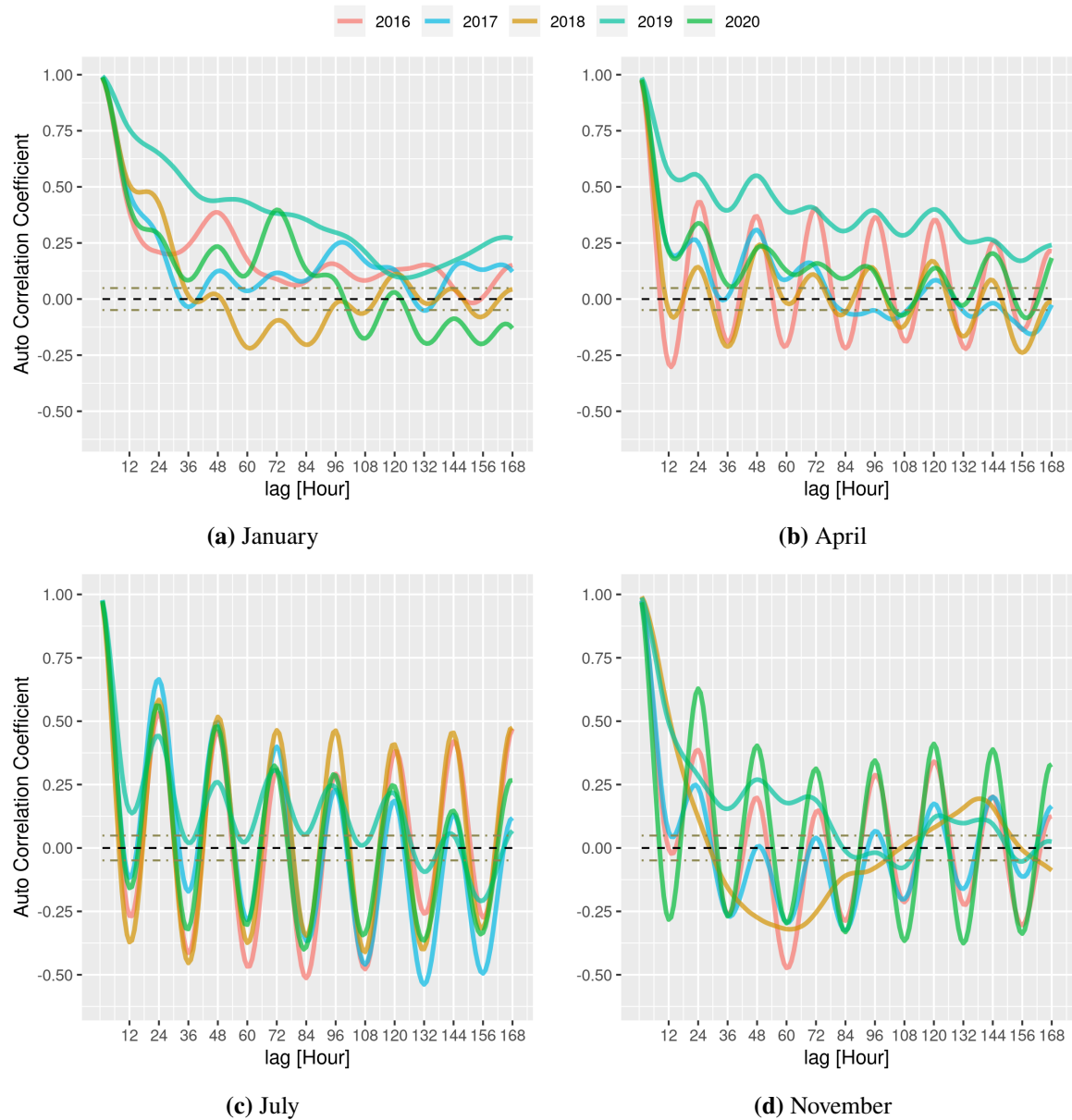how the season to which the data belongs should be accounted for. Figure 4.24c on which the trends are more visible shows data from July. This is the month when the African thunderstorm activity is at its lowest on average [172]. This implies that minimal activity translates as well into minimal variability, so even when lightning activity is at its lower (see Figure 4.20) 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 (Figure 4.24b) and November (Figure 4.24d), These months are theoretically on the falling and rising edges of activity, respectively. By following the previous train of thought, how well the daily trend can be observed in the month of a specific year gives information about how closer to the trough the month is.

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

4.4. Statistical differences quantification

In this section, the technique for quantifying SR intensity variations over time under different situations using statistical methods proposed in Section 3.4 is tested. It will be applied to the BPF data of 2018 and 2019. Only the EW channel is analyzed, since it is strongly influenced by the African thunderstorm center due to its orientation, while the effects of the other two main hotspots of lightning activity (American and Maritime continent) are mitigated [57], thus simplifying pattern detection. The results are evaluated to uncover any additional information that can be extracted from the procedure.

Under the idea that supports this thesis - a SR temporal segment's statistical distribution being representative of the events captured by the signal in that period of time - in this section the hypothesis of segments with low KS distance value having similar characteristics will be tested.

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

4.4.1. Methodology verification

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

An example of each can be seen in the following figures. On one hand, Figure 4.25 shows a stable sample with no significant trend changes and overall low variability along the whole interval whereas Figure 4.26 displays significant variations over the base value.

Figure 4.25: Signal variability on June the 24th, 2018 from 15:00 UTC to 16:00 UTC - segments 5 minute in length, 2 second difference between segments.

The former serves as a baseline to quantify the standard level of randomness that the signal can exhibit, whereas the latter displays several differences when compared with the target segment. The target segment is compared with itself as part of the analysis process and thus is identifiable for being the only data point showing KS distance of 0.0.
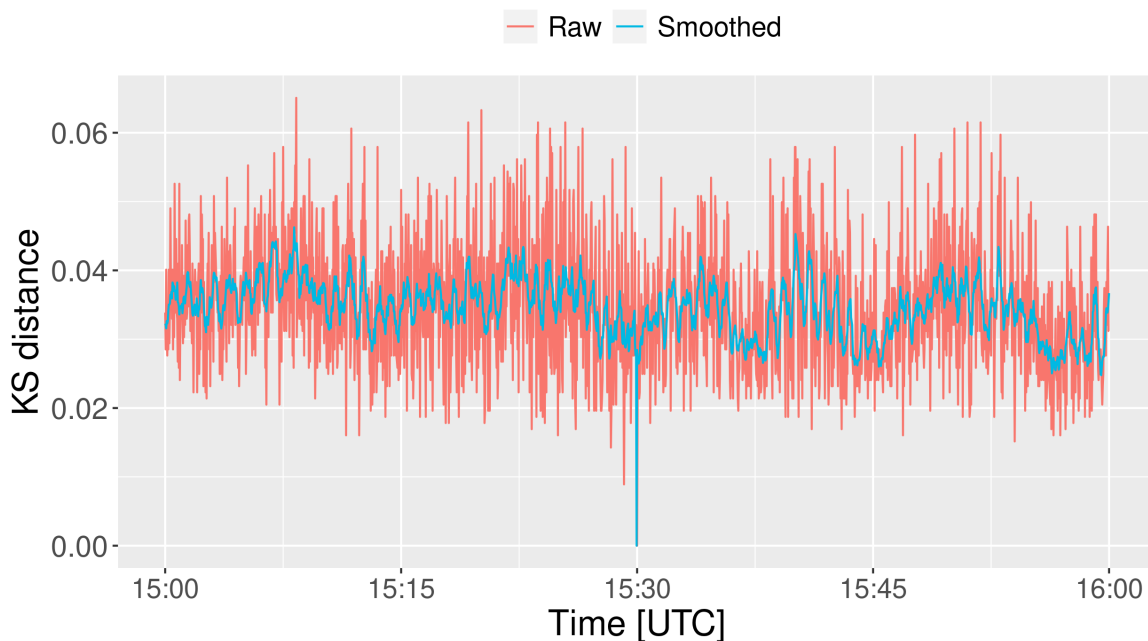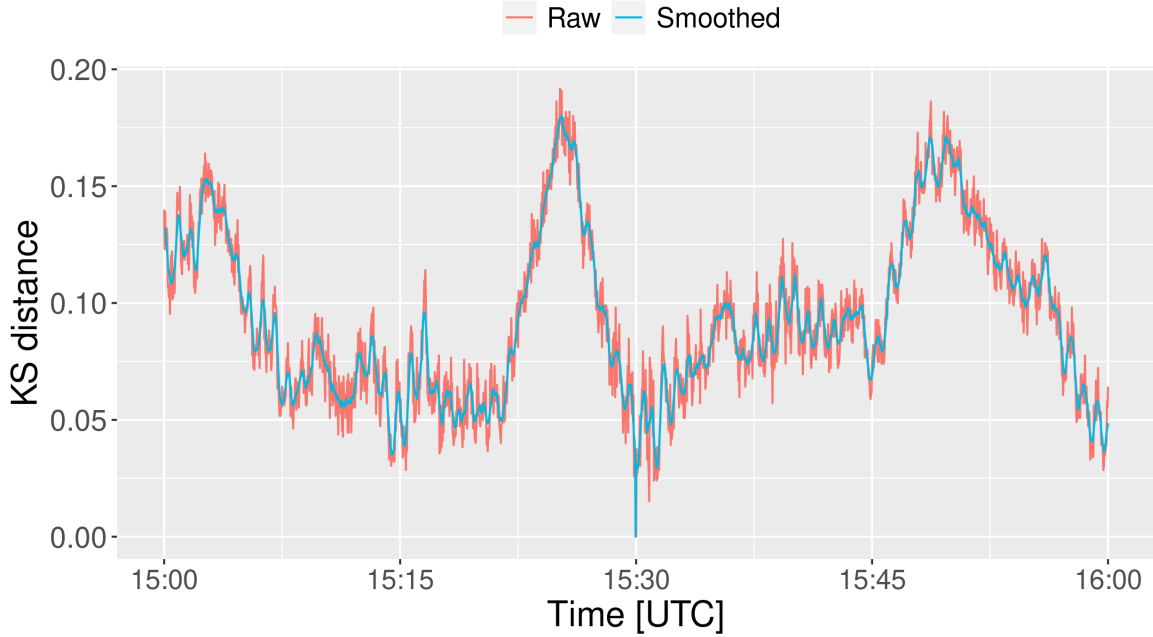
Figure 4.26: Signal variability on April the 6th, 2019 from 15:00 UTC to 16:00 UTC - segments 5 minute in length, 2 second difference between segments.

