Computers & Geosciences 165 (2022) 105148

Contents lists available at ScienceDirect

Computers and Geosciences

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

Schumann resonance data processing programs and four-year measurements from Sierra Nevada ELF station

A. Salinas [a,*], J. Rodríguez-Camacho [b], J. Portí [b], M.C. Carrión [b], J. Fornieles-Callejón [a], S. Toledo-Redondo [c]

[a] Department of Electromagnetism and Matter Physics, University of Granada, Granada, Spain
[b] Department of Applied Physics, University of Granada, Granada, Spain
[c] Department of Electromagnetism and Electronics, University of Murcia, Murcia, Spain

ARTICLE INFO

Keywords:
Schumann resonances
Data processing
ELF station

ABSTRACT

In this work, we present to the scientific community the measurements taken during four years, from March 2013 to February 2017 inclusive, by the Extremely Low Frequency Sierra Nevada station, Spain, together with the data processing programs developed in Python (version 3.8) to extract the Schumann resonance (SR) parameters (i.e., amplitudes, resonant frequencies, resonance widths) in 10 min time periods from these records. The measurements correspond to the voltage induced by the atmospheric electromagnetic field at the north-south and east-west oriented magnetometers of the station. The process comprises four stages. The spectrum, calibrated in the frequency band ranging from 6 Hz to 25 Hz, is obtained at the first stage using the Welch method with Hann windows. The second step eliminates the anthropogenic noise generated by different undesired sources. Next, a non-linear fit of the measured spectrum combining Lorentzian functions together with a linear term is carried out in order to identify the presence of SRs and quantitatively characterize them. This third step is carried out using the Python package Lmfit, which implements the Levenberg-Marquad algorithm. Finally, a compact and easy-to-read output is generated at the fourth stage, using the power of the Numpy arrays and the npz format. In addition, four Jupyter notebooks with the description of the code and the possibility of their use in interactive mode are presented as supplementary material with this paper.

1. Introduction

The Earth's electromagnetic cavity consists of the good dielectric atmosphere, limited by two conducting boundaries: the planet surface and the lower ionosphere. The main electromagnetic source of this cavity is the lightning activity, mainly produced by storms. The Earth's electromagnetic cavity has eigenfrequencies around 8, 14, and 21 Hz, for the first three modes, respectively. Lightning produces a broadband electromagnetic spectrum that reaches the MHz band, but, due to the conductivity of the atmosphere at these frequencies, only Extremely Low Frequency (ELF) components, up to ~50 Hz, are capable of completing various turns around the planet before they almost completely attenuate and become undetectable. Multiple lightning sources are active at any time around the globe, exciting ELF modes which are known as Schumann Resonances (SRs) (Schumann, 1952). SRs were predicted in 1952, and first detected by Balser and Wagner (1962). A historical review on the progress of SR studies can be found in (Besser 2007).

There is plenty of specialized scientific literature on Atmospheric Electrodynamics in general and on SRs in particular. Worthy examples are the books (Nickolaenko and Hayakawa, 2002, 2014) and the reviews (Price, 2016; Simões et al., 2012; or Sátori et al., 2013). The SRs constitute the AC part of the Global Electric Circuit (GEC) (Rycroft et al., 2008) and their variations have been linked to the average temperature of the surface near tropical latitudes (E. R. Williams, 1992) and to the global lightning activity (Füllekrug and Fraser-Smith, 1997; Prácser et al., 2019; Toledo-Redondo et al., 2010; Williams et al., 2014). SRs are also a powerful tool to monitor global climate phenomena, such as El Niño (Williams et al., 2021) and also to study fluctuations in the ionosphere at the scale of years (e.g., (Bozóki et al., 2021; Koloskov et al., 2020; Nickolaenko et al., 2015).

SRs on Earth are a very faint signal, of the order of a few pT for the magnetic field components, i.e., roughly six orders of magnitude below the geo-magnetic field strength at the Earth's surface. Therefore, in order to measure and characterize them, the detectors must have high

* Corresponding author.
E-mail address: asalinas@ugr.es (A. Salinas).

https://doi.org/10.1016/j.cageo.2022.105148
Received 16 November 2021; Received in revised form 7 April 2022; Accepted 16 May 2022
Available online 26 May 2022

sensitivity and must be optimized to provide a good signal to noise ratio (Fornieles-Callejón et al., 2015). Most ELF stations devoted to measure SRs typically comprise two magnetometers which lie in the plane of the surface, plus one vertical electric antenna (e.g., Nickolaenko and Hayakawa, 2014). Nowadays, there are several SR stations spread all around the globe and various studies benefit from multi-point observations (Roldugin et al., 2004; Schlegel and Füllekrug, 2000; Sentman and Fraser, 1991; Shvets and Hayakawa, 2011). Despite the interest in ELF station global networks to share data and findings, the actual situation is that each research group usually works with their own data, since details on the data measured, the signal processing techniques and the codes to implement them are not usually available for the scientific community.

In this sense, two contributions are introduced in this work with regards to previous studies on SR measurements which may be of interest for those research groups that are working or want to start research on SRs and related problems such as the GEC. It can also be useful for those research groups in Atmospheric Sciences and Space Weather to make correlations between different measurements. The first contribution is that the scientific community is granted access to four years of raw and processed data from the Sierra Nevada ELF station measurement records. It seems reasonable to think that providing the data is of high interest to pursue a common goal set by the SR research community in particular, and by the atmospheric electrodynamics research community, in general, as it is the creation of a shared database with records of the different worldwide ELF stations. Our intention is not to set a standard for the processing and the format of the data but to propose a starting point. The second contribution of this work is concerned with the presentation of the codes for obtaining the SR parameters and other related results from these data. The analysis of the four-year long records may demand a huge amount of code lines, which can make it difficult to fully understand the algorithms and, therefore, may hinder their use by other researchers. For this reason, the Jupyter notebooks (Granger and Perez, 2021) are presented together with this paper in the supplementary material, containing the Python code, as well as the explanations needed to understand the processing of the data. Providing the data and the self-explanatory Jupyter notebooks focuses on the concepts and tools behind reporting modern data analyses in a reproducible manner: any supposed scientific achievement must be verified and must serve as a basis for building new achievements. In this sense, the notebooks from Jupyter project allow to publish data analyses in a single document that permit other researchers to easily carry out the same analysis to obtain the same results for the provided data and comparable results for their own measurements.

This paper is structured as follows. Section 2 describes the main characteristics of the Sierra Nevada ELF station and the structure of the data in the different repositories. Section 3 gives the skeleton of the programs and the blocks in which they are structured. In Sections 4 to 7, a detailed description of each block mentioned above is presented. Finally, Section 8 summarizes the main conclusions of this work.

