



**HAL**  
open science

# Detection and characterization of lightning-based sources using continuous wavelet transform: application to audio-magnetotellurics

H Larnier, P. Sailhac, A. Chambodut

► **To cite this version:**

H Larnier, P. Sailhac, A. Chambodut. Detection and characterization of lightning-based sources using continuous wavelet transform: application to audio-magnetotellurics. *Geophysical Journal International*, 2018, 212 (1), pp.103 - 118. 10.1093/gji/ggx418 . hal-01894628

**HAL Id: hal-01894628**

**<https://hal.science/hal-01894628>**

Submitted on 3 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Detection and characterization of lightning-based sources using continuous wavelet transform: application to audio-magnetotellurics

H. Larnier,<sup>1</sup> P. SAILHAC<sup>1,2</sup> and A. Chambodut<sup>1</sup>

<sup>1</sup>*Institut de Physique du Globe de Strasbourg, CNRS UMR-7516, Université de Strasbourg, F-67084 Strasbourg, France. E-mail: [hlarnier@unistra.fr](mailto:hlarnier@unistra.fr)*

<sup>2</sup>*Laboratoire GEOPS, CNRS UMR-8148, Université Paris Sud, F-91405 Orsay, France*

Accepted 2017 September 27. Received 2017 September 25; in original form 2016 October 10

## SUMMARY

Atmospheric electromagnetic waves created by global lightning activity contain information about electrical processes of the inner and the outer Earth. Large signal-to-noise ratio events are particularly interesting because they convey information about electromagnetic properties along their path. We introduce a new methodology to automatically detect and characterize lightning-based waves using a time–frequency decomposition obtained through the application of continuous wavelet transform. We focus specifically on three types of sources, namely, atmospheric, slow tails and whistlers, that cover the frequency range 10 Hz to 10 kHz. Each wave has distinguishable characteristics in the time–frequency domain due to source shape and dispersion processes. Our methodology allows automatic detection of each type of event in the time–frequency decomposition thanks to their specific signature. Horizontal polarization attributes are also recovered in the time–frequency domain. This procedure is first applied to synthetic extremely low frequency time-series with different signal-to-noise ratios to test for robustness. We then apply it on real data: three stations of audio-magnetotelluric data acquired in Guadeloupe, overseas French territories. Most of analysed atmospheric and slow tails display linear polarization, whereas analysed whistlers are elliptically polarized. The diversity of lightning activity is finally analysed in an audio-magnetotelluric data processing framework, as used in subsurface prospecting, through estimation of the impedance response functions. We show that audio-magnetotelluric processing results depend mainly on the frequency content of electromagnetic waves observed in processed time-series, with an emphasis on the difference between morning and afternoon acquisition. Our new methodology based on the time–frequency signature of lightning-induced electromagnetic waves allows automatic detection and characterization of events in audio-magnetotelluric time-series, providing the means to assess quality of response functions obtained through processing.

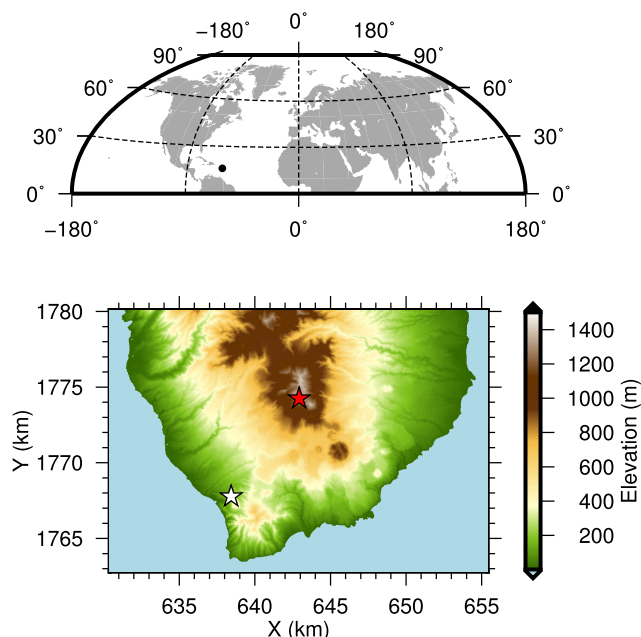
**Key words:** Geomagnetic induction; Magnetotellurics; Time-series analysis; Wavelet transform.

## 1 INTRODUCTION

Natural variation of the magnetic field in the frequency range 1 Hz to 10 kHz mainly originates from global lightning activity, with about 45 lightning strikes per second (Christian *et al.* 2003). Each single lightning strike gives rise to various phenomena. Among those phenomena, emanated electromagnetic (EM) waves propagate through various paths. Most EM waves propagate in the waveguide formed by two conductive layers: the Earth's surface and the ionosphere (Laby *et al.* 1940). Others, as whistlers, follow magnetic field lines from one hemisphere to the other through the magnetosphere. Analysis of such waves' properties supplies information about electric and atmospheric processes of the Earth. Indeed, atmospheric characterization provides information about the ionospheric conductivity (Hughes & Pappert 1975) and the distance to the lightning source (Heydt 1982). Slow tails properties provide information about the

current moment of the lightning strike (Cummer & Inan 2000) and the distance from the source (Mackay & Fraser-Smith 2010). Measurements of the Schumann resonances (Schumann & Koenig 1954) provide insights about the height of the ionosphere, its electron density (Chand *et al.* 2009), or the global ground temperature change (Williams 1992).

EM waves emitted from a lightning strike act as induction sources by creating electric currents in the ground. If the source is sufficiently distant, the emitted waves can be considered as quasi-uniform over the area of measurements. The framework of audio-magnetotellurics (AMT) uses this assumption to give information about the electrical conductivity structure of the subsurface (Strangway *et al.* 1973). The state-of-the-art processing of AMT data typically employs windowed Fourier transform techniques and assumes that source fields are quasi-stationary. Most of the methodological developments have been focused on the introduction of



**Figure 1.** Location of magnetotelluric stations used in Guadeloupe, French Overseas Departments and Territories. Top: global map. Bottom: local map in UTM 20N projection. Red star: location of Guenon and Cratère Sud AMT station. Both stations are too close to be distinguished on this map. White star: Youketi AMT station.

statistical schemes (Chave 2014) or the analysis of stations arrays (Egbert 2002), to exploit and analyse the variability of source fields, but only a few studies actually investigate EM waves and their properties for the processing of magnetotelluric (MT) time-series. Among those studies, a detailed analysis of the seasonal variation of magnetic field amplitude was conducted by Garcia & Jones (2002). They demonstrated that the optimal season to acquire AMT data is summer in the northern hemisphere at mid-latitude. In the framework of AMT processing, Zhang & Paulson (1997) proposed a methodology using a moving threshold in time–frequency domain to detect atmospheric waves before processing. We previously extended Zhang & Paulson method to consider the choice of the wavelet depending on the frequency domain to be processed (Larnier *et al.* 2016).

In this paper, we propose a new methodology: Automatic Detection of Electromagnetic Waves (ADEM), to automatically detect EM waves in the frequency range 10 Hz–10 kHz. ADEM is based on the time–frequency properties of sought waves using continuous wavelet transform (CWT). We first show in Section 2, AMT data acquired in November 2012 in Guadeloupe, French Overseas Territories. They exhibit interesting features of typical EM waves. The ADEM method, including details on the computation of wavelet coefficients and polarization attributes, is given in Section 3. In Section 4, we show the application of the ADEM to real data and discuss about polarization attributes and frequency content of atmospheric waves with consequences on the acquisition and processing of AMT data.

## 2 LIGHTNING-BASED WAVES

### 2.1 Sites locations and data

In this paper, we consider EM time-series from three stations (Fig. 1) acquired during an MT campaign on La Soufriere lava dome

(Guadeloupe, France) in November 2012 (Sailhac *et al.* 2016). All MT data were acquired using the Metronix Geophysics ADU07 data logger as a recording system.