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

Besides the base randomness value and the existence of trends in the statistical variation between temporal segments of a given time slot, other fact strikes as surprising. The KS distance was expected to gradually increase from the 0.0 that indicates the target segment. This is so because the data point belonging to the comparison against a segment two minutes forward shares 3/5 of its information with the target segment. Despite this, it can be seen in Figure 4.26 how segments further away (for example, the minimal values around 15:15) are more similar to the target segment than the ones that share information with the target segments, like the data points around the 15:25 peak.

As explained in the previous section, the relevant source of variations at this temporal resolution is lightning activity. Therefore, the variations shown in both Figure 4.25 and Figure 4.26 can be related to differences in thunderstorm activity over time. On one hand, the minimal variations displayed by Figure 4.25 can be considered as a stable period of lightning activity, since all the analyzed segments are similar. On the other hand, the variations presented in Figure 4.26 mean that during the analyzed hour, lightning activity was inconsistent.

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

Moving onwards, the major point of interest in testing the presented methodology is to offer some insight into the general trend changes such as the ones displayed in Figure 4.26. Consequently, all the results will be smoothed as displayed in the figure. It is worth mentioning how the smoothed line still reaches a 0.0 value for the target segment; this is intended to make it identifiable at a glance.

To further characterize the resulting curve's shape and how the methodology's degrees of freedom affect the results, the parameters of the analysis must be accounted for. It is trivial that the time between segments affects the analysis resolution. Common sense drives the hypothesis of segment duration mitigating the average variation value for a given period, but a simple test reveals that this is not the case. Figure 4.27 displays an analysis result where the length of the segments has been modified while keeping the step size (namely, the amount of time between a segment and the next) constant.
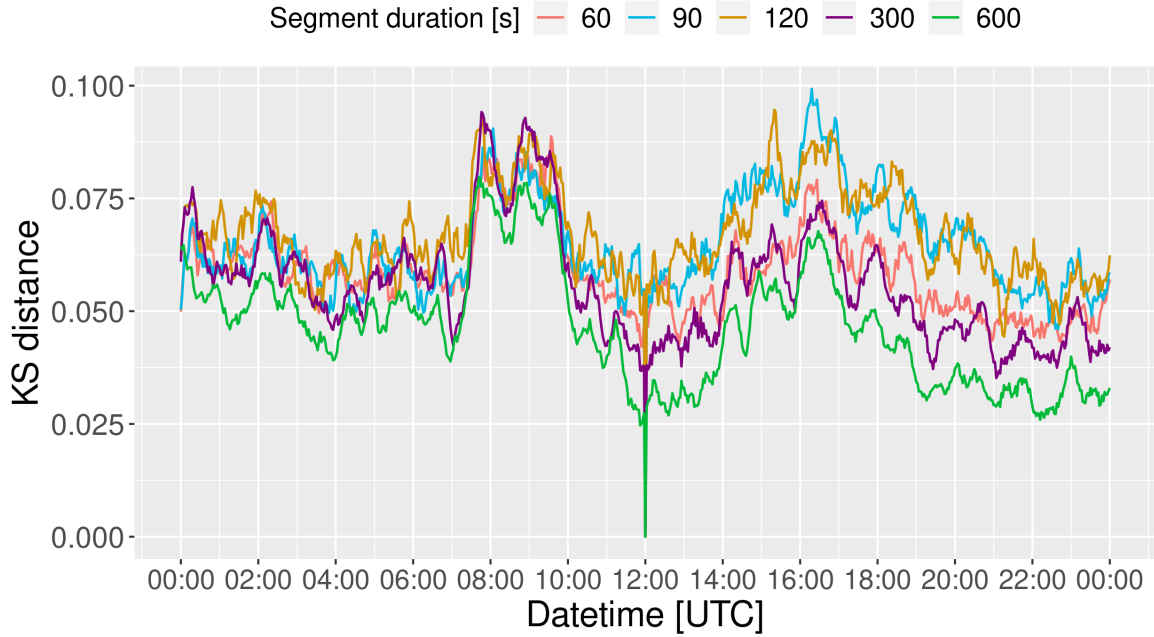
Figure 4.27: Signal variability on December the 27th, 2018 - variable duration, 2 minute difference between segments.

While it is true that modifying segment duration affects the average variation, their relationship is not direct. Figure 4.27 shows how the segment duration that peaks in average variation are 90 and 120 seconds. It is also worth noticing how, on top of the lower average values displayed by the 600 second segments' analysis, it shows a heightened sensitivity to difference detection. Between 08:00 and 10:00 AM, average KS distance is doubled for the 600 second segments, reaching the values shared by the rest of the curves, while the comparative gain on the rest is not that high. On that same interval, it is the 300 second segments' curve that peaks in variance.

To provide further insight, Figure 4.28 displays a similar procedure applied on a day with more pronounced variations.

In this case, the analysis with a segment duration of 120 seconds displays the highest average variation, especially noticeable in the second half of the day where the signal is more stable. On the other hand, during the night the pronounced variations on all curves follow the same pattern, with the range of values being more narrow than during the signal stability.

111
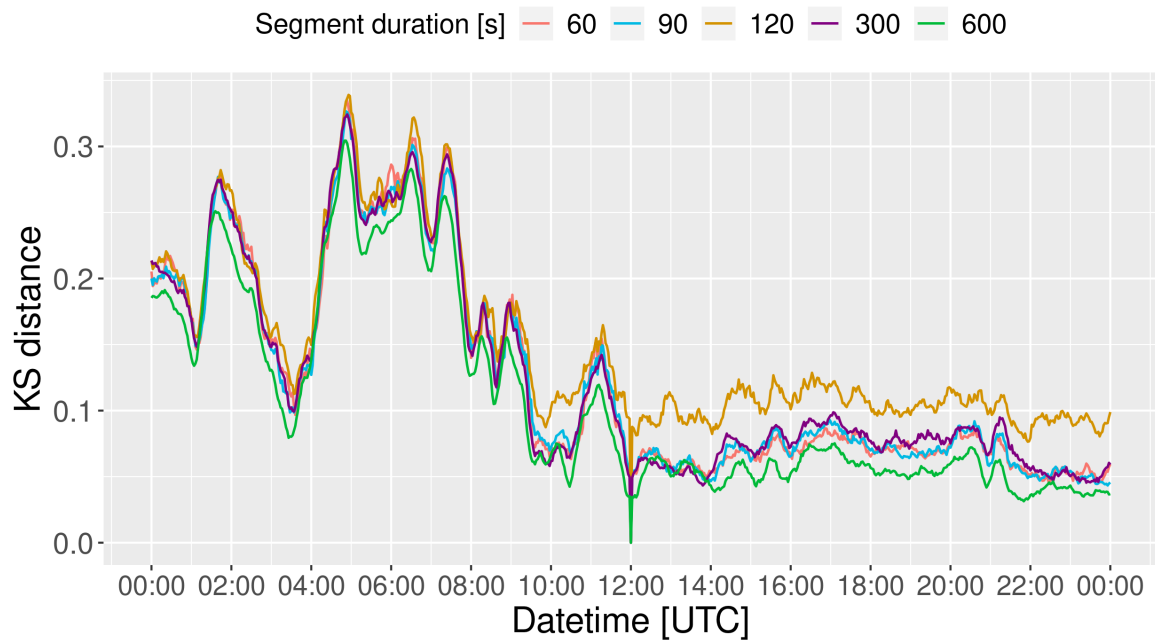
Figure 4.28: Signal variability on March the 13th, 2019 - variable duration, 2 minute difference between segments.

In this initial stage of methodology testing, the reason why the average variability does not linearly decrease with segment duration can only be the subject of speculation. It may be due to the value of the rest of the parameters or related to some disturbance or behavior present on the records. The former requires extensive testing since the many tests realized during the development of the methodology yielded no specific evidence. The latter is suggested by how in Figure 4.27 the curves of different segments display the maximum variability at different periods. It will require a detailed study of a period well characterized with information about the additional variables that define the SR (lightning activity, ionosphere state, etc.).

To consider the weight of the choice of the target segment, an analysis has been performed multiple times targeting different segments inside the same interval. An example of this is shown in Figure 4.29 which displays pronounced variations during the small hours of the day while staying stable for the rest of the day.