2. The station and the data

2.1. The station

The ELF measurement station is located at the heart of the Sierra Nevada National Park, at 2500 m above sea level, in the area surrounding the mountain hut, *Refugio del Poqueira* (37°02′N, 3°19′W) and it has two magnetometers, North-South (NS) and East-West (EW) oriented. The magnetometers have been optimized with the specific purpose of measuring the first three SRs with a very high signal-to-noise ratio (28 dB for a time-varying signal of 1 pT in the band from 6 Hz to 25 Hz).

The station data acquisition system works at a sampling frequency of $f_m = 256$ Hz and generates a raw data file for each sensor and hour. The sampling time interval is 3906 μs, which is slightly less than $1/f_m$. In 1 h file, a total of 921,600 samples are taken, so the time length of measurements of each file is 3599.7696 s, instead of 3600 s, which implies a delay of 0.2304 s every hour. For this reason, the time of the first date of each hour file is different to the next hour, and so on.

The measurements correspond to the output of the sensor, including the amplification and digitization system. Measurements are taken between ±10 V limits, greater values produce saturation. The digitization system employs 16 bits equally distributed. The amplification system uses a Butterworth filter of tenth order to dampen the effect of the 50 Hz network signal. Due to the specific characteristics of the designed filter, it is very inaccurate to obtain the phase of the transfer function. That's why the scaling function shown below only defines the amplitude.

The magnetometers have been calibrated for frequencies between 6 Hz and 25 Hz. Considering the equivalent circuit of the magnetometers, (Fornieles-Callejón et al., 2015), their frequency response corresponds to an RLC circuit that exhibits an almost flat response for frequencies below 25 Hz, approximately, which defines the upper bandwidth limit. In order to avoid the $1/f$ noise, the lower bandwidth limit has been stablished to 6 Hz. To convert the voltage induced in the sensors into a magnetic field, the calibration function must be defined for each frequency of the spectrum. Previously, it is necessary to pass the measurements from the time domain to the frequency domain using the Fourier transform. From the theorical RLC model of the magnetometers, the calibrated magnetic field is given by:

$$B(f) = S_c(f)V(f), \tag{1}$$

with f in the (6 Hz, 25 Hz) interval. $B(f)$ is the magnetic field (in T/ $\sqrt{Hz}$), $V(f)$ is the transformed voltage amplitude and $S_c(f)$ is the calibration coefficient for each frequency f. As mentioned above, the specific design of the Butterword filter presents difficulties in accurately measuring the phase, therefore, we only apply the calibration function to the amplitude of the Fourier transformed voltage and all the quantities in (1) are to be understood as the amplitude of the corresponding quantity.

The function $S_c(f)$ is given by:

$$S_c(f) = \frac{1}{2500}\ \frac{1}{1.9\cdot 10^6}\ \frac{1}{f}\sqrt{\left(1-\left(\frac{f}{f_0}\right)^2\right)^2 + \left(2\delta\frac{f}{f_0}\right)^2}, \tag{2}$$

where 1/2500 is the amplification factor, $1/(1.9\cdot 10^6)$ is the magnetometer sensitivity, $1/f$ is given by Faraday's induction law. Finally, f_0 and δ are the resonance frequency and dumping factor of the equivalent RCL of the measurement system, respectively, and they can be calculated from:

$$f_0 = \frac{1}{2\pi\sqrt{LC_T}}, \quad \delta = \frac{R}{2}\sqrt{\frac{C_T}{L}}, \tag{3}$$

where $R = 320$ kΩ is the magnetometer resistance, $L = 298$ kH is the magnetometer inductance and $C_T = 40$ pF is the magnetometer capacitance. The technical characteristics of the station are described in more detail in (Fornieles-Callejón et al., 2015).

2.2. The data

The Sierra Nevada ELF station was deployed on July 17, 2012 and has been operating almost continuously since March 2013. The measurement period made available with this paper runs from March 2013 to February 2017. The days for which no measurements are available are: 2014, July 2 and 3 and November 5 to 18; 2016, from July 4 to 6.

The raw data of the station corresponding to the aforementioned measurement periods can be downloaded from the following Zenodo repositories (Salinas et al., 2022a, b, c, d). Each repository corresponds to one year of measurements except the first one that contains the data for the years 2013 (from March) and 2017 (January and February).

Each data folder is named with the four digits corresponding to the specific year, e.g., 2014. Each year contains folders for each month,

named with the four digits corresponding to the year and month, e.g., 1412 stands for the raw data during December 2014. Within each month folder, the data and information files for each sensor and each hour are available. Data are stored in files containing measurements corresponding to a time period of approximately 1 h. The filenames begin with a common part, "smplGRTU1_sensor_", followed by a specific part to denote the sensor used (0 for the NS orientation and 1 for the EW orientation), the date and the initial time of the measurements recorded. For example, smplGRTU1_sensor_0_1412010430 stands for data measured by the NS-oriented magnetometer, during the year 2014, month 12, day 01, hour 04, and starting minute 30. The information file has the same name but ends with _info.txt. Each information file, among other data, contains the specific time in which the first sample of the corresponding data file is taken. The starting time for each file is determined using GPS signals and has a resolution of milliseconds. Since this time of the first sample in each file is variable and part of the filename, the filenames in each month folder are not completely determined a priori. For this reason and to ease the task of the different programs used, each month folder includes two files, 'ficheros0′ and 'ficheros1′, for sensor NS and EW respectively, that contain the set of filenames for that specific month. Therefore, for each normally measured day, we have 24 data files and also 24 information files for each sensor. Each hour data file occupies 1.8 MB, so each month has roughly a data volume of 2.6 GB.

The data in each 1 h data file is in 16-bit binary format and its conversion to voltage involves multiplying by a factor defined by $10/2^{15}$, to describe the ± 10 V saturation limits and the use of a 16-bit of the A/D system, one bit for the sign and 15 bits for amplitude.

3. The code: an overview

The software codes to extract the SR parameters presented with this work are programmed in Python (version 3.8) using the modules Numpy,[1] Matplotlib,[2] Lmfit,[3] and Interact[4] (this one is only used in notebooks).

The main blocks and programs developed are the following:

1. 'Reading_Fourier.py'. It reads the raw data in the time domain and obtains the amplitude spectrum in the calibrated band of the measurement system.
2. 'Anthropo.py'. When present, this program eliminates the anthropogenic noise in the amplitude spectrum of the previous step.
3. 'Lorentz_Lmfit.py'. The program applies a non-linear fit using the 'Lmfit' module of Python to generate a Lorentzian fitted spectrum together with its corresponding error parameters.
4. 'Packaging.py'. This program carries out a final storage of the relevant data. The program generates a file for each month and sensor that allows a direct reading and recovering of the structure (arrays of the Numpy package of Python) of all the processed data generated by the previous packages for subsequent analysis purposes.