The first two stations, named Guenon and Cratère Sud respectively, were located on the lava dome. Magnetic fields were recorded using MFS07e induction coils (using chopper off). Guenon measurements employed 41 and 31 m electric dipoles for the NS direction and EW direction respectively. 50 m dipoles were employed in measurements at Cratère Sud.

The third station, named Youketi, was used as a remote station for AMT processing. It is located 8 km off the lava dome and 600 m off the Soufriere Volcano Observatory. This station was composed by 50 m electric dipole for NS and EW measurements and MFS06 induction coils (using chopper off) to measure the magnetic field.

Sample time-series analysed in this paper were sampled at 16384 Hz. Guenon time-series were acquired for 5 min starting at 19 hr 45 min in Coordinated Universal Time (UTC) (14 hr 53 min 9 s solar time), 2012 November 16, in the afternoon. Cratère Sud time-series were acquired for 5 min starting at 19 hr 45 min UTC (14 hr 53 min 34 s solar time), 2012 November 14, and again for 5 min starting at 15 hr 45 min UTC (10 hr 53 min 22 s solar time), 2012 November 15. No high frequency night data were recorded during this acquisition.

Many types of EM waves coexist in the frequency band of 1 Hz up to 10 kHz for the time-series analysed here. In this paper, we focus on three types of EM waves, atmospheric, slow tails and whistlers, all of which are well represented in the analysed time-series, thus providing an excellent basis for testing ADEM.

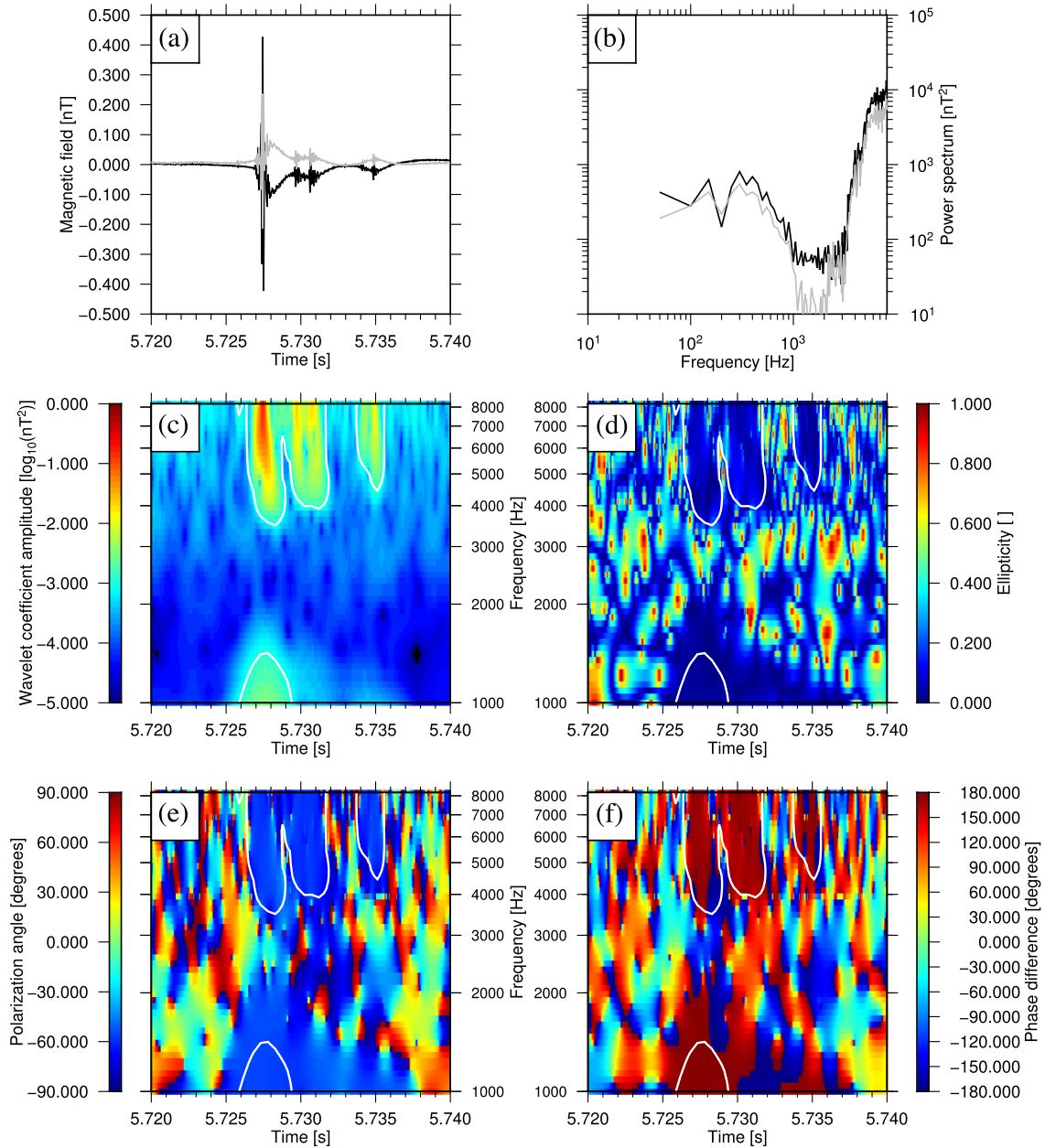
### 2.2 Atmospheric

Atmospherics appear on EM time-series as damped oscillations (Fig. 2a). Their shapes vary, depending on ionospheric conditions (e.g. night or day) and distance from the lightning strike (Rakov & Uman 2003). According to the waveguide mode theory (Wait 1962; Cummer 2000), atmospheric waves undergo dispersion from the source to the measurement point where highest frequencies travel faster than lowest ones. Atmospheric waves have frequency content ranging from 2 kHz up to more than 30 kHz (Fig. 2b). In Fig. 2, we present the CWT analysis of a short portion of Guenon time-series. The largest atmospheric at 5.727 s exhibits broad frequency content, from 3–4 kHz to more than 10 kHz (Fig. 2c). In addition, other atmospheric waves appear later (around 5.730 s). They also have wide frequency content but with lower amplitude and lower signal-to-noise ratio. An interesting feature is the dispersion effect on wavelet coefficients. Indeed, large coefficients at high frequencies around 10 kHz appear to be located a few milliseconds before large coefficients at lower frequencies around 4 kHz.

Polarization attributes (Figs 2d–f) are constant over the area of high signal-to-noise ratio. Atmospheric waves in this typical time sequence display quasi-null ellipticity, a phase difference close to 180 degrees and a polarization angle close to 60 degrees. If such polarization attributes are observed with accuracy, we will consider atmospheric waves to have linear polarization.

### 2.3 Slow tails

Slow tails are also emitted by a lightning strike and arrive slightly later than atmospheric waves due to lower speed propagation. These waves only propagate in quasi-Transverse ElectroMagnetic (TEM) field mode. Characteristics of both atmospheric and slow tails waves