Once again, the temporal location of each target segment is given by the moment where the curve value in the y-axis is 0.0, as well as highlighted in the Figure's legend. An
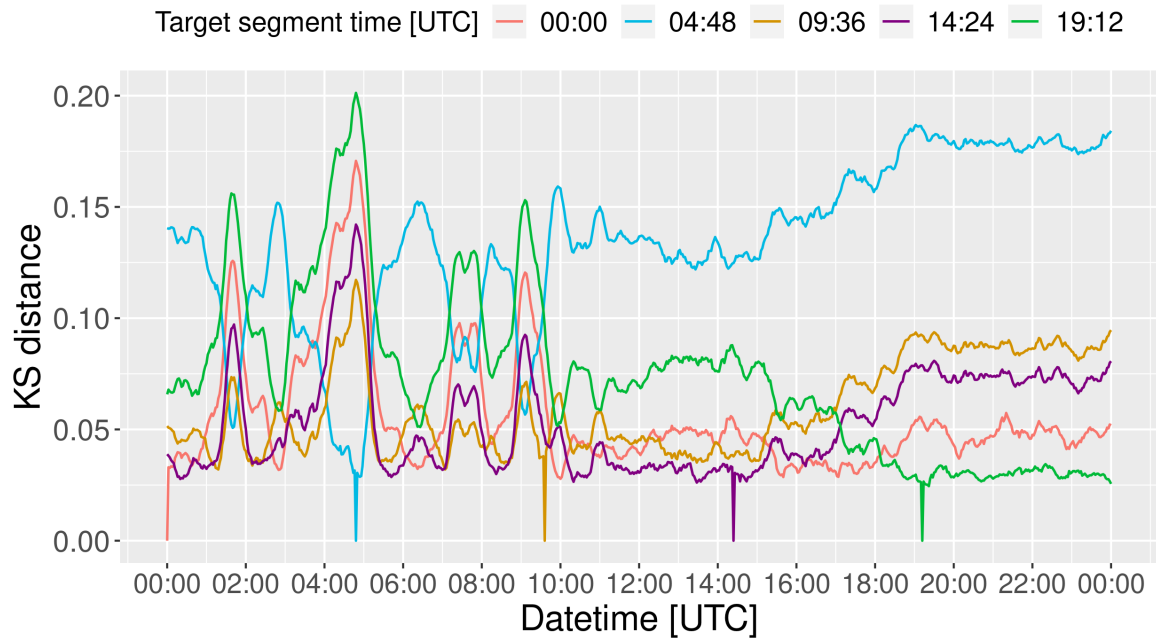
Figure 4.29: Signal variability on February the 4th, 2018 - 10 minute length, 2 minute difference between segments

interesting detail is how the analysis for the target segment of 04:48 UTC (Blue curve) is matched with a maximum variation point for the analysis of 19:12 UTC (Green curve). This causes these two curves to almost mirror each other. In the same fashion, the rest of the segments follow the green curve's tendency during the first half, although the maximum value they reach during each peak has different magnitudes. It is singular how for the end of the segment their maximum values are inverted, with the 09:36 UTC analysis (orange curve) reaching the maximum value with the exception of 04:48 UTC, whereas it had the lowest value on the first half, and with the green curve being the opposite.

In Figure 4.30 the results of an analysis under the same parameters are displayed, but in this case, it shows variation distributed in three clear zones.

The maximum/minimum value inversion behavior is similar to the one displayed in Figure 4.29 but, in this case, no curve clearly mirrors the other. All of them tend to follow the same tendency (especially in the section between 12:00 UTC and 21:00 UTC) with only the curve with the lowest average value per section following a different tendency. The sudden shift that can be appreciated around 11:00 UTC is of interest, since it marks a boundary between
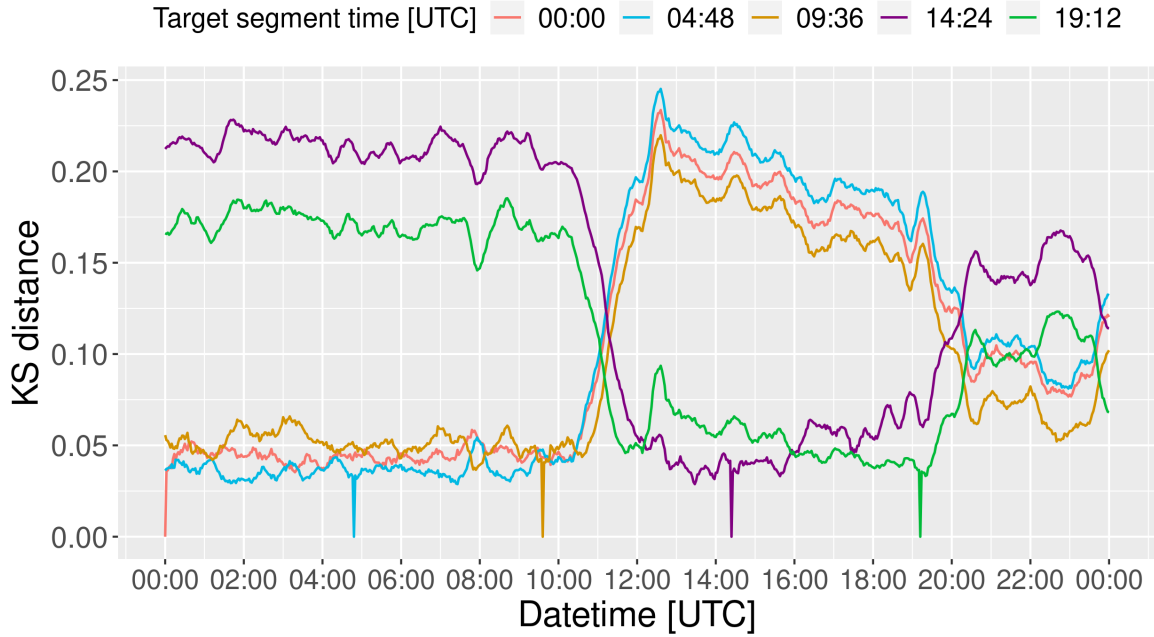
Figure 4.30: Signal variability on June the 2nd, 2018 - 10 minute length, 2 minute difference between segments

the nighttime and daytime realizations. The most likely explanation should be the ionosphere change at sunrise, but the time when the inversion happens takes away some of the weight of the hypothesis. Even more so, considering the analyzed data comes from June, when sunrise happens around 7:00 local time.

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

A variation overview for all the analyzed data is shown in Figure 4.31. All chosen data was subjected to hourly analysis, using segments of 10 minutes in length and with a time difference of 2 minutes between each other. All the results are hourly averaged by month and the resulting mean and statistical deviation by hours is displayed here, for both years.

First, it clearly states the main advantage of averaging a signal such as the SR in monthly segments; the average KS distance stays almost ever below 0.06, being most usually around 0.04. It points out how averaging reduces significantly the variation, allowing many re-
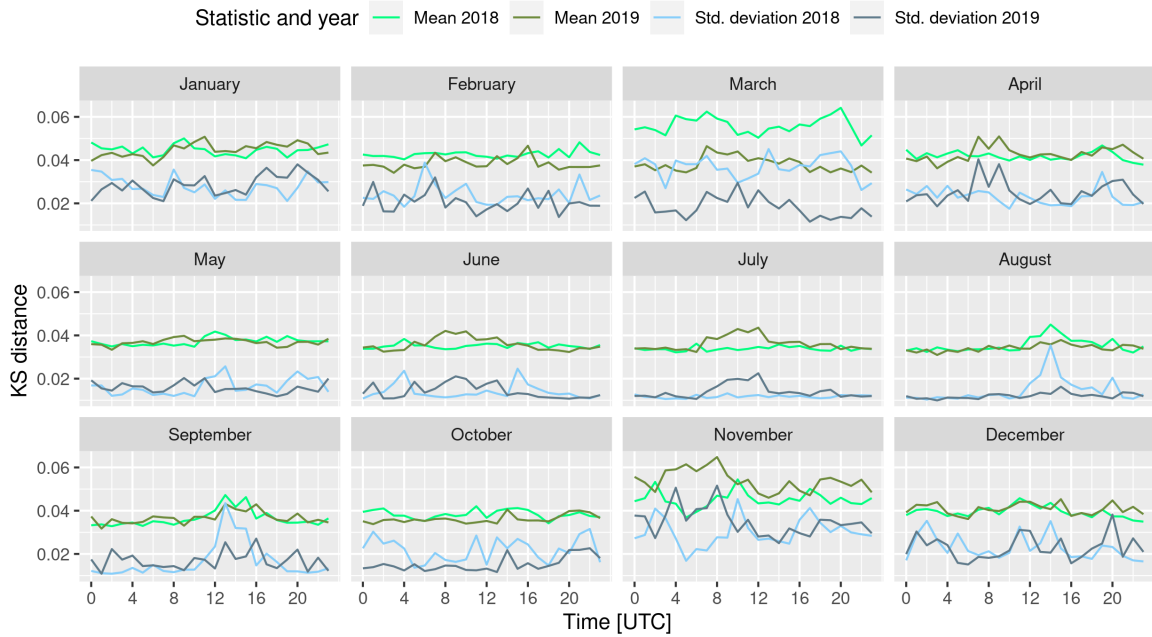
Figure 4.31: General results, showing the hourly mean and statistical deviation of each month, for the two analyzed years

quirements to be fulfilled. That being said, it is interesting how the statistical deviation is more sensitive than the average, the latter following the trend of the former almost to a fault but acquiring higher relative values as well. Lastly, it is worth noticing how the variations, although still reduced, are more pronounced in the first and last quadrimester of the year, whereas in the middle quadrimester the signal is more stable. This is consistent with the properties of the African thunderstorm cycle, which reaches a minimum activity level during late spring and summer. On the other hand, the peak of daily activity for this thunderstorm center happens at 15:00 UTC [94], but the variation analysis does not show a clear maximum at this time, or even around it.

Lastly, a simple test is performed to present in a practical way the effects of this measurement, its results displayed in Figure 4.32.