The full set of four programs may be of interest to those researchers who are interested in the direct measurements of the Sierra Nevada station or in processing their own raw data. They are welcome to use the program of block 1 using the ELF station raw data that can be downloaded from the Zenodo repositories (Salinas et al., 2022a, b, c, d) as mentioned above. But the code is developed in modular form, thus, as regards those researchers interested in working with the different phases of the spectrum, with or without anthropogenic noise, with the spectrum fitted or not, they can use the output files from programs in block 2 or 3.

[1] https://numpy.org/.
[2] https://matplotlib.org/.
[3] https://lmfit.github.io/lmfit-py/.
[4] https://pypi.org/project/interact/.

There are several repositories that can help in this task. The University of Granada repositories http://hdl.handle.net/10481/73978 and http://hdl.handle.net/10481/73978 contain all the files of the complete process (blocks 1 to 4) corresponding to December 2014 and March 2015 respectively. They have the same structure of years and months as the raw data folder. Within each month, output files for each step of the processing can be found. The code files, notebooks and one day (March 1, 2015) input/output files can be downloaded from the GitHub repository https://github.com/asalinas62/Sierra_Nevada_SR.

If the researcher's interest is exclusively concentrated in the use of the final results of the analysis, they can directly use the files resulting from block 4. These files can be downloaded from the University of Granada repository https://digibug.ugr.es/handle/10481/71563. This folder stores the final SR results of the raw data processing from the period of time between March 2013 to February 2017. These files have been used to carry out the Schumann resonance long-term analysis presented in (Rodríguez-Camacho et al., 2022), where a first analysis of the data provided with this work can be found.

The input and output files associated to the programs used for each intermediate processing stage are:

1. 'Reading_Fourier.py': the amplitude spectra and related information generated by this program are stored in the following files (December 2014 is used as an example):
   - SR1412_media_0,1: it contains the amplitude spectra within the calibrated band of the magnetometers (6 Hz–25 Hz), for the NS sensor (sensor 0) and for the EW sensor (sensor 1), respectively. The writing format is ASCII in one column.
   - SR1412_satper_0,1: it stores the percentage of those 10 s intervals that show saturation within each 10 min interval. There is a file for each sensor. The writing format is ASCII in one column.
2. 'Anthropo.py': the anthropogenic noise detected in the measurements is eliminated from the spectrum. The following output file is generated:
   - SR1412_mediaNA_0,1: similar to SR1412_media_0,1 but with the anthropogenic noise filtered from the amplitude spectra. The writing format is ASCII in one column.
3. 'Lorentz_Lmfit.py': a Lorentzian fitting of the amplitude spectrum is made. The output files for this step are:
   - SR1412_mediaLO_0,1: it contains the parameters of the Lorentzian adjustment of the spectra in amplitude for each 10 min interval and sensor. The writing format is ASCII in one column.
   - SR1412_mediaLOC_0,1: it includes the fitting error parameter of the χ^2 function in each spectrum. The writing format is ASCII in one column.
   - SR1412_mediaLOE_0,1: it contains the errors that the fitting program assigns to each calculated fitting parameter. The writing format is ASCII in one column.

As a summary of the whole process, for each month of measurements (December 2014 and NS sensor is used as an example) and each sensor the procedure is as follows. The raw data from the station is arranged in files 1 h long. The 'Reading_Fourier.py' program calculates the amplitude spectrum for the frequency band (6 Hz–25 Hz) and the saturation index for each 10 min interval and they are saved in two files, SR1412_media_0 and SR1412_satper_0, respectively. The first file is used to filter the anthropogenic noise with the 'Anthropo.py' program and the output spectra are stored in the file SR1412_mediaNA_0. This file in turn serves as input for the 'Lorentz_Lmfit.py' program, which returns the parameters of the Lorentzian fit. Finally, selected outputs of each block are stored.

The general procedure for processing data from an ELF station with the code presented in this work may have many particularities that must be taken into account if it is to be adapted to data from other stations. Nevertheless, we can establish the following general inputs to the programs that can be adapted to each particular ELF station:

1. Define the measurement interval generated by the station: in our case, one month is divided in 1 h data files.
2. Define the characteristics of the A/D system: sampling frequency, measurement range and number of bits.
3. Define the calibration function to convert voltage into magnetic field. In the case of the Sierra Nevada ELF station and due to the characteristics of the sensors, this step can only be taken in the frequency domain.
4. In case there is anthropogenic noise with characteristics similar to those found in Sierra Nevada: define the frequency bands with possible anthropogenic noise.

A more detailed description of the programs corresponding to this overview is carried in the following sections.

3. Signal processing step 1: the Reading_Fourier.py program

This program reads the raw data and obtains the amplitude spectrum for each component of the horizontal magnetic field. The code and details of this stage are described in the notebook 'Reading_Fourier_docu.ipynb'.

The data for each hour and sensor is processed in 10 min intervals. Since the noise level is large due to the low power of SRs, in order to get a natural stabilization of the SR spectra, we apply the Welch method to improve the signal to noise ratio. The 10 min time signal is divided into 10 s segments, overlapped by 5 s, and a Fast Fourier Transform (FFT) is calculated for each of them after applying a Hann window. Following Nickolaenko and Hayakawa, 2002, pg. 170, the resulting power spectra are then averaged, and the calibration function is applied to the obtained amplitude spectrum, see also (Price and Melnikov, 2004), providing physical units of pT/ $\sqrt{Hz}$.

To apply Welch's method, the FFT of a set of n data is made, taken in groups of m data and with a shift of p data. The number of intervals generated from n, m and p is $[(n-m)/p]+1$, where '[ ]' stands for the integer part. The value of n corresponds to the amount of data in a 10 min interval. The value of $m = 2560$ corresponds to the number of measurements in a 10 s sub-interval and p corresponds to a 5 s shift interval, which means that consecutive FFTs have and overlap of 5 s.

As mentioned above, we use the Hann's window with m samples that is defined by:

$$h(i) = 1 - \cos^2(\pi(i-1)/m) \tag{4}$$

for $i = 1,\ldots,m$, with the normalizing factor $\sqrt{\sum_{i=1}^{m} h(i)^2/m}$. The m data window is expanded with zeros to 10^{23} samples. This new window size and the sampling frequency define a frequency increment of 31.25 mHz.

The FFT is done with the Numpy subroutine 'fft.rfft' that takes advantage of the real character of the data. The Hann window and Welch method have been easily implemented using Numpy arrays.