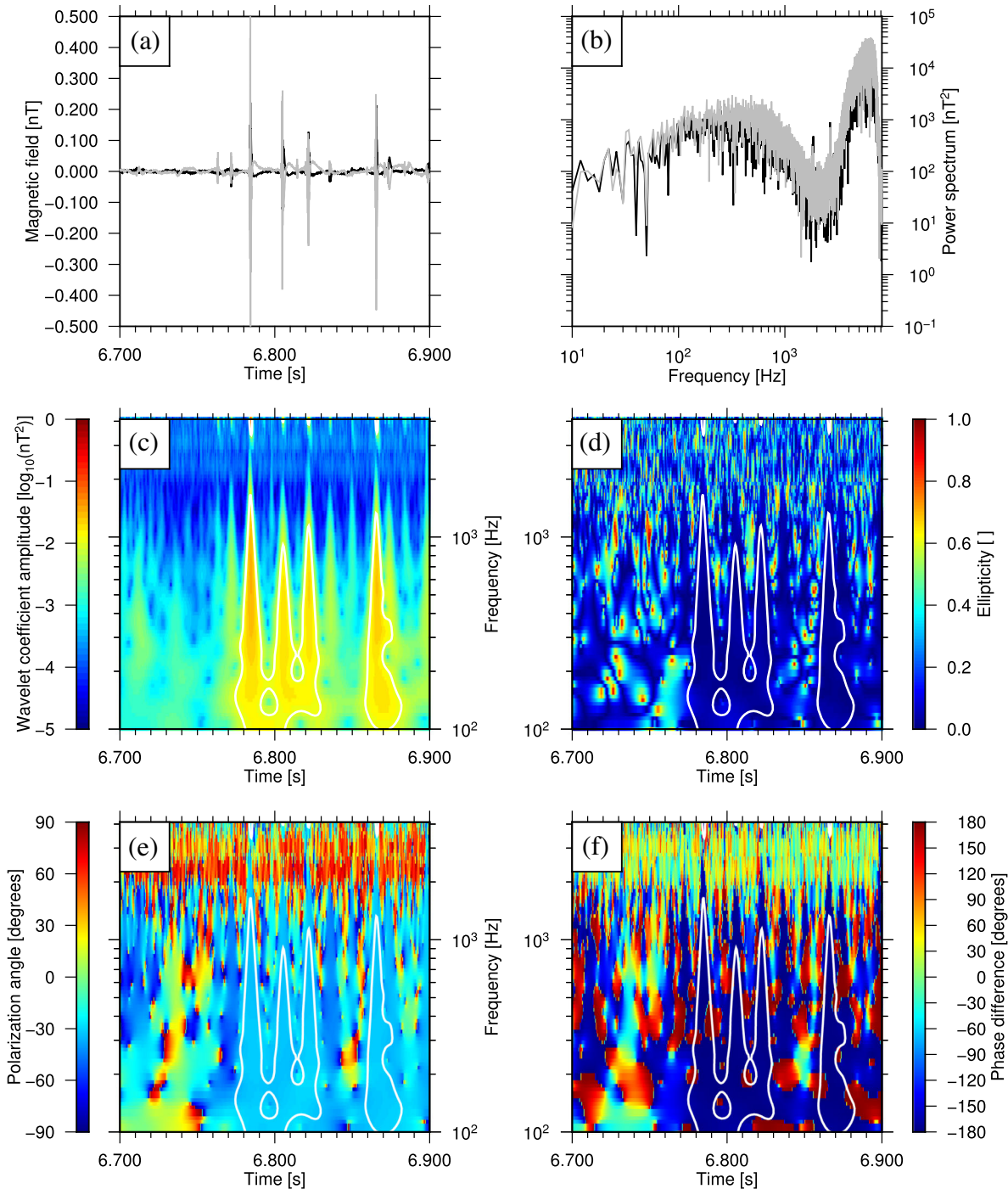
**Figure 2.** Atmospherics observed at Guenon station, on 2012 November 16. Time-series start at 14 hr 53 min 9 s solar time. Displayed time is the delay from this date. (a) Time-series. Black line: north–south magnetic field  $H_x$ . Grey line: east–west magnetic field  $H_y$ . (b) Power spectrum. Black line: north–south magnetic field. Grey line: east–west magnetic field. The blue area indicates the frequency domain covered by the wavelet analysis. (c) Spectrogram of wavelet coefficients. (d) Ellipticity coefficients. (e) Polarization angle coefficients. (f) Phase difference coefficients. (Find definitions of these parameters in Section 3.). A 0.001 nT<sup>2</sup> contour is plotted on subplots (c)–(f) to illustrate the polarization coefficients of atmospherics.

have been used in several studies to determine the distance from the source lightning (Wait 1960; Mackay & Fraser-Smith 2010) or its current moment (Cummer & Inan 2000). Their frequency content ranges between 30 Hz and 3 kHz (Figs 3a and b). Fig. 3(c) shows wavelet analysis of a short portion of Guenon time-series showing several large slow tails. Similar to the atmospherics, these slow tails have a broad frequency content ( $\sim 50$  Hz to  $\sim 1$  kHz). Amplitude variability of wavelet coefficients also affects the frequency content. By comparison, the slow tail located at 6.835 s has narrower frequency content, from less than 2 kHz down to 300 Hz, whereas the slow tail at 6.785 s has high amplitude wavelet coefficients from 2.5 kHz down to 100 Hz.

Polarization attributes (Figs 3d–f) are constant over the region of large amplitude coefficients indicating constant polarization attributes. The largest slow tail has ellipticity close to 0, a phase difference of  $\pm 180$  degrees and a polarization of about  $-30$  degrees. Like atmospherics, we consider these attributes as signature of linear polarization.

## 2.4 Whistlers

Whistlers are EM waves from lightning strikes that propagate along field lines instead of through the waveguide. They follow field lines between hemispheres (Rakov & Uman 2003) and a subgroup,