In this figure, the results of a variation analysis can be seen, with the variation's excursions showing how the time interval is one where the signal changed noticeably in small amounts of time. Below, three sets of frequency spectra calculated with the FFT algorithm using the same signal segments involved in the variation analysis are shown. These have 7 seconds

Figure 4.32: Practical analysis to present the effects of variation, showing the signal variability for April the 20th, 2018 with segments length of 7 seconds, 7 second difference between segments, and the FFTs calculated from different subsets of the sample.

length and 7 seconds between each other, closely matching the recommendations from [4]. Nonetheless, the FFTs has been calculated using different sets of segments. For instance, the spectrum colored in orange has been calculated using the segments with low variation (segments whose variation is under the ora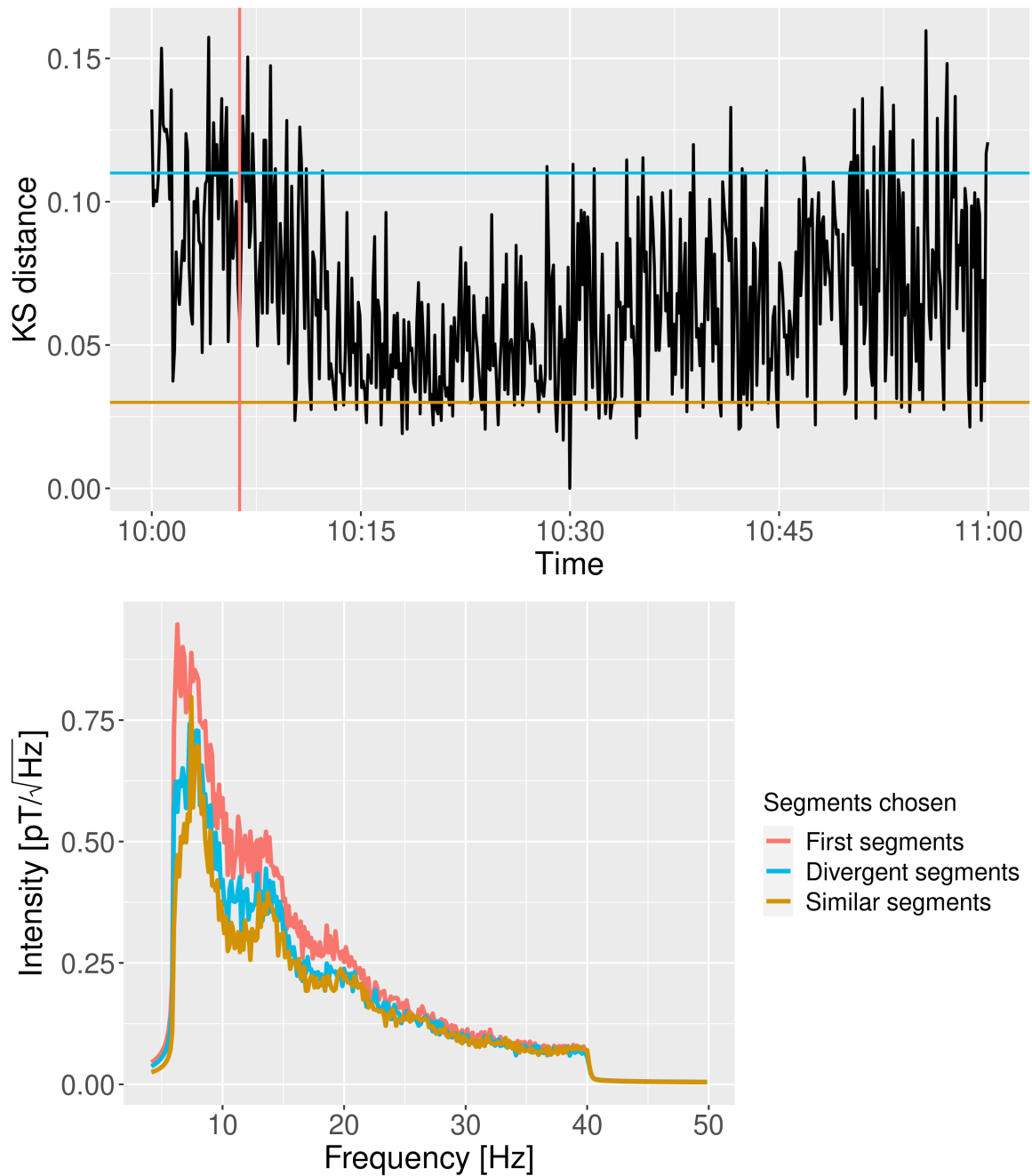nge line, a total of 52 segments) whereas the blue one comes from those with high variation (the 58 segments whose variation places them over the blue line). Finally, the pale red spectrum was calculated using the first 58 segments of the hour.

It is mentioned in [4] how the first four modes are clearly visible after averaging 7 minutes' worth of segments, and recognizable after 2 minutes. It is understood that this rule of thumb is applicable to clear segments with no interference. In this test, the FFTs were calculated using almost 7 minutes of data, (albeit a bit less for the segments with lower values of variation) but, due to the noisy nature of the segment, the SR modes are hard to notice in the FFT calculated just by averaging the spectra of the segments sequentially. On the contrary, the other two spectra show the first three modes, showing how by choosing segments with a similar variation the number of segments to produce a clear spectrum is reduced. That being said, the spectrum produced by using similar segments (that is, those with lower values of variation) is clearer than the one calculated using the divergent segment, despite having fewer samples. An explanation can be reasoned under the constraints of this analysis. There is no doubt that segments with a low KS distance will be similar to the target segment, and by extension similar to each other. On the other hand, segments with a high variation are different from the target segment, but they are not necessarily similar between them.

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

remarked how the utility of variation calculation may be boosted if the characteristics of the target segment are properly identified.

Chapter 5

Conclusions

Along this thesis, the development, test, and results interpretation of two methodologies to study the temporal aspect of the *Schumann Resonance* has been presented. An intensive temporal analysis of the electromagnetic phenomenon is a study that, to the best of the author's knowledge, has not been performed to this date. The developed methodologies can support the traditional frequency based analysis techniques by contributing additional and valuable information.

The first methodology performs an automatic classification of the *Schumann Resonance* registers by segmenting them and finding the best fit among three target distributions. By studying the classification results and comparing it with the known patterns underlying the Earth-ionosphere electromagnetic activity, it has been concluded that a relationship exists between the best fit distribution for a segment and the state of the *Schumann Resonance*.

- Segments classified under the Logistic distribution are related with the signal background noise. These are segments where lightning activity is under expected parameters, without any unusual activity or big discharges.

- A significant correlation has been found between the number of Gaussian classified segments per hour and lightning activity intensity, which gives way to the conclusion of the former being an indicative of the latter.

- Laplacian classified segments are related with high excursions of the signal, mostly produced by high powered lightning discharges. It has been shown how the appearance of Laplacian segments on both channels of the sensors was highly correlated, pointing out how these segments manifest when lightning activity is close to the station.

The effect of the methodology's configuration parameters have been tested as well, revealing the analysis' sensitivity to segment duration; choosing long segments makes them most representative of the signal's background noise (most Logistic segments), whereas short segments are more selective, producing a more diverse classification and revealing the different states of the signal in a more precise way.

The second methodology relies on *Kolmogorov-Smirnov* distance to quantify the short term variations between *Schumann Resonance* segments. The study of the signal with a segment size in which lightning activity is the main factor to consider gives way to consider it a stochastic phenomenon. To do so, a segment is chosen as main sample and the rest segments of a time period are compared with it. The result of an analysis performed in this fashion is the differences between all the analyzed segments and the main sample.

The quantification results show distinct effects that could be attributed to lightning activity and even to day-light differences in the ionosphere. It is also exposed how segments with similar *Kolmogorov-Smirnov* distance values display similarities between each other, which enhances the methodology's usefulness as an automatic tool to select segments with particular characteristics or detect sudden variations that could help in pinpointing specific transient events.

5.1.   Scientific production

This thesis manuscript collects the most significant results of the methodologies developed. From these, the initial exploration of the methodology for statistical classification of *Schumann Resonance* has been published under the title "*A new way of analyzing the schu-*

mann resonances: A statistical approach". This article serves as indicator for the scientific quality of this manuscript.

- Manuel Soler-Ortiz, Manuel Fernández Ros, Nuria Novas Castellano, and Jose A Gázquez Parra. A new way of analyzing the schumann resonances: A statistical approach. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021. pages 64

Besides, two other manuscripts have been written and are in revision stage at the date of presenting this manuscript.

- Manuel Soler-Ortiz, Manuel Fernández Ros, Nuria Novas Castellano, and Jose A Gázquez Parra. Study of the statistical footprint of lightning activity on the Schumann Resonance. *Computers & Geosciences*, UNDER REVISION
- Manuel Soler-Ortiz, Manuel Fernández Ros, Nuria Novas Castellano, and Jose A Gázquez Parra. Quantifying Schumann resonances' variation over time through statistical differences. *Journal of Atmospherics and Solar-Terrestrial Physics*, UNDER REVISION

The first gathers the most relevant aspects of the additional inquiries performed to further interpret the relationship between the results of the statistical classification methodology and lightning activity, The second presents the variability quantification technique based on *Kolmogorov-Smirnov* distance, along with its results and validation.

5.2. Future works

Given the novelty of the methodologies developed in the framework of this thesis, there are still many aspects to be explored.

The statistical classification methodology had demonstrated to be a powerful automated tool to analyze *Schumann Resonance* registers. This is especially so considering the evidence in favor of the relationship between the chosen distribution for any given segment being related with specific states of the signal, as the contents of this thesis has proven. Nonetheless, only the core of the methodology - its qualitative aspect in the way of segment tagging by their best fit distribution - has been tested, and, despite the good results, a lot of potential lies in the quantitative features.

The future research regarding this methodology will proceed as follows:

1. The first step is to complete the methodology's description of events. A considerable amount of segments that were not properly described by their best fit should be further studied to introduce another level in the classification. Just as well, the frequency domain information of each analyzed segment could be processed and stored as part of the classification process

2. In this point of development, the methodology's ability to analyze great quantities of data can produce valuable results if used to locate specific events and attend to their classification. If any event could be tied to its statistical footprint reliably, ways to automate their capture in the records could be devised and included in the methodology, which in turn would speed up other research.

3. The stored data about the analyzed segments - the mean, statistical deviation, skewness, and kurtosis - has been reviewed on this thesis, but their relevance has only been outlined as the central moments' presentation. It follows that comparing these with data records of the main variables affecting the signal, such as lightning activity and ionospheric conductivity, is a necessary step. In case of uncovering quantitative relationships, the methodology's value would be asserted.

4. Finally, the methodology's versatility makes possible to apply it on an entirely different set of time series. Adjusting the target distributions to fit the shapes on another temporal records histograms could be of interest, extending the results produced on the *Schumann Resonance* to other fields.