In the analysis of each 10 min interval through the FFT of the 10 s sub-intervals, each 10 s time span including saturated signals is eliminated and not considered in the average for its corresponding 10 min interval. According to the number of 10 s time portions eliminated, the percentage of saturation for each 10 min span is defined. This saturation level may be indicative of the presence of transient sources in the ELF band. Due to the large size of the Earth's electromagnetic cavity, it is likely that its influence in global quantities is not relevant but may be indicative and interesting to detect local phenomena. The study of the possible correlation of this percentage of saturation with global or local variations of the different parameters associated with the SRs, amplitudes, resonance frequencies and resonance widths is a pending task which is beyond the scope of this work.

The following directory tree is considered in this program. The input files for this code contain the raw data measured at the station, which are stored in a folder named 'S_N_Data' (Sierra Nevada Data). The output files are located in the folder named 'S_N_FD' (Sierra Nevada, Frequency Domain) which contains another sub-folder with the month under study in a scheme which resembles that of the 'S_N_Data' folder. As previously mentioned, the 'Reading_Fourier.py' code generates two output files: one stores the spectrum and the other includes the percentage of saturation for each 10 min interval in that specific month. There are independent files for each magnetometer. The output file is organized as one column data using ASCII format. As regards the spectrum, the file includes an amplitude value for each frequency between 6 Hz and 25 Hz with a frequency resolution of 31.25 mHz, while there is a unique value of the saturation level for each 10 min interval.

Fig. 1 shows an example of the raw data measured by one of the sensors, at 256 samples/s, which is the input for program 'Reading_Fourier.py'. The ordinate axis represents volts at the magnetometers due to the fact that the calibration function $Sc(f)$ required to translate to magnetic field values is defined in the frequency domain and the data are shown in time domain. Fig. 2 shows an example of the amplitude spectra for the two sensors obtained as output of the program 'Reading_Fourier.py' for a specific 10 min interval spectrum, such as that shown in Fig. 1. The three SR peaks can be clearly distinguished at roughly 8, 14, and 20 Hz. In addition, there are fluctuations of the spectra due to the low energy of the signal (below 0.4 pT/ $\sqrt{Hz}$ for the whole 6 Hz–25 Hz band). As stated before, the electromagnetic field energy that can travel few turns around the cavity is very small, due to the losses caused by the conductivity of the Earth's atmosphere.

4. Signal processing step 2: elimination of the anthropogenic noise

Despite the site to set the Sierra Nevada ELF station was carefully chosen to avoid human activity, an increase of this activity occurred in the vicinity of the station soon after its deployment, and, since then, the appearance of anthropogenic signals could be detected in the measurements at certain hours. This noise is not observed in the example of Fig. 2 but in other 10 min intervals, such as in the example shown in the top of Fig. 3, the amplitude spectrum of the signal includes clear narrow peaks with an amplitude much higher than that of the natural signal. These peaks are fundamentally observed in three bands, (14.7 Hz, 15.00 Hz), (15.00 Hz, 15.48 Hz), and (16.45 Hz, 16.90 Hz), which have an anthropogenic origin. The first two peaks are closely located and appear while a diesel generator that supplies electricity to a mountain hut near the station is running. The third peak contains the operating frequency of the Central European train network. The appearance of this

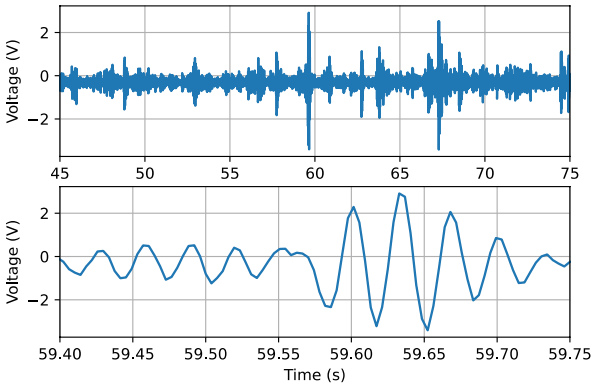

Figure 1. Example of the raw data generated by the station data system. The signal corresponds to voltage in the time domain recorded by the NS-oriented magnetometer at a rate of 256 samples/s. The upper plot corresponds to the time interval after 45 s from the instant 00:49 UTC on March 1, 2015. The lower plot shows in more detail the time interval ranging from 59.4 s to 59.75 s.

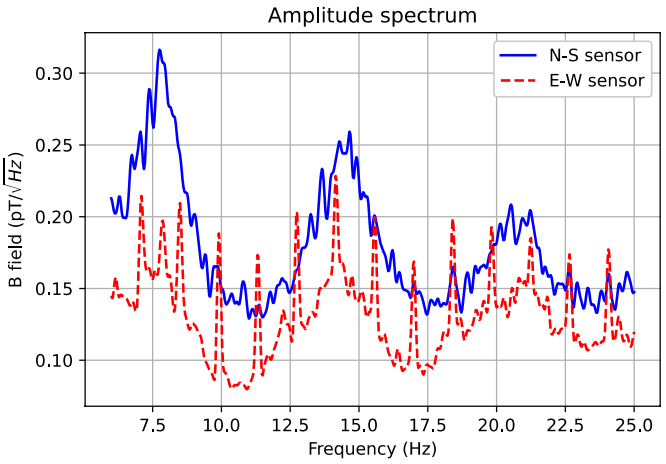

Figure 2. Example of output from the program 'Reading_Fourier.py', which shows the typical amplitude spectrum for the two sensors during a given 10 min interval. The signal corresponds to the 10 min interval associated to the time 00:49:34 UTC of March 1, 2015.

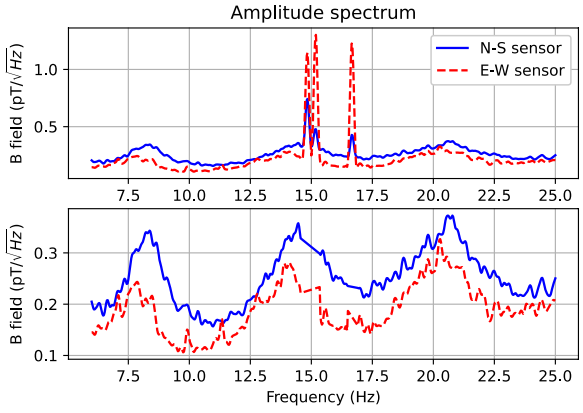

Figure 3. Example of application of anthropogenic noise filtering. (Top) Amplitude spectrum generated by the program 'Reading_Fourier.py' for a 10 min time period in which anthropogenic noise is present. Blue and red lines correspond to the NS and EW sensors, respectively. (Bottom) The spectra in the top panel are processed with the program 'Antropo.py'. The resulting spectra are plotted showing that the anthropogenic noise has been removed. The signal corresponds to the 10 min interval associated to the 17:10 UTC of March 29, 2015.

anthropogenic noise does not follow a fixed hourly pattern, so its elimination requires a detection process to avoid filtering in time intervals for which there is no noise. Furthermore, the presence of noise in the different bands is neither simultaneous nor permanent. Thus, there are some 10 min intervals for which the three peaks can be observed, only one or two peaks are present in another intervals, while other periods do not present any anthropogenic signal at all. Due to the proximity of the two lower bands, the combined frequency band (14.55 Hz, 15.35 Hz) has been considered to deal with the anthropogenic noise originated by the diesel generator of the hut.