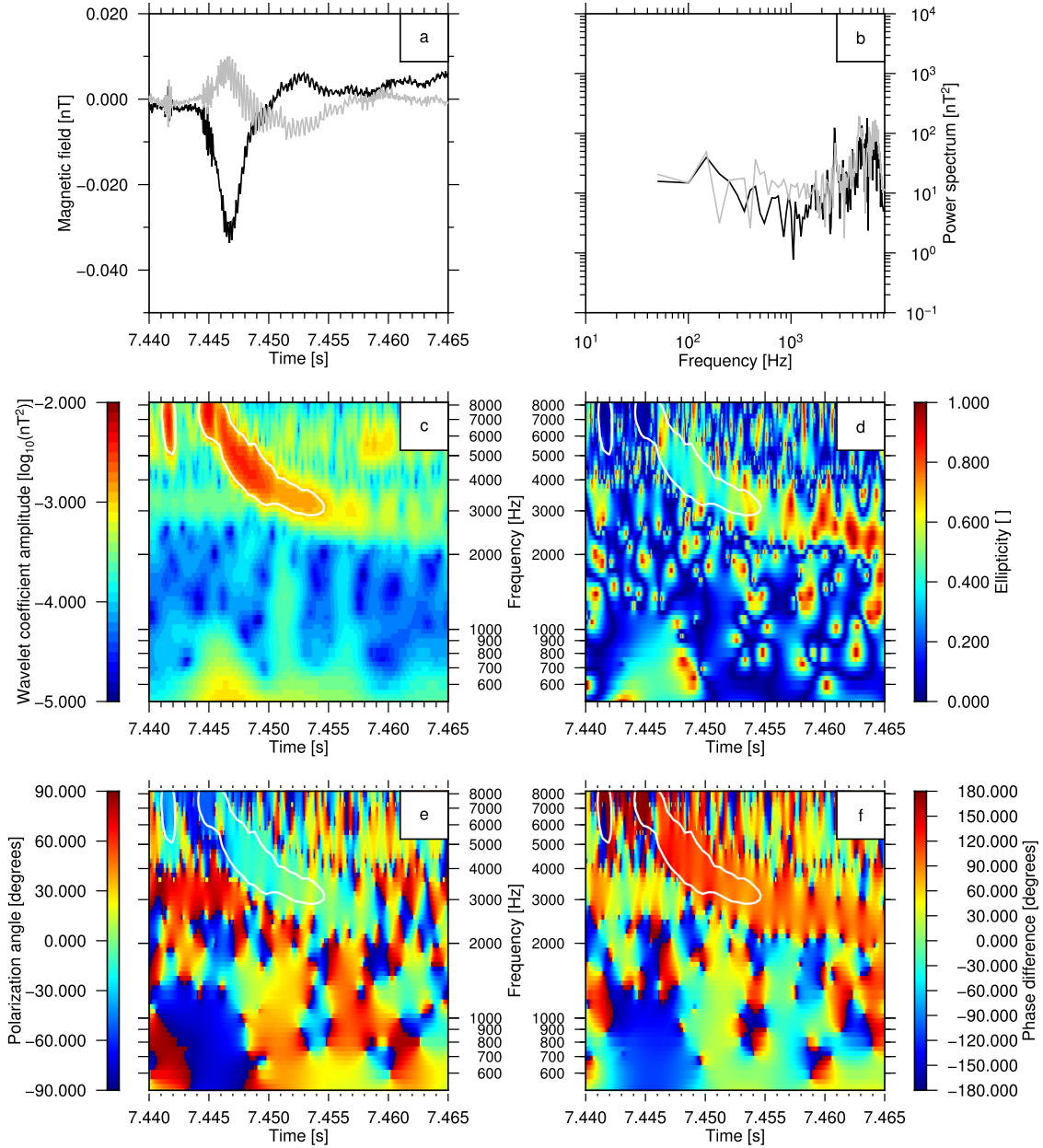


**Figure 3.** Slow tails observed at Guenon station on 2012 November 16. Time-series start at 14 hr 53 min 9 s solar time. Displayed time is the delay from this date. (a) Time-series. Black line: north–south magnetic field  $H_x$ . Grey line: east–west magnetic field  $H_y$ . (b) Power spectrum. Black line: north–south magnetic field. Grey line: east–west magnetic field. The blue area indicates the frequency domain covered by the wavelet analysis. (c) Spectrogram of wavelet coefficients. (d) Ellipticity coefficients. (e) Polarization angle coefficients. (f) Phase difference coefficients. A  $0.01 \text{ nT}^2$  contour is plotted on subplots (c)–(f) to illustrate the polarization coefficients of slow tails.

‘ducted’ whistlers are observable at the Earth’s surface, dispersed during their propagation through ionosphere and magnetosphere. Thus high frequencies arrive significantly sooner than low frequencies. A whistler may contain frequencies from 100 Hz up to 10 kHz (Rakov & Uman 2003). Our experience is that the typical frequency content of whistlers has a lower limit of 1 kHz. Frequency content below 1 kHz is mostly observed on satellite measurements

(Huang *et al.* 2004). Figs 4(a) and (b) show a whistler in time and frequency domain.

Fig. 4(c) shows the wavelet analysis of a whistler observed in Guenon data. The most distinctive feature in the wavelet spectrogram is the dispersion effect. This dispersion affects the wavelet coefficient such that high amplitude coefficients at low frequencies (around 3 kHz) are located 10 ms after the high frequencies.



**Figure 4.** Whistlers observed at Guenon station on 2012 November 16. Time-series start at 14 hr 53 min 9 s solar time. Displayed time is the delay from this date. (a) Time-series. Black line: north–south magnetic field  $H_x$ . Grey line: east–west magnetic field  $H_y$ . (b) Power spectrum. Black line: north–south magnetic field. Grey line: east–west magnetic field. The blue area indicates the frequency domain covered by the wavelet analysis. (c) Spectrogram of wavelet coefficients. (d) Ellipticity coefficients. (e) Polarization angle coefficients. (f) Phase difference coefficients. A  $0.001 \text{ nT}^2$  contour is plotted on subplots (c)–(f) to illustrate the polarization coefficients of the whistler wave.

Polarization attributes (Figs 4d–f) indicate elliptical polarization with ellipticity of about 0.5, phase difference around 120 degrees positive, and polarization angle of  $-40$  degrees.

### 3 METHODS

#### 3.1 Wavelet analysis

We recall that the CWT is a mathematical tool that uses a special class of functions called wavelets to decompose a signal  $s$  into time–frequency representation. The basis functions are derived from the mother wavelet that is dilated by a dimensionless factor, the scale  $a$  (corresponding to the inverse of frequency) and translated by a

factor  $b$  (corresponding to time). The signal  $s$  is then decomposed into coefficients by

$$\mathcal{W}_\psi(a, b)[s] = \int_{-\infty}^{+\infty} dt \frac{1}{a} \psi_{a,b}^*(t) s(t), \quad (1)$$

where  $*$  represents the complex conjugate and  $\psi_{a,b}(t)$  is the dilated and translated function given by eq. (2),

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right). \quad (2)$$

For discrete time-series, the function  $\mathcal{W}_\psi(a, b)[s]$  is sampled on discrete values for  $a$  and  $b$ , constituting a matrix of wavelet coefficients.

First introduced in seismic processing (Goupillaud *et al.* 1984), the CWT has already been widely used in acoustics and climate studies to analyse non-stationary processes and proven to be an efficient tool to study magnetotelluric data (Zhang & Paulson 1997; Garcia & Jones 2008; Escalas *et al.* 2013; Larnier *et al.* 2016).

A large number of mother wavelets is available (Torrence & Compo 1998; Holschneider 1995). Following Zhang & Paulson (1997), we use the Morlet wavelet for its properties of good accuracy both in time and frequency. It is defined in time domain by

$$\psi_m(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}, \quad (3)$$

and in frequency domain by

$$\widehat{\psi}_m(a\omega) = \pi^{-1/4} H(\omega) e^{-(a\omega - \omega_0)^2/2}, \quad (4)$$

where  $H$  is the Heaviside function and  $\omega_0$  the wavelet parameter.  $\omega_0$  can be modified to adjust the time–frequency resolution of the analysing wavelet. In this paper,  $\omega_0$  is set to the common value of 6, close to the smallest value for the so-called admissibility criteria of the Morlet wavelet (Larnier *et al.* 2016).

### 3.2 Detection and characterization of EM waves

Lightning EM waves display different time–frequency patterns according to their nature (Figs 2–4). Some have wavelet coefficients with large amplitude on a wide frequency band whereas some display narrow frequency content. The detection of these waves in the time–frequency domain may be based on the characteristics shown by the CWT. We thus set two criteria:

- (i) A minimum amplitude threshold in the time–frequency domain (signal-to-noise ratio).
- (ii) A specific pattern in time–frequency (e.g. dispersion).

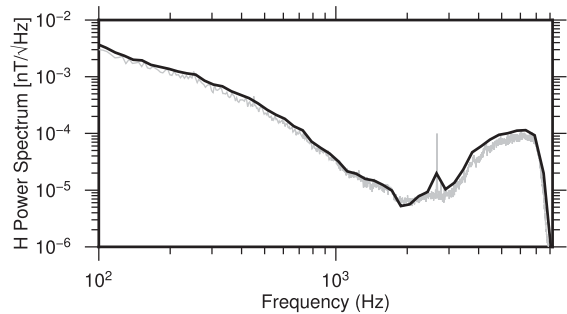
#### 3.2.1 First criterion of significant amplitude

We want to detect time–frequency patterns of atmospheric waves by comparing their amplitude to the background noise and, thus, determine scale-dependent thresholds that are generally useful for selecting segments of data with high signal-to-noise ratios. Background noise power spectrum is estimated by the method of Torrence & Compo (1998). They have used Monte Carlo methods and considered stationary time-series  $s$  of length  $N$  described by order one Auto-Regressive (AR-1) process or Gaussian white noise. They demonstrated that the square modulus wavelet distribution:  $D(a_k) = \{|\mathcal{W}_\psi(a_k, b_i)[s]|^2, i = 1 \dots N\}$  at each scale  $a_k$ , is related to the  $\chi_2^2$  distribution.  $D(a_k)$  is distributed as  $P_k \sigma^2 \chi_2^2/2$  (with same variance and confidence levels), where  $\sigma$  is the noise standard deviation and  $P_k$  the mean normalized Fourier background energy spectrum of signal  $s$  at frequency  $f_k$  corresponding to the scale  $a_k$  (such that  $a_k = (\omega_0 + \sqrt{2 + \omega_0^2})/4\pi f_k$ ). The subscript 2 on the  $\chi_2^2$  distribution corresponds to the number of degrees of freedom (2 for a complex mother wavelet, 1 for a real mother wavelet).