Regarding the technique to quantify the signal's variability in time, comparative studies must be taken to extract further meaning from this variation. It should be possible to combine it with the statistical classification methodology, in order to provide an in-detail level of description if deemed appropriate - for example, to keep track of unusual events when working with long segments, which would be otherwise masked in the histogram.

Bibliography

[1] Marcelo Bender Perotoni. Eigenmode prediction of the schumann resonances. *IEEE Antennas and Wireless Propagation Letters*, 17(6):942–945, 2018. pages xi, 1, 2

[2] B. P. Besser. Synopsis of the historical development of schumann resonances. *Radio Science*, 42(2), 2007. doi: https://doi.org/10.1029/2006RS003495. pages xi, 2, 3, 5, 6, 43

[3] M Balser and CA Wagner. Observations of earth–ionosphere cavity resonances. *Nature*, 188(4751), 1960. pages xi, 1, 6, 7, 10

[4] Alexander Nickolaenko and Masashi Hayakawa. *Schumann resonance for tyros: Essentials of global electromagnetic resonance in the earth-ionosphere cavity*. 2014. ISBN 9784431543589. doi: 10.1007/978-4-431-54358-9. pages xi, 10, 11, 32, 44, 83, 84, 87, 89, 117

[5] Toshio Ogawa, Yoshikazu Tanaka, Teruo Miura, and Michihiro Yasuhara. Observations of natural elf and vlf electromagnetic noises by using ball antennas. *Journal of geomagnetism and geoelectricity*, 18(4), 1966. pages xi, 4, 10, 12, 17, 97

[6] Manoj Kumar Paras and Pooja Rani. Survey on electrical activity in earth's atmosphere. *Advanced Electromagnetics*, 7(4):34–45, 2018. ISSN 21190275. doi: 10.7716/aem.v7i4.737. pages xi, 21

[7] International Telecommunications Union (ITU). Nomenclature of the frequency and wavelength bands used in telecommunications. *Recommendation ITU/RV*, pages 431–438, 2015. pages 1

[8] Eran Greenberg and Colin Price. Diurnal variations of ELF transients and background noise in the Schumann resonance band. *Radio Science*, 42(2), 2007. doi: 10.1029/ 2006rs003477. pages 2, 32, 33, 34, 52

[9] Joseph John Thomson. *Notes on recent researches in electricity and magnetism: intended as a sequel to Professor Clerk-Maxwell's Treatise on electricity and magnetism.* Cambridge University Press, 1893. pages 2

[10] 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*, 7(2), 1952. ISSN 18657109. doi: 10.1515/zna-1952-0202. pages 2, 6

[11] Toshio OGAWA and Yutaka MURAKAMI. Schumann resonance frequencies and the conductivity profils in the atmosphere. *Contributions of the Geophysical Institute, Kyoto University*, 13:13–20, 1973. pages 3, 9

[12] PV Bliokh, Yu P Galyuk, EM Hünninen, AP Nikolaenko, and LM Rabinovich. Resonance effects in the earth-ionosphere cavity. *Radiophysics and Quantum Electronics*, 20(4):339–345, 1977. pages 3

[13] Earle Williams and Eugene Mareev. Recent progress on the global electrical circuit. *Atmospheric Research*, 135:208–227, 2014. pages 3

[14] AP Nickolaenko. Modern aspects of schumann resonance studies. *Journal of Atmospheric and Solar-Terrestrial Physics*, 59(7), 1997. pages 3, 4, 23, 32, 90

[15] Colin Price. ELF electromagnetic waves from lightning: The schumann resonances. *Atmosphere*, 7(9), 2016. ISSN 20734433. doi: 10.3390/atmos7090116. pages 4

[16] Toshio OGAWA, Yoshikazu TANAKA, and Teruo MIURA. On the frequency response of the ball antenna for measuring elf radio signals. *Special Contributions of the Geophysical Institute, Kyoto University*, 6:9–12, 1966. pages 4

125

[17] T Nakamura, C Hirose, R Hirose, S Hirooka, and H Sasaki. Observation of electric fields in the shallow sea using the stainless steel electrode antenna system. *Physics and Chemistry of the Earth, Parts A/B/C*, 31(4-9):352–355, 2006. pages 4

[18] Michael L Burrows. *ELF communications antennas*. Number 5. P. Peregrinus Stevenage, 1978. pages 4, 12, 31, 90

[19] Rosa M García Salvador, José A Gázquez Parra, and Nuria Novas Castellano. Characterization and modeling of high-value inductors in elf band using a vector network analyzer. *IEEE Transactions on Instrumentation and Measurement*, 62(2):415–423, 2012. pages 4, 12, 47

[20] DD Sentman and BJ Fraser. Simultaneous observations of schumann resonances in california and australia: Evidence for intensity modulation by the local height of the d region. *Journal of Geophysical Research: Space Physics*, 96(A9), 1991. pages 4, 20, 32, 60

[21] C Cano-Domingo, M Fernández-Ros, N Novas, and JA Gazquez. Diurnal and seasonal results of the schumann resonance observatory in sierra de filabres, spain. *IEEE Transactions on Antennas and Propagation*, 2021. pages 4, 28

[22] Tamás Bozóki, Gabriella Sátori, Earle Williams, Irina Mironova, Péter Steinbach, Emma C Bland, Alexander Koloskov, Yuri M Yampolski, Oleg V Budanov, Mariusz Neska, et al. Solar cycle-modulated deformation of the earth–ionosphere cavity. *Frontiers in Earth Science*, 9:689127, 2021. pages 4, 60

[23] Zenon Nieckarz, Stanisław Zięba, Andrzej Kułak, and Adam Michalec. Study of the periodicities of lightning activity in three main thunderstorm centers based on schumann resonance measurements. *Monthly weather review*, 137(12):4401–4409, 2009. pages 4, 60, 104

[24] Heng Yang and Victor P Pasko. Power variations of schumann resonances related to el nino and la nina phenomena. *Geophysical research letters*, 34(11), 2007. pages 5

[25] E Anyamba, E Williams, J Susskind, A Fraser-Smith, and Martin Fullekrug. The manifestation of the madden–julian oscillation in global deep convection and in the schumann resonance intensity. *Journal of the atmospheric sciences*, 57(8):1029–1044, 2000. pages 5

[26] Ciarán D Beggan and Malgorzata A Musur. Is the madden–julian oscillation reliably detectable in schumann resonances? *Journal of Atmospheric and Solar-Terrestrial Physics*, 190:108–116, 2019. pages 5, 104

[27] Soiram Ernesto Silva Artigas. Stochastic modeling of lightning occurrence by non-homogeneous poisson process. In *2012 International Conference on Lightning Protection (ICLP)*, pages 1–7. IEEE, 2012. pages 5, 43

[28] DA Chrissan and AC Fraser-Smith. A clustering poisson model for characterizing the interarrival times of sferics. *Radio Science*, 38(4):17–1, 2003. pages 5, 60

[29] Irina G Kudintseva, Alexander P Nickolaenko, Michael J Rycroft, and Anna Odzimek. Ac and dc global electric circuit properties and the height profile of atmospheric conductivity. *Annals of geophysics*, 59(5), 2016. pages 5, 26

[30] Earle R Williams. The schumann resonance: A global tropical thermometer. *Science*, 256(5060), 1992. pages 5, 20, 25

[31] Colin Price and Mustafa Asfur. Can lightning observations be used as an indicator of upper-tropospheric water vapor variability? *Bulletin of the American Meteorological Society*, 87(3), 2006. pages 5, 25

[32] DD Sentman. Schumann resonance effects of electrical conductivity perturbations in an exponential atmospheric/ionospheric profile. *Journal of Atmospheric and Terrestrial Physics*, 45(1), 1983. pages 5, 18

[33] Masashi Hayakawa, Kenji Ohta, Alexander P Nickolaenko, and Yoshiaki Ando. Anomalous effect in schumann resonance phenomena observed in japan, possibly associated with the chi-chi earthquake in taiwan. In *Annales Geophysicae*, volume 23. Copernicus GmbH, 2005. pages 5, 29

[34] George Neville Watson. The transmission of electric waves round the earth. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 95(673):546–563, 1919. pages 5

[35] WO Schumann. Über die dämpfung der elektromagnetischen eigenschwingungen des systems erde—luft—ionosphäre. *Zeitschrift für Naturforschung A*, 7(3-4):250–252, 1952. pages 6

[36] WO Schumann. Über die ausbreitung sehr langer elektrischer wellen um die erde und die signale des blitzes. *Il Nuovo Cimento (1943-1954)*, 9(12):1116–1138, 1952. pages 6

[37] WO Schumann and H König. Über die beobachtung von "atmospherics" bei geringsten frequenzen. *Naturwissenschaften*, 41(8):183–184, 1954. pages 6

[38] H Konig. Ultra-low-frequency atmospherics. In *1958 IRE International Convention Record*, volume 8, pages 128–133. IEEE, 1966. pages 6

[39] M Balser and CA Wagner. Measurements of the spectrum of radio noise from 50 to 100 cycles per second. *J. Res. Natl. Bur. Stand. US, Sect. D*, 64, 1960. pages 6, 10

[40] KG Budden. Liii. the propagation of very low frequency radio waves to great distances. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(352):504–513, 1953. pages 7

[41] James R Wait. The mode theory of vlf ionospheric propagation for finite ground conductivity. *Proceedings of the IRE*, 45(6):760–767, 1957. pages 7

[42] Jerry Shmoys. Long-range propagation of low-frequency radio waves between the earth and the ionosphere. *Proceedings of the IRE*, 44(2):163–170, 1956. pages 7

[43] Janis Galejs. Elf waves in the presence of exponential ionospheric conductivity profiles. *IRE Transactions on antennas and propagation*, 9(6):554–562, 1961. pages 7