If an anthropogenic signal exists in the corresponding noisy band, a straight line is fitted between these limits. That is to say, for a set of n noisy measurements of the magnetic field in this band, $B_N(i)$, at a frequency $f_N(i)$, where $i = 1, \ldots, n$, with $i = 1$ and $i = n$ define de lower and upper limits of the band, respectively, the filtering process consists of substituting the actual measurements, $B_N(i)$, by the filtered values, $B_{FN}(i)$ corresponding to $f_N(i)$ for the straight line connecting the noisy band limits, $f_N(1)$ and $f_N(n)$:

$$B_{FN}(i) = \frac{B_N(n) - B_{Ni}(1)}{f_N(n) - f_N(1)} (f_N(i) - f_N(1)) + B_N(1). \quad (5)$$

The problem raised is determining if the anthropogenic noise is present or not to avoid filtering a signal unaffected by this type of noise. To this aim, a neighbor band at which anthropogenic noise has never been detected is used as reference. At this band, the set of reference measured magnetic field, $B_R(i)$, at the reference frequencies, $f_R(i)$, $i = 1, \ldots, n$, helps in detecting the anthropogenic noise presence. To do so, interpolated values $B_{FR}(i)$ at the reference frequencies for the straight line connecting the reference band limits are calculated through the expression:

$$B_{FR}(i) = \frac{B_R(n) - B_R(1)}{f_R(n) - f_R(1)} (f_R(i) - f_R(1)) + B_R(1). \quad (6)$$

Then, quantities d_N and d_R to compare interpolated and original values are defined for both the anthropogenic noise band and the reference one in the following manner:

$$d_{N,R} = \mathrm{Max}\left\{\frac{B_{N,R}(i) - B_{FN,FR}(i)}{B_{FN,FR}(i)}\right\} + \mathrm{Min}\left\{\frac{B_{N,R}(i) - B_{FN,FR}(i)}{B_{FN,FR}(i)}\right\}, \quad (7)$$

where the N and R subscripts refers to the noise and the reference bands, respectively. The way in which d_N and d_R allow detecting the presence of anthropogenic noise is as follows. In the absence of anthropogenic noise in the considered band and due to the fact that the ELF signal detected at the station includes a natural noise component, measurements will be randomly distributed above and below the straight line and the sum of the positive maximum and negative minimum terms in (7) tend to compensate. This also applies in the reference band, which is defined close to the band with anthropogenic noise to preserve a similar behavior. In contrast, when anthropogenic noise is present, the maximum positive term in (7) will be considerable higher and will not be cancelled by the minimum negative value which will not be substantially affected. For this reason, detecting a value of d_N clearly higher than that of d_R indicates the presence of anthropogenic noise and the noise filter must be applied, and the actual measurements $B_N(i)$ are substituted by the filtered ones, $B_{FN}(i)$ at each frequency measured.

More detailed explanations are given in 'Antropo_Docu.ipynb'. The process is automatic, but the singularity of the noise advises including the possibility of a second and direct correction if the filtered signal is not as good as expected after this first automatic filtering process. To this end, the notebook allows carrying out a visual inspection of the signal to apply a second-stage filtering. To perform this task, the 'Interact' package within the Python module 'ipywidgets'[5] is used.

Fig. 3 shows an example of the anthropogenic noise removal. The top panel shows the output of the program 'Reading_Fourier.py' for a 10 min period time with anthropogenic noise. Three large spikes can be clearly observed, two around 15 Hz and one around 16.5 Hz. The bottom panel shows the same spectra after being processed with the program 'Antropo.py'. The anthropogenic noise has been removed. As a final result of this step, we have a file with the signal amplitude spectrum for each sensor and month, which is free of anthropogenic noise within the calibrated band of the magnetometers. In Fig. 2 and for the EW sensor, a somehow regular spike sequence can be also observed but their amplitude is comparable to that of the measured signal. These peaks are a result of the noisy nature of the ELF signal received and they do not require the application of a filtering process as it is the case of the anthropogenic peaks of much higher amplitude observed at the top of Fig. 3.

The input files for 'Antropo.py' program are the output files of 'Reading_Fourier.py' program, i.e., the amplitude spectrum for each 10 min interval of the month under study and its percentage of saturation.

[5] https://pypi.org/project/ipywidgets/.

The output of 'Antropo.py' is a file containing the spectrum without anthropogenic noise but maintaining the structure of the input file, that is to say, ASCII format and one column. In case this filtering stage is not required, it will suffice changing the name of the output file of program 'Reading_Fourier.py' or, alternatively, to change the name of the input file of program 'Lorenzt_Lmfit.py' of the next step.

5. Signal processing step 3: curve fitting using Lorentzian functions

Once the two previous steps have been applied, the amplitude spectrum of the horizontal magnetic field components, free of anthropogenic noise, is available for each 10 min interval. Unfortunately, this spectrum shows important fluctuations due to its low amplitude with respect to the amplitude of the environmental and instrumental noises, as it can be seen in Figs. 2 and 3. The bandwidth of the magnetometers at the Sierra Nevada station allows the measurement of the first three SRs. Each mode is quantitatively defined through three parameters: amplitude or local maximum value of the spectrum, frequency at which this maximum is observed, and bandwidth of each resonance. These parameters are difficult to define in the noisy spectra of stages 1 and 2. To overcome this difficulty, we can adjust the noisy spectrum to an analytical function that allows defining them with the least possible ambiguity. The resonant behavior expected is in the origin of the Lorentzian fit first proposed in (Sentman, 1987). Since then, a fit combining Lorentzian functions, plus a straight line to allow taking into account a possible slope in the spectrum, is often used by other authors (Mushtak and Williams, 2009; Nickolaenko and Hayakawa, 2002, and references therein). The comparison of Lorentzian fits with Gaussian ones is considered in (Rodríguez-Camacho et al., 2018), which confirms lower errors between the adjusted and experimental data for the case of the Lorentzian fit.

Therefore, the goal of this part of the process is to fit the noisy spectrum provided by step 2 for each 10 min interval using a Lorentzian function for each peak in the frequency interval within the calibrated bandwidth of the magnetometers. The code and notebook associated with this step is 'Lorentz_Lmfit.py' and 'Lorentz_Lmfit_docu.ipynb' respectively.