Thus, one can define a threshold value  $\lambda(a_k)$  at each scale  $a_k$  with confidence level  $(1 - \alpha)$  which depends on

$$\lambda(a_k, \alpha) = \frac{1}{2} P_k \sigma^2 \chi_2^2(\alpha). \quad (5)$$

The confidence level  $(1 - \alpha)$  refers to a probability confidence level where  $0 < \alpha < 1$  is the complementary probability. All wavelet coefficients above this threshold are considered as statistically significant. Unfortunately, knowledge of the electromagnetic background



**Figure 5.** Comparison between Fourier spectrum  $H_k$  (thin grey line) and global wavelet spectrum (thick black line) of north–south magnetic field for an extract of 262 144 samples.

spectrum is not trivial since background EM atmospheric activity does not follow any classical shape such as white or red noise processes. For example, it presents deviations such as the so-called MT and AMT dead bands (Chave & Jones 2012), to be taken into consideration. We follow the suggestion of Torrence & Compo (1998), and use the global wavelet spectrum of the magnetic field as background spectrum. The global wavelet spectrum is defined as:

$$G_\psi^2(a)[s] = \int_{-\infty}^{+\infty} db |\mathcal{W}_\psi(a, b)[s]|^2. \quad (6)$$

Fig. 5 presents an example of the comparison between the global wavelet spectrum and the Fourier power spectrum  $H_k$ . The global wavelet spectrum appears as a smoothed Fourier spectrum reducing the effects of abrupt variations in the Fourier spectrum. Considering the short temporal support of the studied EM waves, this spectrum will mostly reflect the energy of the background activity.

Thus using the global wavelet spectrum as background spectrum  $P_k$ , the wavelet modulus distribution at each scale  $a_k$  is constrained by the following threshold at confidence level  $1 - \alpha$ :

$$\lambda(a_k, \alpha) = \frac{1}{2} G_\psi^2(a_k)[s] \chi_2^2(\alpha). \quad (7)$$

All wavelet coefficients above this threshold will be considered as statistically significant and kept for further analysis. As discussed by Torrence & Compo (1998) and detailed by Maraun *et al.* (2007), it can be shown that even pure white or red noise processes create local maxima of high magnitude in the time–frequency domain that will be kept after thresholding. Another criterion based on dispersion patterns is thus necessary to assess with accuracy the existence of EM waves in MT time-series.

#### 3.2.2 Second criterion, extent in time–frequency (dispersion pattern)

Another criterion we use to discriminate EM waves from noise, is supported by their frequency band and their dispersion behaviour. Indeed, as illustrated in Figs 2–4, different patterns in time–frequency domains are clearly identified. We use these characteristic patterns as a second criterion which also yields automatic segmentation. All sought EM waves have broad frequency content. This is characterized by a large number of high wavelet coefficients along the scale axis ( $a$  values) within a very short time ( $b$  interval). Atmospherics and whistlers are also affected by dispersion. The maximum in correlation with a large-scale wavelet (lower frequency) is then located later than those with a short-scale wavelet (higher frequency). Dispersion affects this series of high wavelet coefficients by a shift along the time axis ( $\Delta b$  of a few milliseconds

as seen in Figs 2 and 4). Two main behaviours are thus expected and are used to discriminate EM waves from noise:

- (i) Large frequency content and no dispersion for slow tails, with vertical maxima chains.
- (ii) Large frequency content and dispersion effect for atmospherics and whistlers, with curved maxima chains.

From the wavelet coefficients matrix, we compute and chain local modulus maxima along the scale axis (see Appendices B and C). The chain rule is wave-dependent, but the general rule is to chain two local maxima along a dispersion curve controlled by a dispersion coefficient  $D$ . Flexibility is allowed and controlled by the correlation kernel of the analysing wavelet (Farge 1992; Maraun *et al.* 2007). The exact chain rule implemented in the algorithm is described in Appendix C.

Owing to the diversity of EM wave intensities and polarizations, we restrict our searches to specific maxima chains with fixed extent in the time–frequency domain. For example, we may decide to look for atmospherics having significant content in the frequency range 4–10 kHz, or, be more ambitious and only look for atmospherics having significant content in the frequency range 2–16 kHz. The number of detected events will be highly dependent on this choice. Detecting more events is achieved by lowering the confidence level  $\alpha$  or by reducing the expected frequency range in the maxima chains.

### 3.3 Polarization attributes

Each monochromatic plane wave is characterized by specific polarization parameters in the horizontal plane. The polarization represents the space evolution of the electric and magnetic components with time. This evolution describes an ellipse in the orthogonal plane which ellipse can be characterized by three attributes,  $e$  the ellipticity,  $\theta$  the polarization angle and  $\Delta\phi$  the phase difference between orthogonal components.

The ellipticity  $e$  represents the ratio of the minor  $r$  and the major  $R$  axis with:  $e = (r/R) \in [0, 1]$ . The polarization angle  $\theta$  is the angle between the horizontal axis  $y$ , geomagnetic east (right-handed down common to MT measurements) and the major axis with  $\theta \in [-\pi/2, \pi/2]$ . The phase difference is computed between the horizontal components, with  $\Delta\phi \in [-\pi, \pi]$ .

Many methods exist to recover the polarization attributes by retrieving polarizations parameters on hodograms or in Fourier domain (Fowler *et al.* 1967). We choose the wavelet based determination of polarization attributes described by Diallo *et al.* (2006). Let the signal  $c$  be defined by

$$c(t) = h_y(t) + ih_x(t), \quad (8)$$

where  $h_x$  and  $h_y$  are the north–south and east–west component of the magnetic field respectively. The time–frequency polarization attributes are given by

$$e(a, b) = \frac{|\mathcal{W}_\psi^+(a, b)[c]| - |\mathcal{W}_\psi^-(a, b)[c]|}{|\mathcal{W}_\psi^+(a, b)[c]| + |\mathcal{W}_\psi^-(a, b)[c]|} \quad (9)$$

$$\theta(a, b) = \frac{1}{2} \arg(\mathcal{W}_\psi^+(a, b)[c] \otimes \mathcal{W}_\psi^-(a, b)[c]) \quad (10)$$

$$\Delta\phi(a, b) = \arg\left(\frac{\mathcal{W}_\psi^+(a, b)[c] + (\mathcal{W}_\psi^-(a, b)[c])^*}{\mathcal{W}_\psi^+(a, b)[c] - (\mathcal{W}_\psi^-(a, b)[c])^*}\right) + \frac{\pi}{2}, \quad (11)$$

where  $\mathcal{W}_\psi^+(a, b)[c]$  and  $\mathcal{W}_\psi^-(a, b)[c]$  are the coefficients obtained using a progressive and regressive wavelet respectively

(Holschneider 1995). Progressive and regressive wavelets are respectively defined by

$$\widehat{\psi}_m^+(f) = \begin{cases} \widehat{\psi}_m(f) & f \geq 0 \\ 0 & f < 0 \end{cases} \quad (12)$$

$$\widehat{\psi}_m^-(f) = \begin{cases} 0 & f \geq 0 \\ \widehat{\psi}_m(f) & f < 0 \end{cases} \quad (13)$$

where  $\widehat{\psi}_m$  is the Fourier transform of the Morlet wavelet (eq. 4). Diallo *et al.* (2006) emphasize that these instantaneous coefficients should be analysed within time–frequency windows to properly characterize source signals. This method has been previously used to describe cultural noise sources in Controlled-Source AMT experiments by Escalas *et al.* (2013) or to determine the polarization attributes of geomagnetic pulsations by Kulesh *et al.* (2007).