[44] T Madden and W Thompson. Low-frequency electromagnetic oscillations of the earth-ionosphere cavity. *Reviews of Geophysics*, 3(2):211–254, 1965. pages 7, 24

[45] D Llanwyn Jones. Schumann resonances and elf propagation for inhomogeneous, isotropic ionosphere profiles. *Journal of Atmospheric and Terrestrial Physics*, 29(9): 1037–1044, 1967. pages 8

[46] T Røed Larsen and A Egeland. Fine structure of the earth-ionosphere cavity resonances. *Journal of Geophysical Research*, 73(15), 1968. pages 8

[47] Janis Galejs. Frequency variations of schumann resonances. *Journal of Geophysical Research*, 75(16), 1970. pages 8

[48] EA Mechtly, SA Bowhill, and LG Smith. Changes of lower ionosphere electron concentrations with solar activity. *Journal of Atmospheric and Terrestrial Physics*, 34(11):1899–1907, 1972. pages 8, 18

[49] A Tran and C Polk. Schumann resonances and electrical conductivity of the atmospher and lower ionosphere—i. effects of conductivity at various altitudes on resonance frequencies and attenuation. *Journal of Atmospheric and Terrestrial Physics*, 41(12): 1241–1248, 1979. pages 9

[50] A Tran and C Polk. Schumann resonances and electrical conductivity of the atmosphere and lower ionosphere—ii. evaluation of conductivity profiles from experimental schumann resonance data. *Journal of Atmospheric and Terrestrial Physics*, 41(12): 1249–1261, 1979. pages 9

[51] D. Hogg. The measurement of atmospheric radio noise in south africa in the low-frequency band. *Transactions of the South African Institute of Electrical Engineers*, 41(7):209–225, 1950. pages 10

[52] V. H. Rumsey. Traveling wave slot antennas. *Journal of Applied Physics*, 24(11): 1358–1365, 1953. doi: 10.1063/1.1721178. URL `https://doi.org/10.1063/1.1721178`. pages 11

[53] J. Bolljahn and R. Reese. Electrically small antennas and the low-frequency aircraft antenna problem. *Transactions of the IRE Professional Group on Antennas and Propagation*, 1(2):46–54, 1953. doi: 10.1109/T-AP.1953.27328. pages 11

[54] A. A. Hemphill. A magnetic radio compass antenna having zero drag. *IRE Transactions on Aeronautical and Navigational Electronics*, ANE-2(4):17–22, 1955. doi: 10.1109/TANE3.1955.4201422. pages 11

[55] C Polk and F Fitchen. Schumann resonances of the earthionosphere cavity–extremely low frequency reception at kingston, ri. *J. Res. Nat. Bur. Stand. US, Sect. D*, 66: 313–318, 1962. pages 12

[56] ROGER GENDRIN and ROBERT STEFANT. Magnetic records between 0.2 and 30 c/s. In *Propagation of Radio Waves at Frequencies Below 300 Kc/s: Proceedings of the Seventh Meeting of the AGARD Ionospheric Research Committee, Munich, 1962*, page 371. on behalf of Advisory Group for Aeronautical Research and Development. North . . . , 1964. pages 12

[57] GG Belyaev, A Yu Schekotov, AV Shvets, and AP Nickolaenko. Schumann resonances observed using poynting vector spectra. *Journal of atmospheric and solar-terrestrial physics*, 61(10):751–763, 1999. pages 12, 107

[58] D Llanwyn Jones and David T Kemp. Experimental and theoretical observations on the transient excitation of schumann resonances. *Journal of Atmospheric and Terrestrial Physics*, 32(6):1095–1108, 1970. pages 12, 17, 32

[59] DD Sentman. Magnetic elliptical polarization of schumann resonances. *Radio Science*, 22(04):595–606, 1987. pages 13, 19

[60] F Horner and J Harwood. An investigation of atmospheric radio noise at very low frequencies. *Proceedings of the IEE-Part B: Radio and Electronic Engineering*, 103 (12):743–751, 1956. pages 14

[61] AD Watt and EL Maxwell. Characteristics of atmospheric noise from 1 to 100 kc. *Proceedings of the IRE*, 45(6):787–794, 1957. pages 14

[62] K Furutsu and T Ishida. On the theory of amplitude distribution of impulsive random noise. *Journal of Applied Physics*, 32(7):1206–1221, 1961. pages 14

[63] Petr Beckmann. Amplitude probability distribution of atmospheric radio noise. *Radio science*, 68(6):723–736, 1964. pages 14, 31, 52

[64] James E Evans. Preliminary analysis of elf noise. Technical report, MAS-SACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1969. pages 15, 31, 52

[65] James R Wait. Project sanguine. *Science*, 178(4058):272–275, 1972. pages 15

[66] Andrew S Griffiths. Elf noise processing. Technical report, MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1972. pages 15, 31, 52

[67] J Evans and A Griffiths. Design of a sanguine noise processor based upon world-wide extremely low frequency (elf) recordings. *IEEE transactions on communications*, 22 (4), 1974. pages 16

[68] M Burrows and C Niessen. Elf communication system design. In *Ocean 72-IEEE International Conference on Engineering in the Ocean Environment*, pages 95–109. IEEE, 1972. pages 16, 31

[69] Eliot Marshall. Project elf finally wins a vote: But the navy's 25-year battle to build an antenna for submarine communication is far from over; researchers say it is rapidly becoming obsolete. *Science*, 221(4611):630–631, 1983. pages 16

[70] Anthony J Ferraro, Hai S Lee, Thomas W Collins, Maurice Baker, Douglas Werner, FM Zain, and Ping J Li. Measurements of extremely low frequency signals from modulation of the polar electrojet above fairbanks, alaska. *IEEE transactions on antennas and propagation*, 37(6), 1989. pages 16

[71] CHARLES Polk. Relation of elf noise and schumann resonances to thunderstorm activity. *Planetary electrodynamics*, 2, 1969. pages 17

[72] D Llanwyn Jones and DT Kemp. The nature and average magnitude of the sources of transient excitation of schumann resonances. *Journal of Atmospheric and Terrestrial Physics*, 33(4):557–566, 1971. pages 17, 32

[73] MD Clayton and Ch Polk. Diurnal variation and absolute intensity of world-wide light-ning activity, september 1970 to may 1971. In *Electrical Processes in Atmospheres*, pages 440–449. Springer, 1976. pages 18

[74] RB Anderson. Lightning research in south africa. *Transactions of the South African Institute of Electrical Engineers*, 68(2):30–39, 1977. pages 18

[75] BN Turman. Analysis of lightning data from the dmsp satellite. *Journal of Geophysical Research: Oceans*, 83(C10):5019–5024, 1978. pages 18

[76] Pavel Viktorovich Bliokh, Aleksandr Pavlovich Nicholaenko, and Iu F Fillipov. Schu-mann resonances in the earth-ionosphere cavity. *Stevenage*, 9, 1980. pages 18

[77] Charles Polk. Schumann resonances. *CRC Handbook of Atmospherics. Volume 1*, 1: 111–178, 1982. pages 18

[78] PS Cannon and Michael J Rycroft. Schumann resonance frequency variations during sudden ionospheric disturbances. *Journal of Atmospheric and Terrestrial Physics*, 44 (2), 1982. pages 18

[79] DD Sentman. Detection of elliptical polarization and mode splitting in discrete schumann resonance excitations. *Journal of atmospheric and terrestrial physics*, 51 (6), 1989. pages 19, 32

[80] M Sekiguchi, M Hayakawa, AP Nickolaenko, and Y Hobara. Evidence on a link between the intensity of schumann resonance and global surface temperature. In *Annales Geophysicae*, volume 24. Copernicus GmbH, 2006. pages 20

[81] Dennis J Boccippio, Earle R Williams, Stan J Heckman, Walter A Lyons, Ian T Baker, and Robert Boldi. Sprites, elf transients, and positive ground strokes. *Science*, 269 (5227), 1995. pages 20, 27, 32

[82] Eran Greenberg, Colin Price, Yoav Yair, Michal Ganot, József Bór, and Gabriella Sátori. Elf transients associated with sprites and elves in eastern mediterranean winter thunderstorms. *Journal of atmospheric and solar-terrestrial physics*, 69(13), 2007. pages 21, 32

[83] Jose A Gazquez, Rosa M Garcia, Nuria N Castellano, Manuel Fernandez-Ros, Alberto-Jesus Perea-Moreno, and Francisco Manzano-Agugliaro. Applied engineering using schumann resonance for earthquakes monitoring. *Applied Sciences*, 7(11), 2017. pages 21

[84] Constantinos I Votis, Giorgos Tatsis, Vasilis Christofilakis, Panos Kostarakis, Ch Repapis, et al. Design and implementation of schumann resonances sensor platform. *Journal of Engineering Science & Technology Review*, 9(6), 2016. pages 21

[85] Andrzej Kulak, Jerzy Kubisz, Slawomir Klucjasz, Adam Michalec, Janusz Mlynarczyk, Zenon Nieckarz, Michal Ostrowski, and Stanislaw Zieba. Extremely low frequency electromagnetic field measurements at the hylaty station and methodology of signal analysis. *Radio Science*, 49(6), 2014. pages 21, 34, 52

[86] Michał Dyrda, Andrzej Kulak, Janusz Mlynarczyk, Michał Ostrowski, Jerzy Kubisz, Adam Michalec, and Zenon Nieckarz. Application of the schumann resonance spectral decomposition in characterizing the main african thunderstorm center. *Journal of Geophysical Research: Atmospheres*, 119(23), 2014. pages 22

[87] Anirban Guha, Earle Williams, Robert Boldi, Gabriella Sátori, Tamás Nagy, József Bór, Joan Montanya, and Pascal Ortega. Aliasing of the schumann resonance background signal by sprite-associated q-bursts. *Journal of Atmospheric and Solar-Terrestrial Physics*, 165, 2017. pages 22, 32