The Lorentzian fit for adjusting the signal corresponding to three modes used in this stage consists in minimizing the mean square error between the experimental spectrum and the function defined by:

$$L(f) = \sum_{i=1}^{3} \frac{A_i}{1 + \left(\frac{f - f_i}{\sigma_i}\right)^2} + Bf + C, \tag{8}$$

where A_i stands for the amplitude of the i-th mode, resonating at frequency f_i and with a bandwidth of σ_i, while parameters B and C correspond to an added linear adjusting term in the frequency spectrum.

In addition to these 11 parameters, we can define a global mode amplitude, P_i, as the value of the fitting function at the resonance frequencies, f_i. Namely:

$$P_i = L(f_i). \tag{9}$$

These values describe the amplitude of each resonance peak and takes into account the simultaneous effect of the three Lorentzian terms combined with the correcting effect of the linear term.

The procedure therefore consists in minimizing a function χ^2 that accounts for the difference between the data and the model by varying the parameters of the fitting:

$$\chi^2 = \sum_{i=1}^{N} (F(f_i) - L(f_i))^2, \tag{10}$$

where $F(f_i)$ is the value of the measured spectrum for the frequency f_i, L is the fitting function, namely, the sum of Lorentzian functions and a linear term, also for the frequency f_i, and N is the total number of samples of the spectrum to be analyzed. Therefore, the function χ^2 is a scalar function that depends on 11 parameters and the optimization problem results in finding the vector with the values of the 11 parameters for which function χ^2 has a minimum value.

In the developed code, we use the packages 'numpy', 'lmfit.minimize', and 'lmfit.Parameters'. The module 'lmfit'[6] is not installed by default within Python. This module provides a high-level interface for performing non-linear adjustments in Python. It is an extension of 'scipy.optimize', the first optimization module developed for this language. We chose the Levenberg-Marquardt fit algorithm, which provides good results with less computing demands than other methods (Rodríguez-Camacho et al., 2018).

In order to carry out the fit, the method needs the estimation of a seed, or initial value, for each one of the 11 parameters. The seeds are chosen as follows for each 10 min interval:

1. We start from preset values for the central frequencies, width of each of the first three resonances and the linear term. This choice is based on averaged measurements of these quantities. Details can be found in (Rodríguez-Camacho et al. (2018).
2. The initial guesses for peak amplitudes are determined by averaging the month of data and taking the average amplitude at the predefined central frequency. This choice of initial amplitudes is a simplification of the process described in (Rodríguez-Camacho et al., 2018), but providing similar results.

The code executes the following steps:

1. Reading of the output spectra of the 'Anthopo.py' program of block 2, i.e., the calibrated frequency band for each 10 min interval.
2. Definition of the spectrum for the range of fitting frequency band that is contained in the range of calibrated frequency interval. The bandwidth (6.35 Hz − 23.75 Hz) has been chosen because it produces a lower number of parameters outside the acceptable range described in Table 1 in (Rodríguez-Camacho et al., 2018) when compared to other possible frequency regions defined inside the calibrated bandwidth of the station magnetometers.
3. Call to 'lmfit' function and determination of 11 fitting parameters for each 10 min interval of the month.
4. Evaluation of the three global mode amplitude.
5. Output of the results obtained.

The outputs of program 'Lorentz_Lmfit.py' for each 10 min interval of the month and sensor are:

1. A list of 14 fitting parameters.
2. A list of the 11 errors associated with the Lorentzian fitting.
3. A list of the mean square errors (χ^2 function) of the fitting.

The error parameters are outputs of program Lmfit. The study of the correlation of these errors with variations of the parameters associated with the SRs is still pending.

Fig. 4 shows the results of fitting an amplitude spectrum using three Lorentzian fits. The signal in red line shows the anthropogenic noise filtered spectrum obtained after applying the program 'Anthropogenic.py' at stage two. This signal is the input for the program 'Lorentz_Lmfit.py' of stage three, which provides the fit shown with blue-dashed line.

7. Step 4: Data packaging

[6] https://lmfit.github.io/lmfit-py/model.html.

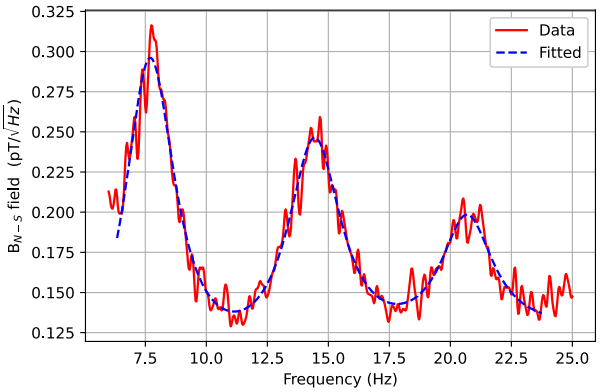

Figure 4. Example of fitted amplitude spectrum, using three Lorentzian curves plus a linear term. The red line shows the output of stage two, program 'Anthropo.py', and the blue line shows the fitted spectrum resulting from applying program 'Lorentz_Lmfit.py' of stage 3. The signal corresponds to the 10 min interval associate to time 00:49 UTC of March 21, 2015.

In the final part of the process, a file is generated in numpy format (npz[7]) with all the information for the month. This is the file that could be shared between ELF stations, containing all the information necessary for the study of the SRs for each month. Information about the beginning in universal time of each 10 min interval of the month is also incorporated.

For each 10 min interval, the following readings from the previous steps are done:

- Percentage of saturation.
- Amplitude spectrum.
- Parameters of the Lorentzian fitting.
- Value of the χ^2 function.
- Value of the error in the fitting for each parameter.

In addition, six new parameters associated with the Lorentzian fitting are obtained from the fitted spectra. They correspond to the maximum amplitude values for each mode and corresponding frequencies for each resonance. The maximum values of this fitted function, the local maximum amplitudes, and their corresponding frequencies, local maximum frequencies, also could define the SR amplitudes and frequencies, respectively. This calculation has been included in this part in order to make it optional with respect to the Lorentzian fitting made in the previous step.

The output file has npz format for writing/reading data. The file is in binary format, but it maintains the numpy array structure so that its reading and the recovery of the structure of each array is immediate. This makes that the size of each file is considerably reduced and, in addition, that no specific program is required to relate data with each output array. As a drawback, the reading depends on Python language and the Numpy package.

There is a file for each month of measurements and for each sensor. The filename makes reference to the year, month and sensor used. The name for December 2014 and sensor NS is SN_1412_0.npz, for instance. Each file has a size of 43.3 MB, therefore, all the months can be stored in the same directory in order to facilitate the reading for a later analysis.