To recover the polarization attributes of each event, we use the maxima chain kept at the detection stage. For maxima chain  $i$ , at each scale  $a_j$ , let  $b_{i,j}$  be the time position of the local maxima in the maxima chain. We define a distance  $d_j$  based on the correlation kernel of the mother wavelet (Appendix A). We then store wavelet polarization coefficients according to

$$\begin{aligned} D_{e_{i,j}} &= \{e(a_j, b), |b - b_{i,j}| < d_j\} \\ D_{\theta_{i,j}} &= \{\theta(a_j, b), |b - b_{i,j}| < d_j\} \\ D_{\Delta\phi_{i,j}} &= \{\Delta\phi(a_j, b), |b - b_{i,j}| < d_j\}. \end{aligned} \quad (14)$$

For each event  $i$ , the polarization attributes  $e_i$ ,  $\theta_i$  and  $\Delta\phi_i$  are determined by taking the median value of the distributions  $D_{e_{i,j}}$ ,  $D_{\theta_{i,j}}$ ,  $D_{\Delta\phi_{i,j}}$  respectively for all scales. Note that careful attention is paid to the determination of circular quantities such as  $\theta$  and  $\Delta\phi$  using the methodology of directional statistics (Fisher 1995).

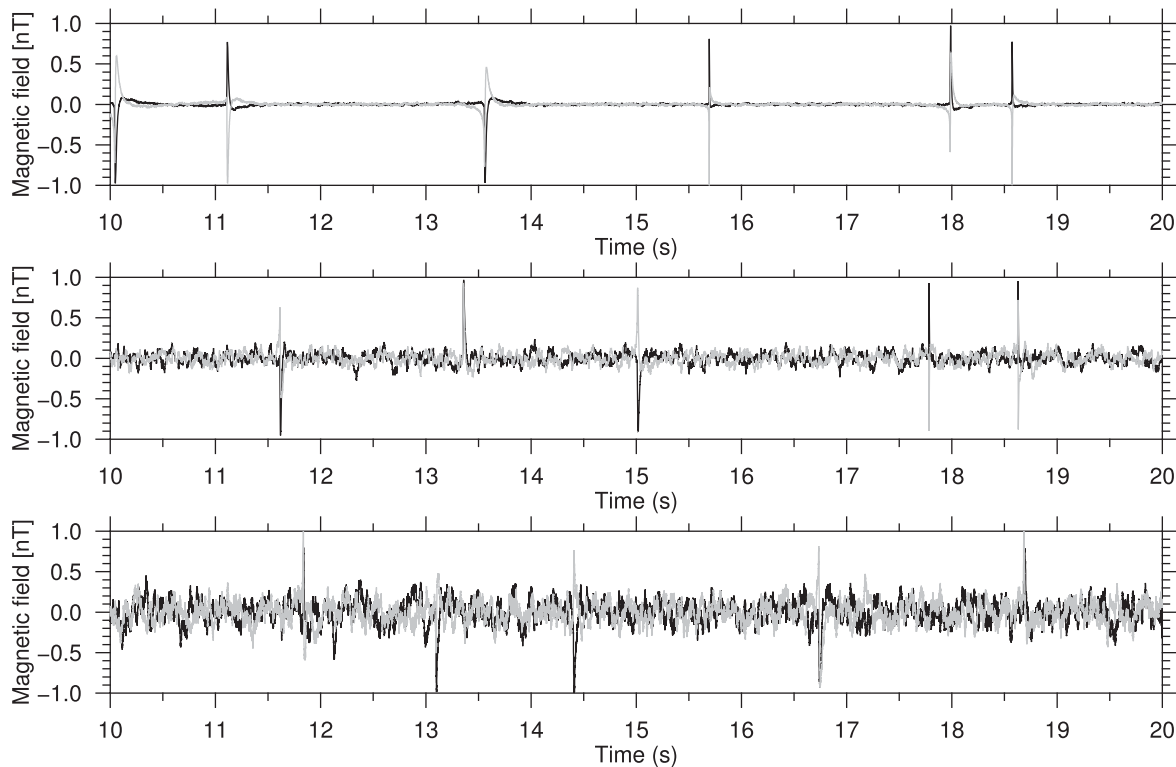
### 3.4 Test on synthetic data

We first apply the ADEM algorithm on synthetic data created by following Wait (1960), as described in Appendix C. Three types of slow tails time-series are synthesized with different noise level (Fig. 6). The lowest noise level is typical of most AMT data while the moderate noise level corresponds to noisy AMT data. The highest noise level is  $\sigma = 0.02$  nT and is used to test the robustness of the method but is not representative of most AMT data. Higher values for  $\sigma$  result in time-series where the noise level is too high and unrealistic of field AMT data. This test was carried out to assess the efficiency of the code in the detection and characterization of non-monochromatic waves. The synthetic time-series are 40 s long. For each event we attribute a random time index and random polarization attributes. Results are described in Table 1.

Let us now discuss about results without remote site first. For low noise level, only one true event is not detected and no false event is detected by the code. The characterization of polarization attributes is efficient with almost no difference between recovered attributes and true attributes. For moderate noise level, only one true event is not detected and no false event is detected. Error slightly increases but remains low. Large noise level leads to the recovery of only 70 per cent of true events. Even worse for automatic detection, false events are detected. In spite of the trouble with false events, polarization attributes are correct for the true events recovered by the process.

When a remote time-series is used in the detection, it removes every false detection made by the ADEM algorithm. The remote station is therefore critical in the detection of atmospherics to





**Figure 6.** Synthetic time-series. Top: low level of noise. Middle: medium noise level. Bottom: high noise level. Black: north–south magnetic field. Grey: east–west magnetic field.

**Table 1.** Results of synthetic tests. Subscript  $m$  indicates the model parameters. Subscript  $d$  indicates the code results. SS indicates single site detection. RR indicates remote reference detection.

Noise level	Low ( $\sigma = 0.001$ nT)		Moderate ( $\sigma = 0.01$ nT)		High ( $\sigma = 0.02$ nT)	
	SS	RR	SS	RR	SS	RR
Number of detected true events (over a total of 40)	39	39	39	39	30	27
Number of detected false events	0	0	0	0	3	0
Ellipticity difference median: ( $e_m - e_d$ )	-0.002	-0.002	0.012	0.011	0.046	0.050
Standard deviation on distribution	0.004	0.004	0.045	0.045	0.086	0.086
Polarization angle difference median: ( $\theta_m - \theta_d$ ) ( $^\circ$ )	0.022	0.066	0.78	0.39	-2.16	-3.00
Standard deviation on distribution	0.6	0.76	12.9	12.7	9.8	10
Phase shift difference median: ( $\Delta\phi_m - \Delta\phi_d$ ) ( $^\circ$ )	-0.187	-.210	-0.659	-0.962	-2.32	0.692
Standard deviation on distribution	0.773	0.960	4.91	5.00	8.66	8.45

prevent any false detection from the time-series. The use of a remote station is standard process in AMT acquisition so does not act as a burden in the field acquisition scheme.