[88] Manuel Fernández-Ros, Jose A. Gázquez Parra, Rosa M. García Salvador, and Nuria Novas Castellano. Optimization of the periodogram average for the estimation of the power spectral density (PSD) of weak signals in the ELF band. *Measurement*, 78, jan 2016. ISSN 0263-2241. doi: 10.1016/J.MEASUREMENT. 2015.10.006. URL https://www.sciencedirect.com/science/article/pii/S0263224115005369. pages 22, 32, 51

[89] Jose A.Gazquez Parra, Manuel Fernandez Ros, Nuria Novas Castellano, and Rosa M.Garcia Salvador. Techniques for Schumann Resonance Measurements:

A Comparison of Four Amplifiers with a Noise Floor Estimate. *IEEE Transactions on Instrumentation and Measurement*, 64(10), 2015. ISSN 00189456. doi: 10.1109/TIM.2015.2420376. pages 22, 47

[90] Toshio Ogawa and Masayuki Komatsu. 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, 2010. pages 22

[91] 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*, 54, 2019. ISSN 1944-799X. pages 22, 24, 97

[92] James R Wait. *Electromagnetic waves in stratified media: Revised edition including supplemented material*, volume 3. Elsevier, 2013. pages 23

[93] SJ Heckman, E Williams, and B Boldi. Total global lightning inferred from schumann resonance measurements. *Journal of Geophysical Research: Atmospheres*, 103(D24), 1998. pages 23

[94] AP Nickolaenko, G Sátori, B Zieger, LM Rabinowicz, and IG Kudintseva. Parameters of global thunderstorm activity deduced from the long-term schumann resonance records. *Journal of Atmospheric and Solar-Terrestrial Physics*, 60(3), 1998. pages 23, 90, 115

[95] Alexander P Nickolaenko, Masashi Hayakawa, and Yasuhide Hobara. Long-term periodical variations in global lightning activity deduced from the schumann resonance monitoring. *Journal of Geophysical Research: Atmospheres*, 104(D22), 1999. pages 23

[96] Yoshiaki Ando, Masashi Hayakawa, Alexander V Shvets, and Alexander P Nickolaenko. Finite difference analyses of schumann resonance and reconstruction of lightning distribution. *Radio Science*, 40(2), 2005. pages 24

[97] VV Surkov and M Hayakawa. Schumann resonances excitation due to positive and negative cloud-to-ground lightning. *Journal of Geophysical Research: Atmospheres*, 115(D4), 2010. pages 24, 43

[98] Olga Pechony and Colin Price. Schumann resonance parameters calculated with a partially uniform knee model on earth, venus, mars, and titan. *Radio science*, 39(5), 2004. pages 24, 30

[99] Robert Boldi, Earle Williams, and Anirban Guha. Determination of the global-average charge moment of a lightning flash using schumann resonances and the lis/otd lightning data. *Journal of Geophysical Research: Atmospheres*, 123(1):108–123, 2018. pages 25

[100] E Williams, Tamás Bozóki, Gabriella Sátori, C Price, Péter Steinbach, A Guha, Y Liu, CD Beggan, M Neska, R Boldi, et al. 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, 2021. pages 25

[101] Ernő Prácser and Tamás Bozóki. On the reliability of the inversion aimed to reconstruct global lightning activity based on schumann resonance measurements. *Journal of Atmospheric and Solar-Terrestrial Physics*, 235:105892, 2022. pages 25

[102] Vadim C Mushtak and Earle R Williams. Elf propagation parameters for uniform models of the earth–ionosphere waveguide. *Journal of atmospheric and solar-terrestrial physics*, 64(18), 2002. pages 26

[103] A Kulak, Stanisław Zikeba, Stanisław Micek, and Zenon Nieckarz. Solar variations in extremely low frequency propagation parameters: 1. a two-dimensional telegraph equation (tdte) model of elf propagation and fundamental parameters of schumann resonances. *Journal of Geophysical Research: Space Physics*, 108(A7), 2003. pages 26

[104] 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*, 193, 2019. pages 26

[105] 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*, 30(9), 2012. ISSN 09927689. doi: 10.5194/angeo-30-1321-2012. pages 26, 32

[106] IG Kudintseva, Yu P Galuk, AP Nickolaenko, and M Hayakawa. Modifications of middle atmosphere conductivity during sudden ionospheric disturbances deduced from changes of schumann resonance peak frequencies. *Radio Science*, 53(5), 2018. pages 27, 28

[107] Toshio Ogawa and Masayuki Komatsu. Analysis of q burst waveforms. *Radio Science*, 42(2), 2007. pages 27, 32

[108] AP Nickolaenko and LM Rabinowicz. Study of the annual changes of global lightning distribution and frequency variations of the first schumann resonance mode. *Journal of Atmospheric and Terrestrial Physics*, 57(11), 1995. pages 27

[109] X-Y Ouyang, Z Xiao, Y-Q Hao, and D-H Zhang. Variability of schumann resonance parameters observed at low latitude stations in china. *Advances in Space Research*, 56(7), 2015. pages 27, 28

[110] K Maki and T Ogawa. Elf emissions associated with earthquakes. *Journal of Atmospheric Electricity*, 3(1):41–44, 1983. pages 28

[111] Masashi HAYAKAWA. Electromagnetic phenomena related to earthquake prediction. *Terra Sci. Pub. Comp., Tokyo, 1994*, 1994. pages 28

[112] Antony C Fraser-Smith, Arman Bernardi, PR McGill, M_E Ladd, R_A Helliwell, and OG Villard Jr. Low-frequency magnetic field measurements near the epicenter of the ms 7.1 loma prieta earthquake. *Geophysical Research Letters*, 17(9), 1990. pages 28

[113] JY Liu, YI Chen, SA Pulinets, YB Tsai, and YJ Chuo. Seismo-ionospheric signatures prior to m 6.0 taiwan earthquakes. *Geophysical research letters*, 27(19), 2000. pages 28

[114] Kenji Ohta, Nobuo Watanabe, and Masashi Hayakawa. Survey of anomalous schumann resonance phenomena observed in japan, in possible association with earthquakes in taiwan. *Physics and Chemistry of the Earth, Parts A/B/C*, 31(4-9), 2006. pages 29

[115] AP Nickolaenko, M Hayakawa, M Sekiguchi, Y Ando, and K Ohta. Model modifications in schumann resonance intensity caused by a localized ionosphere disturbance over the earthquake epicenter. In *Annales Geophysicae*, volume 24. Copernicus GmbH, 2006. pages 29

[116] Yi Wang and Qunsheng Cao. Analysis of seismic electromagnetic phenomena using the fdtd method. *IEEE Transactions on Antennas and Propagation*, 59(11):4171–4180, 2011. pages 29

[117] Vasilis Christofilakis, Giorgos Tatsis, Constantinos Votis, Ioannis Contopoulos, Christos Repapis, and Vasilis 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*, 182, 2019. ISSN 1364-6826. doi: https://doi.org/10.1016/j.jastp.2018.11.009. URL `http://www.sciencedirect.com/science/article/pii/S1364682617305278`. pages 29

[118] Yuri P Galuk, Irina G Kudintseva, Alexander P Nickolaenko, and Masashi Hayakawa. Modifications of schumann resonance spectra as an estimate of causative earthquake magnitude: The model treatment. *Journal of Atmospheric and Solar-Terrestrial Physics*, 209, 2020. pages 29

[119] M Hayakawa, AP Nickolaenko, Yu P Galuk, and IG Kudintseva. Scattering of extremely low frequency electromagnetic waves by a localized seismogenic ionospheric perturbation: Observation and interpretation. *Radio Science*, 55(12), 2020. pages 29

[120] Carlos Cano-Domingo, Ruxandra Stoean, Nuria Novas-Castellano, Manuel Fernandez-Ros, Gonzalo Joya, and Jose A Gázquez-Parra. On the prospective use of

deep learning systems for earthquake forecasting over schumann resonances signals. *Engineering Proceedings*, 18(1):15, 2022. pages 29

[121] YURDANUR Tulunay, E Altuntas, E Tulunay, C Price, T Ciloglu, Y Bahadırlar, and ET Şenalp. 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), 2008. pages 30

[122] Tamás Bozóki, Ernő Prácser, Gabriella Sátori, Gergely Dálya, Kornél Kapás, and János Takátsy. Modeling Schumann resonances with schupy. *Journal of Atmospheric and Solar-Terrestrial Physics*, 196(October), 2019. ISSN 13646826. doi: 10.1016/j. jastp.2019.105144. pages 30

[123] 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)*, 76(9), 2017. doi: 10.1615/TelecomRadEng.v76.i9.60. pages 30, 82

[124] IG Kudintseva, SA Nikolayenko, AP Nickolaenko, and M Hayakawa. Synthesis of schumann resonance background signal in time domain. *International Journal of Electronics and Applied Research (IJEAR)*, 4(1):1–23, 2017. pages 30

[125] M. Ralchenko, M. Svilans, C. Samson, and M. Roper. Finite-difference time-domain modelling of through-the-earth radio signal propagation. *Computers & Geosciences*, 85:184–195, 2015. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo. 2015.09.018. URL https://www.sciencedirect.com/science/article/pii/ S0098300415300601. pages 30

[126] Martin Füllekrug. Simulation of earth-ionosphere cavity resonances with lightning flashes reported by otd/lis. *Journal of Geophysical Research: Atmospheres*, 126(24): e2021JD035721, 2021. pages 30

[127] Joanna Kozakiewicz, Andrzej Kulak, and Janusz Mlynarczyk. Analytical modeling of schumann resonance and elf propagation parameters on mars with a multi-layered ground. *Planetary and Space Science*, 117, 2015. pages 30

[128] Alexander P. Nickolaenko, Bruno P. Besser, and Konrad Schwingenschuh. Model computations of schumann resonance on titan. *Planetary and Space Science*, 51(13), 2003. ISSN 0032-0633. doi: https://doi.org/10.1016/S0032-0633(03) 00119-3. URL `https://www.sciencedirect.com/science/article/pii/ S0032063303001193`. pages 30