The list of saved arrays is:

- 'sat': saturation percentage,
- 'fre': list of frequencies in the calibrated band,
- 'freajus': list of frequencies in the adjustment band,
- 'rs': spectra obtained from the measurements,
- rs_LO: spectra obtained from the fitting process,
- PARLO: parameters obtained from the adjustment (peaks, amplitudes, frequencies, widths, slope, and ordinate at the origin of the linear term),
- parLOC: χ^2 function,
- parLOE: error in the parameters,
- pr: maximum values for each resonance,
- time: date and time of the beginning of each 10 min interval of the month.

[7] https://numpy.org/doc/stable/reference/generated/numpy.savez.html.

6. Conclusions

Measuring the natural ELF electromagnetic field in the atmosphere is a complicated task. The ELF stations around the Earth are scarce and the techniques to extract the SRs from them are also complex. Sharing these measurements, procedures and codes may help to spread the findings of a particular research group to the rest of scientific community and to stablish a global network to study SRs and their applications.

In this work, we present the algorithms that allow characterizing each SR mode starting from the raw data measured by the ELF station at Sierra Nevada, Spain, recorded between March 2013 and February 2017. This software is written in Python language (version 3.8), because it allows easy access to the different stages of the procedure and, therefore, it facilitates obtaining intermediate results as well as including modifications, adaptations and even improvements of the method. A detailed and comprehensive description of the code is given through several Jupyter notebooks that are provided as supplementary material. In addition, the time raw data series measured by the Sierra Nevada station during a four year period are provided in this paper as well as the outputs of the different parts of the procedure.

The procedure starts from the time series raw data measured by the ELF station and ends with the obtention of three parameters to quantitatively describe each resonance: peak amplitude, central frequency and resonance width. The code is divided into four stages, which can be run independently in order to optimize the results of each stage without affecting the other. Each block can be run on a personal computer requiring a CPU time of a few minutes.

The programs and basic tasks carried out in each stage are the following:

1. First, 'Reading_Fourier' reads the raw data obtained by the stations and generates an initial spectrum for each 10 min period of each month.
2. The program 'Anthropo.py' is applied to filter the human noise present in some of the spectra obtained at the first stage.
3. In the third stage, a Lorentianz fit is obtained through the program 'Lorentz_Lmfit'.
4. Finally, the last program, 'Packaging.py', generates an output file in 'npz' format with all the results.

Authorship contribution statement

All authors conceived the idea, provided software supervision and participated in essential manuscript reviews and editing. The code has been developed mainly by A. Salinas and J. Rodríguez- Camacho. The original manuscript has been written by A. Salinas and J. Portí. S. Toledo-Redondo, J. Fornieles-Callejón and M.C. Carrión provided project resources and contributed to data curation and analysis.

Code availability section

Name of the code/library: Sierra_Nevada_SR
Contact: asalinas@ugr.es; +34 958242312.
Hardware requirements: Laptop or PC running Linux, Mac or

Windows operating systems.

Program language: Python (version 3.8).

Software required: Python and Jupyter notebooks.

Program size: 27 kB (total for the four programs).

The source codes, Jupyter notebooks and one day of data (March 1, 2015) can be found at: https://github.com/asalinas62/Sierra_Nevada_SR.

The Sierra Nevada ELF raw data station from the period March 2013 to February 2017 can be downloaded from the Zenodo repositories [Salinas et al. a, b, c, d].

The final proceeded data from the previous raw data using the program 'Packaging.py' in npz format can be download from de University of Granada repository http://hdl.handle.net/10481/71563.

The raw data, intermediate proceeded and final data of the December 2014 and March 2015 can be download from the University of Granada repositories http://hdl.handle.net/10481/73978 and http://hdl.handle.net/10481/73977 respectively.

Declaration of competing interest

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

Acknowledgments

This work has been supported by the investigation research project FIS2017-90102-R, of the Ministry of Economy and Competitiveness (MINECO) co-financed by the Fund European Regional Development (FEDER), and the Ministry of Education, Science and Sport of Spain through the FPU grants for PhD studentship (reference: FPU15/04291. We acknowledge support of the Ministry of Science and Innovation of Spain (grant PID2020-112805GA-I00). We are grateful to Parque Nacional Sierra Nevada for providing support to the project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cageo.2022.105148.

References

Balser, M., Wagner, C.A., 1962. On frequency variations of the Earth-ionosphere cavity modes. J. Geophys. Res. 67 (10), 4081–4083. https://doi.org/10.1029/JZ067i010p04081.

Besser, B.P., 2007. Synopsis of the historical development of Schumann resonances. Radio Sci. 42 (2), RS2S02 https://doi.org/10.1029/2006RS003495.

Bozóki, T., Sátori, G., Williams, E., Mironova, I., Steinbach, P., Bland, E.C., Koloskov, A., Yampolski, Y.M., Budanov, O.V., Neska, M., Sinha, A.K., Rawat, R., Sato, M., Beggan, C.D., Toledo-Redondo, S., Liu, Y., Boldi, R., 2021. Solar cycle-modulated deformation of the earth–ionosphere cavity. Front. Earth Sci. 9, 735. https://doi.org/10.3389/feart.2021.689127.

Fornieles-Callejón, J., Salinas, A., Toledo-Redondo, S., Portí, J., Méndez, A., Navarro, E. A., Morente-Molinera, J.A., Soto-Aranaz, C., Ortega-Cayuela, J.S., 2015. Extremely low frequency band station for natural electromagnetic noise measurement. Radio Sci. 50, 191–201. https://doi.org/10.1002/2014RS005567.

Füllekrug, M., Fraser-Smith, A.C., 1997. Global lightning and climate variability inferred from ELF magnetic field variations. Geophys. Res. Lett. 24 (19), 2411–2414. https://doi.org/10.1029/97GL02358.

Granger, B.E., Perez, F., 2021. Jupyter: thinking and storytelling with code and data. Comput. Sci. Eng. 23, 7–14. https://doi.org/10.1109/MCSE.2021.3059263, 02.

Koloskov, A.V., Nickolaenko, A.P., Yampolsky, Y.M., Hall, C., Budanov, O.V., 2020. Variations of global thunderstorm activity derived from the long-term Schumann resonance monitoring in the Antarctic and in the Arctic. J. Atmos. Sol. Terr. Phys. 201, 105231 https://doi.org/10.1016/j.jastp.2020.105231.

Mushtak, V.C., Williams, E.R., 2009. An improved Lorentzian technique for evaluating resonance characteristics of the Earth-ionosphere cavity. Atmos. Res. 91 (2), 188–193. https://doi.org/10.1016/j.atmosres.2008.08.013.

Nickolaenko, A.P., Hayakawa, M., 2002. Resonances in the Earth-Ionosphere Cavity. Kluwer Acad., Dordrecht, Netherlands.

Nickolaenko, A.P., Hayakawa, M., 2014. Schumann Resonance for Tyros. Springer, Japan.