## 4 APPLICATION

### 4.1 Atmospherics

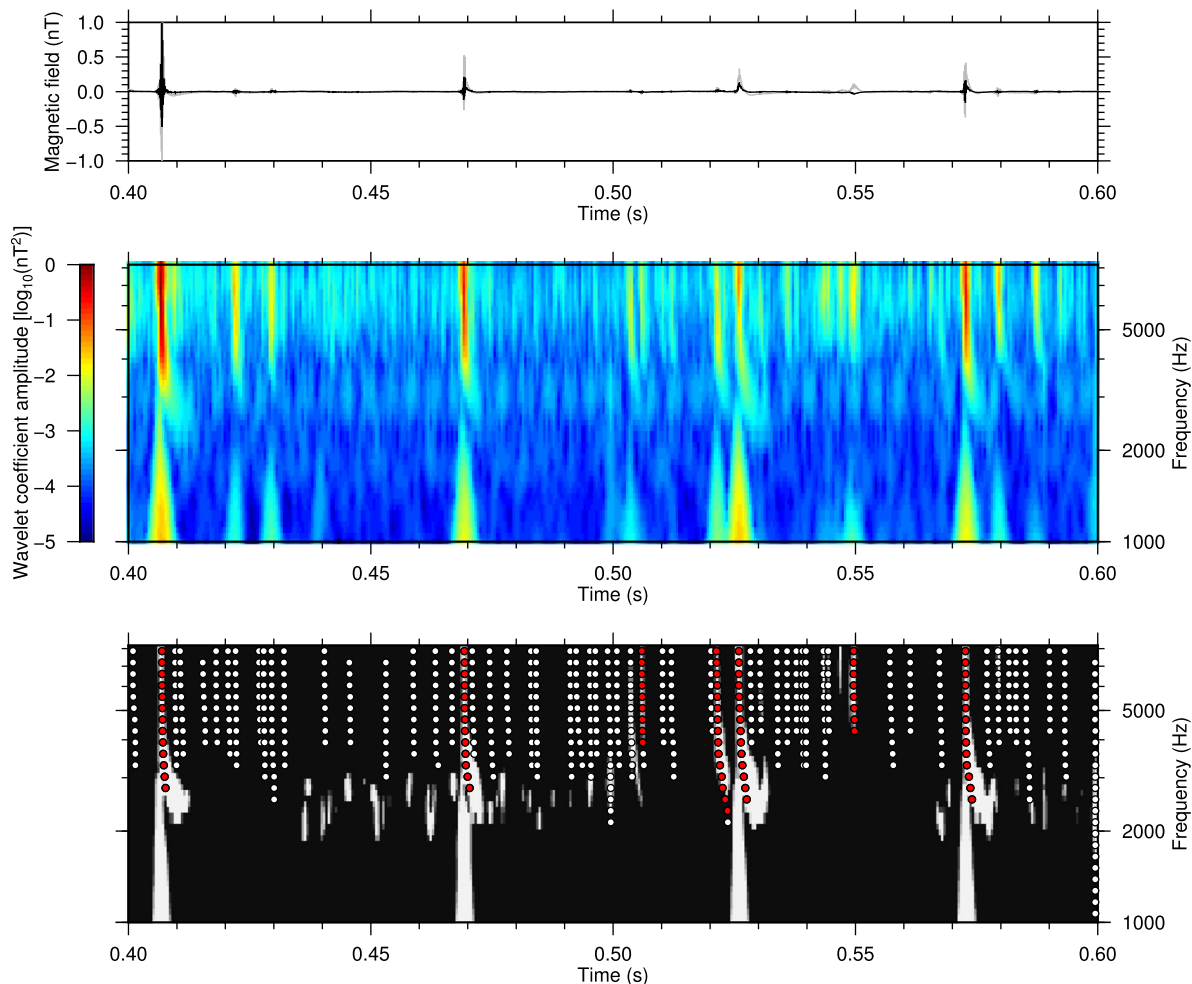
#### 4.1.1 Characterization

We first apply the ADEM method to Guenon time-series; Youketi being the remote station in the detection process. The detection was set using a 90 per cent confidence level. Sought maxima chains were expected to have significant content from 8 kHz down to 3 kHz. The correlation kernel was obtained using a critical value of 0.90. On this basis, 172 atmospherics are detected. To illustrate the detection

process, we have represented in Fig. 7 the two stages described previously on 0.2 s of the analysed time-series. The distribution of polarization attributes is illustrated Fig. 8. Most of detected waves have quasi-null ellipticity and phase difference around 0 modulo  $\pi$ . This indicates linear polarization for all events. The polarization angle indicates a large distribution of wave polarization, with two principal directions of +60 and -40 degrees. Computation of the difference in polarization attributes for local and remote sites common events gives difference below 10 degrees in polarization angle, 15 degrees for phase difference and 0.1 for ellipticity (Fig. 9).

#### 4.1.2 Frequency content and occurrence

The variability of frequency content in magnetic time-series may be illustrated by the systematic use of our algorithm to detect atmospherics with variable frequency content. Several detections are



**Figure 7.** Illustration of detection processes on Guenon time-series by CWT. Top: orthogonal magnetic time-series, black line: north–south magnetic field, grey line: east–west magnetic field. Middle: sum of wavelet spectrograms of time-series. Bottom: white: significant coefficients kept after Torrence & Compo (1998) detection scheme. White and red chains: Maxima chains detected. Red chains: Maxima chains kept after application of both criteria.

performed on the three available time-series: atmospheric with significant frequency content from 5 to 8 kHz, from 4 to 8 kHz, from 3 to 8 kHz and finally from 2.5 to 8 kHz. Only simultaneous events detected at the local (Guenon or Cratère Sud) and Youketi remote station are kept.

Time-series of the flow of detected events (number of events detected by second) at Guenon station are illustrated Fig. 10. Every second, between 15 and 50 atmospheric with frequency content from 5 to 8 kHz are detected. On average, about 33 events with such characteristics per second are detected over the whole minute. Atmospheric from 4 to 8 kHz occur less often, with an average of about 22 detections per second. The search for atmospheric with frequency content from 3 to 8 kHz yields even less detected events: 3 on average per second, with periods of no occurrence at all. Finally, atmospheric with frequency content from 2.5 to 8 kHz occur on average 0.2 times per second. The detection here emphasizes the scarcity of events with significant frequency content inside the AMT dead band.

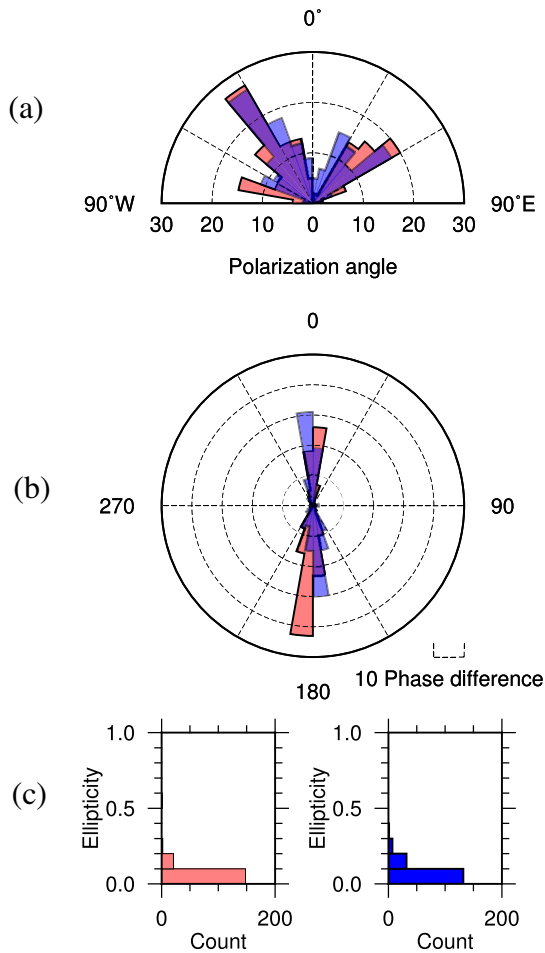
Time-series of the flow of detected events at Cratère Sud station in the morning and in the afternoon are illustrated Fig. 11. The scarcity of wide frequency content events is still displayed but the major feature is the large difference in occurrence rate between morning time and afternoon for each type of event. Morning is characterized

by a very low number of events with frequency content in the AMT dead band with only two events with frequency content from 2.5 to 8 kHz during 60 s. In the afternoon, the occurrence rate of wide frequency range events with frequency content between 2.5 to 8 kHz is larger with about 1 event per second on average.

## 4.2 Slow tails

### 4.2.1 Characterization

We repeat the methodology on Guenon time-series but this time to detect slow tails. The detection was set using 90 per cent confidence level. Maxima chains with significant frequency content from 64 Hz up to 1 kHz are sought. The critical kernel was built using a critical value of 0.9. As a result, 287 events are detected (Fig. 12). The principal direction of polarization is about  $-45$  degrees west. As for atmospheric's case, the distribution of the ellipticity and phase difference indicates linear polarization for most of the slow tails with some departure that may be due to noise in the data. Common detected waves with Guenon and Youketi have similar features. Differences in polarization attributes for common events on both sites show values below 10 degrees for polarization, 20 degrees for phase difference and 0.1 for ellipticity (Fig. 13).