[129] F. Simões, R. Grard, M. Hamelin, J.J. López-Moreno, K. Schwingenschuh, C. Béghin, J.-J. Berthelier, B. Besser, V.J.G. Brown, M. Chabassière, P. Falkner, F. Ferri, M. Fulchignoni, R. Hofe, I. Jernej, J.M. Jeronimo, G.J. Molina-Cuberos, R. Rodrigo, H. Svedhem, T. Tokano, and R. Trautner. A new numerical model for the simulation of elf wave propagation and the computation of eigenmodes in the atmosphere of titan: Did huygens observe any schumann resonance? *Planetary and Space Science*, 55 (13), 2007. ISSN 0032-0633. doi: https://doi.org/10.1016/j.pss.2007.04.016. Titan as seen from Huygens. pages 30

[130] Ralph D. Lorenz and Alice Le Gall. Schumann resonance on titan: A critical re-assessment. *Icarus*, 351, 2020. ISSN 0019-1035. doi: https://doi.org/10.1016/j. icarus.2020.113942. pages 30

[131] D. Iudin A. Nickolaenko, P. Colin. Hurst exponent derived from natural terrestrial radio noise in Schumann resonance band. *Geophysical Research Letters*, 27(19), 2000. pages 31, 32, 34, 52, 65, 78

[132] M Hayakawa, K Hattori, AP Nickolaenko, and LM Rabinowicz. Relation between the energy of earthquake swarm and the hurst exponent of random variations of the geomagnetic field. *Physics and Chemistry of the Earth, Parts A/B/C*, 29(4-9):379–387, 2004. pages 32

[133] A Tzanis and D Beamish. Time domain polarization analysis of schumann resonance waveforms. *Journal of atmospheric and terrestrial physics*, 49(3):217–229, 1987. pages 32

[134] Daniel Labendz. Investigation of schumann resonance polarization parameters. *Journal of atmospheric and solar-terrestrial physics*, 60(18):1779–1789, 1998. pages 32

[135] M Hayakawa and T Otsuyama. Fdtd analysis of elf wave propagation in inhomogeneous subionospheric waveguide models. Technical report, UNIVERSITY OF ELECTRO-COMMUNICATIONS TOKYO (JAPAN), 2002. pages 32

[136] Huan Hao, Huali Wang, Liang Chen, Jun Wu, Longqing Qiu, and Liangliang Rong. Initial results from squid sensor: analysis and modeling for the elf/vlf atmospheric noise. *Sensors*, 17(2):371, 2017. pages 32

[137] Carlos Cano-Domingo, Nuria Novas Castellano, Manuel Fernandez-Ros, and Jose Antonio Gazquez-Parra. Segmentation and characteristic extraction for schumann resonance transient events. *Measurement*, 194:110957, 2022. pages 33

[138] Mourad Khayati, Ines Arous, Zakhar Tymchenko, and Philippe Cudré-Mauroux. Orbits: online recovery of missing values in multiple time series streams. *Proceedings of the VLDB Endowment*, 14(3):294–306, 2020. pages 33

[139] Eben Afrifa-Yamoah, Ute A Mueller, SM Taylor, and AJ Fisher. Missing data imputation of high-resolution temporal climate time series data. *Meteorological Applications*, 27(1):e1873, 2020. pages 34

[140] Mohammad Maghrour Zefreh and Adam Torok. Single loop detector data validation and imputation of missing data. *Measurement*, 116:193–198, 2018. pages 34

[141] Yifan Zhang and Peter J Thorburn. A dual-head attention model for time series data imputation. *Computers and Electronics in Agriculture*, 189:106377, 2021. pages 34

[142] Getnet D Betrie, Rehan Sadiq, Solomon Tesfamariam, and Kevin A Morin. 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, 2016. pages 34

[143] Christian Seiler and Christian Heumann. Microdata imputations and macrodata implications: Evidence from the ifo business survey. *Economic Modelling*, 35:722–733, 2013. pages 34

[144] Scott J Cook and Nils B Weidmann. Lost in aggregation: Improving event analysis with report-level data. *American Journal of Political Science*, 63(1):250–264, 2019. pages 34

[145] Olivier Alata, Christian Olivier, and Yannis Pousset. Law recognitions by information criteria for the statistical modeling of small scale fading of the radio mobile channel. *Signal processing*, 93(5):1064–1078, 2013. pages 34

[146] Vance W Berger and YanYan Zhou. Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*, 2014. pages 35

[147] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951. pages 35

[148] Fabio Baselice, Giampaolo Ferraioli, Vito Pascazio, and Antonietta Sorriso. Denoising of mr images using kolmogorov-smirnov distance in a non local framework. *Magnetic resonance imaging*, 57:176–193, 2019. pages 36

[149] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 1948. pages 36, 61

[150] Ken P Burnham and David R Anderson. P values are only an index to evidence: 20th-vs. 21st-century statistical science. *Ecology*, 95(3), 2014. pages 36

[151] In Jae Myung. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1):90–100, 2003. pages 37, 38

[152] Anthony WF Edwards. The history of likelihood. *International Statistical Review/Revue Internationale de Statistique*, pages 9–15, 1974. pages 38

[153] Stephen M Stigler. The epic story of maximum likelihood. *Statistical Science*, pages 598–620, 2007. pages 38

[154] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974. pages 39

[155] Eric-Jan Wagenmakers and Simon Farrell. Aic model selection using akaike weights. *Psychonomic bulletin & review*, 11(1), 2004. pages 39

[156] 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*, 11(9), 2016. ISSN 1932-6203. pages 39

[157] Xiaoming Chen. Using akaike information criterion for selecting the field distribution in a reverberation chamber. *IEEE Transactions on Electromagnetic Compatibility*, 55 (4), 2013. ISSN 00189375. doi: 10.1109/TEMC.2012.2225107. pages 39, 42

[158] Long Yan, Hong Wang, Xuan Zhang, Ming-Yue Li, and Juan He. Impact of meteorological factors on the incidence of bacillary dysentery in beijing, china: A time series analysis (1970-2012). *PLOS ONE*, 12(8), 08 2017. doi: 10.1371/journal.pone. 0182937. URL `https://doi.org/10.1371/journal.pone.0182937`. pages 39

[159] Petre Stoica and Yngve Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004. pages 40

[160] Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228, 2012. pages 40

[161] Jie Ding, Vahid Tarokh, and Yuhong Yang. Model selection techniques: An overview. *IEEE Signal Processing Magazine*, 35(6):16–34, 2018. pages 40

[162] Ming Yang Jeremy Tan. *Critical study of AIC model selection techniques*. PhD thesis, University of Illinois at Urbana-Champaign, 2012. pages 40

[163] Manuel Fernández Ros. *Técnicas para la adquisición y tratamiento optimizado de señales débiles enmascaradas con ruido en la banda elf*. PhD thesis, Universidad de Almería, 2015. pages 47

[164] Ali Abdi, Cihan Tepedelenlioglu, Mostafa Kaveh, and Georgios Giannakis. On the estimation of the k parameter for the rice fading distribution. *IEEE Communications letters*, 5(3):92–94, 2001. pages 53

[165] Michéle Basseville. Divergence measures for statistical data processing—an annotated bibliography. *Signal Processing*, 93(4):621–633, 2013. pages 61

[166] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951. pages 61

[167] Magnus Ekström. Alternatives to maximum likelihood estimation based on spacings and the kullback–leibler divergence. *Journal of Statistical Planning and Inference*, 138(6):1778–1791, 2008. pages 61

[168] Tryphon T Georgiou and Anders Lindquist. Kullback-leibler approximation of spectral density functions. *IEEE Transactions on Information Theory*, 49(11):2910–2917, 2003. pages 61

[169] Manuel Soler-Ortiz, Manuel Fernández Ros, Nuria Novas Castellano, and Jose A Gázquez Parra. A new way of analyzing the schumann resonances: A statistical approach. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021. pages 64

[170] Carlos Cano Domingo, Manuel Fernandez Ros, Nuria Novas Castellano, and José A Gázquez Parra. Diurnal and seasonal results of the schumann resonance observatory in sierra de filabres, spain. *IEEE Transactions on Antennas and Propagation*, 69(10): 6680–6690, 2021. pages 86

[171] Toshio Ogawa, Yoshikazu Tanaka, and Michihiro Yasuhara. 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, 1969. ISSN 00221392. doi: 10.5636/jgg.21.447. pages 97

[172] Rachel I Albrecht, Steven J Goodman, Dennis E Buechler, Richard J Blakeslee, and Hugh J Christian. Where are the lightning hotspots on earth? *Bulletin of the American Meteorological Society*, 97(11):2051–2068, 2016. pages 104, 107

The ELF spectrum is the range of frequencies contained between 3 and 300 Hz. In this range the Schumann resonance can be measured; an electromagnetic phenomenon that manifests in Earth's lower atmosphere, its main source being global lightning activity.

The lack of recent studies that analyze the temporal component of the signal motivate the development of this thesis, in which statistical methods have been applied to analyze over 5 years of Schumann resonance data captured by the extremely low frequency observatory that research group TIC-019 installed and currently mantains in Sierra de los Filabres, Almería.

A brand new methodology to work with natural electromagnetic signals emerges from this work, allowing process automatically ELF records by splitting them into segments, then classifying each segment under the statistical distribution that best fits the data and storing its most relevant parameters for posterior analysis. This methodology represents an additional way to extract information from the ELF records, complementing the commonly used frequency based approaches.

Analyzing the methodology results it is shown that each distribution is related to identifiable states of the signal, effectively performing an automatic classification that allows to discern events contained within the segments.

A technique to quantify the signal variation over time based on Kolmogorov-Smirnov distance was also tested, showing a way of measuring the signal stability to select specific segments in high precision studies and hinting at a different, automated way to measure lightning activity from the signal.

UNIVERSIDAD DE ALMERÍA

ESCUELA SUPERIOR DE INGENIERÍA