Nickolaenko, A.P., V, K.A., Hayakawa, M., M, Y.Y., V, B.O., E, K.V., 2015. 11-year solar cycle in Shumann resonance data as observed in Antartica. Sun and Geosphere 10 (1), 39–49.

Prácser, E., Bozóki, T., Sátori, G., Williams, E., Guha, A., Yu, H., 2019. Reconstruction of global lightning activity based on schumann resonance measurements: model description and synthetic tests. Radio Sci. 54 (3), 254–267. https://doi.org/10.1029/2018RS006772.

Price, C., 2016. ELF electromagnetic waves from lightning: the schumann resonances. Atmosphere 7 (9), 116. https://doi.org/10.3390/atmos7090116.

Price, C., Melnikov, A., 2004. Diurnal, seasonal and inter-annual variations in the Schumann resonance parameters. J. Atmos. Sol. Terr. Phys. 66 (13–14), 1179–1185. https://doi.org/10.1016/j.jastp.2004.05.004.

Rodríguez-Camacho, J., Fornieles, J., Carrión, M.C., Portí, J.A., Toledo-Redondo, S., Salinas, A., 2018. On the need of a unified methodology for processing schumann resonance measurements. J. Geophys. Res. Atmos. 123 (23), 13,277–13,290. https://doi.org/10.1029/2018JD029462.

Rodríguez-Camacho, J., Salinas, A., Carrión, M.C., Portí, J., Fornieles-Callejón, J., Toledo-Redondo, S., 2022. Four year study of the schumann resonance regular variations using the Sierra Nevada station ground-based magnetometers. J. Geophys. Res. Atmos. 127 (6), e2021JD036051 https://doi.org/10.1029/2021JD036051.

Roldugin, V.C., Maltsev, Y.P., Vasiljev, A.N., Schokotov, A.Y., Belyajev, G.G., 2004. Diurnal variations of Schumann resonance frequency in NS and EW magnetic components. J. Geophys. Res.: Space Phys. 109 (A8), A08304 https://doi.org/10.1029/2004JA010487.

Rycroft, M.J., Harrison, R.G., Nicoll, K.A., Mareev, E.A., 2008. An overview of Earth's global electric circuit and atmospheric conductivity. Space Sci. Rev. 137, 83–105.

Salinas, A., Rodríguez-Camacho, J., Portí, J., Carrión, M.C., Fornieles-Callejón, J., Toledo-Redondo, S., 2022a. Four-year Measurements from Sierra Nevada ELF Station. Zenodo. https://doi.org/10.5281/zenodo.6348691. *Year 2014* [Data set].

Salinas, A., Rodríguez-Camacho, J., Portí, J., Carrión, M.C., Fornieles-Callejón, J., Toledo-Redondo, S., 2022b. Four-year Measurements from Sierra Nevada ELF Station. Zenodo. https://doi.org/10.5281/zenodo.6348773. *Year 2015* [Data set].

Salinas, A., Rodríguez-Camacho, J., Portí, J., Carrión, M.C., Fornieles-Callejón, J., Toledo-Redondo, S., 2022c. Four-year Measurements from Sierra Nevada ELF Station. Zenodo. https://doi.org/10.5281/zenodo.6348838. *Year 2016* [Data set].

Salinas, A., Rodríguez-Camacho, J., Portí, J., Carrión, M.C., Fornieles-Callejón, J., Toledo-Redondo, S., 2022d. Four-year Measurements from Sierra Nevada ELF Station. Years 2013, p. 2017. https://doi.org/10.5281/zenodo.6348930 [Data set]. Zenodo.

Sátori, G., Rycroft, M., Bencze, P., Märcz, F., Bór, J., Barta, V., Nagy, T., Kovács, K., 2013. An overview of thunderstorm-related research on the atmospheric electric field, schumann resonances, sprites, and the ionosphere at sopron, Hungary. Surv. Geophys. 34 (3), 255–292. https://doi.org/10.1007/s10712-013-9222-6.

Schlegel, K., Füllekrug, M., 2000. Diurnal harmonics in schumann resonance parameters observed on both hemispheres. Geophys. Res. Lett. 27 (17), 2805–2808. https://doi.org/10.1029/2000GL003774.

Schumann, W.O., 1952. Über die stralungslosen Eigenschwingungen einer leitenden Kugel die von einer Luftschicht und einer Ionospärenhle umgeben ist. Z. Naturforsch. 7a, 149–154. https://doi.org/10.1515/zna-1952-0202.

Sentman, D.D., 1987. Magnetic elliptical polarization of Schumann resonances. Radio Sci. 22 (4), 595–606. https://doi.org/10.1029/RS022i004p00595.

Sentman, D.D., Fraser, B.J., 1991. Simultaneous observations of Schumann resonances in California and Australia: evidence for intensity modulation by the local height of the D region. J. Geophys. Res.: Space Phys. 96 (A9), 15973–15984. https://doi.org/10.1029/91JA01085.

Shvets, A., Hayakawa, M., 2011. Global lightning activity on the basis of inversions of natural ELF electromagnetic data observed at multiple stations around the world. Surv. Geophys. 32 (6), 705. https://doi.org/10.1007/s10712-011-9135-1.

Simões, F., Pfaff, R., Berthelier, J.-J., Klenzing, J., 2012. A review of low frequency electromagnetic wave phenomena related to tropospheric-ionospheric coupling mechanisms. Space Sci. Rev. 168 (1–4), 551–593. https://doi.org/10.1007/s11214-011-9854-0.

Toledo-Redondo, S., Salinas, A., Portí, J., Morente, J.A., Fornieles, J., Méndez, A., Galindo-Zaldívar, J., Pedrera, A., Ruiz-Constán, A., Anahnah, F., 2010. Study of Schumann resonances based on magnetotelluric records from the western Mediterranean and Antarctica. J. Geophys. Res. Atmos. 115 (D22114) https://doi.org/10.1029/2010JD014316.

Williams, E., Bozóki, T., Sátori, G., Price, C., Steinbach, P., Guha, A., Liu, Y., Beggan, C. D., Neska, M., Boldi, R., Atkinson, M., 2021. Evolution of global lightning in the transition from cold to warm phase preceding two super El Niño events. J. Geophys. Res. Atmos. 126 (3), e2020JD033526 https://doi.org/10.1029/2020JD033526.

Williams, E.R., 1992. The Schumann Resonance: a global tropical thermometer. Science 256 (5060), 1184–1187. https://doi.org/10.1126/science.256.5060.1184.

Williams, E.R., Mushtak, V., Guha, A., Boldi, R., Bór, J., Nagy, T.G., Sátori, G., 2014. Inversion of multi-station schumann resonance background records for global lightning activity in absolute units. AGU Fall Meet. Abstr. 1, AE24A-08.