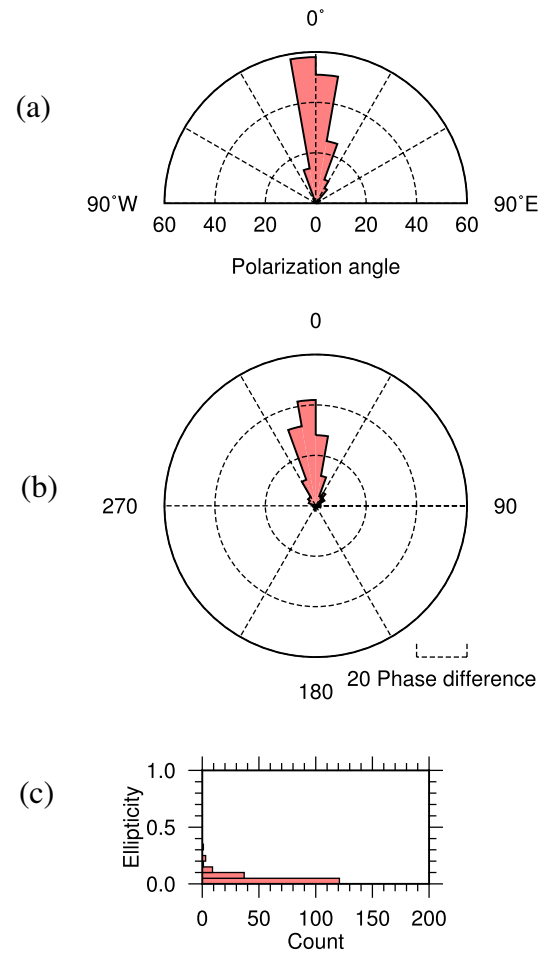
**Figure 8.** Distributions of atmospheric polarization attributes obtained at Guenon and Youkети AMT stations. Red: Guenon. Blue: Youkети. (a) Polarization angle. (b) Phase difference. (c) Ellipticity.

#### 4.2.2 Frequency content and occurrence

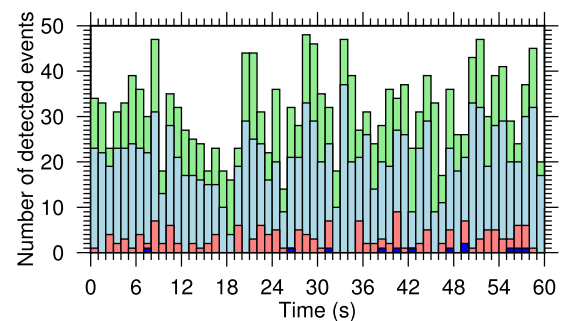
Similar to the process performed on atmospheric, we study the relationship between frequency content and occurrence. We use 1 min of measurement at 16384 Hz. Detection of slow tails is performed with frequency content from 128 to 500 Hz, from 128 Hz to 1 kHz, from 128 Hz to 1.5 kHz, and finally from 128 Hz to 2.5 kHz.

Time-series for the flow of detected events (number of events detected by second) on Guenon station are illustrated Fig. 14. Concerning slow tails with significant frequency content from 128 Hz to 500 Hz, such events occur quite often, 7.1 events per second on average. Events with frequency content from 128 Hz to 1 kHz occur on average  $5.15 \text{ s}^{-1}$  and events with frequency content from 128 Hz to 1.5 kHz occur with an average of  $3.7 \text{ s}^{-1}$ . Finally, the largest events with frequency content from 128 Hz to 2.5 kHz do not occur often with an average of  $0.83 \text{ s}^{-1}$ , with periods of no occurrence.

Time-series of the flow of detected slow tails at Cratère Sud station in the morning and afternoon are shown Fig. 15. As for the atmospheric case, the difference in occurrence rate for all types of events is considerable. Morning time-series display only a few events with large frequency content (about one every 3 s on average) while afternoon time-series display more than 1 event per second on average.



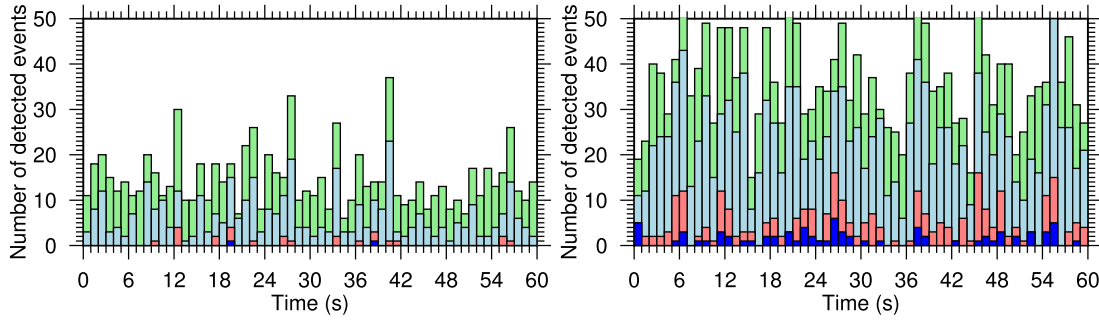
**Figure 9.** Difference of common atmospheric polarization attributes obtained at Guenon and Youkети AMT stations. (a) Polarization angle. (b) Phase difference. (c) Ellipticity.



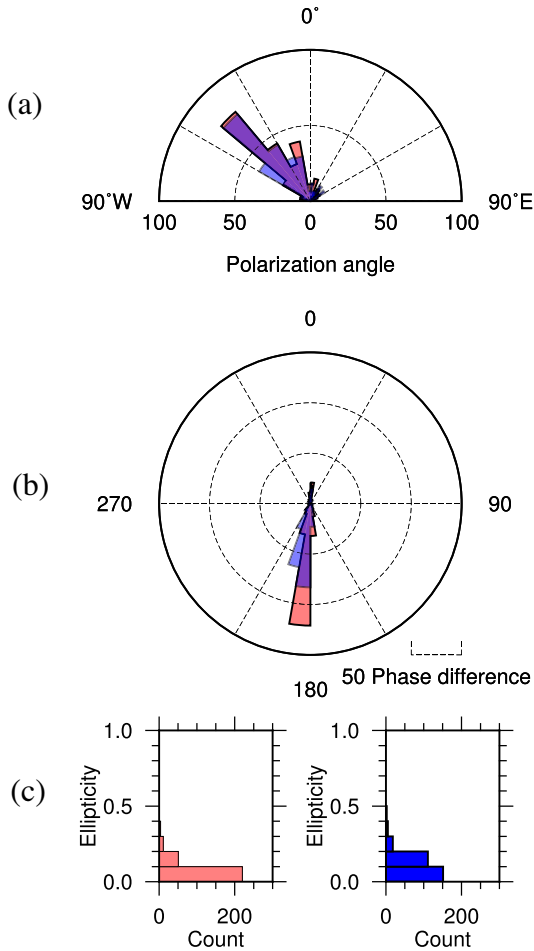
**Figure 10.** Flows of detected atmospheric at Guenon station depending on expected frequency content. Green: from 5 to 8 kHz; light blue: from 4 to 8 kHz; red: from 3 to 8 kHz; dark blue: from 2.5 to 8 kHz.

#### 4.3 Whistlers

Whistlers occur only a few times in our data sets. We are not able to provide extensive analysis over several tens of events such as in the case of atmospheric or slow tails. Nevertheless, the whistler shown in Fig. 4 is detected and analysed by the ADEM code using a confidence level of 0.9. The event is characterized by an ellipticity of  $e = 0.53$ , a polarization angle of  $\theta = 1.8$  in degrees and a phase difference  $\Delta\phi = 73$  degrees.



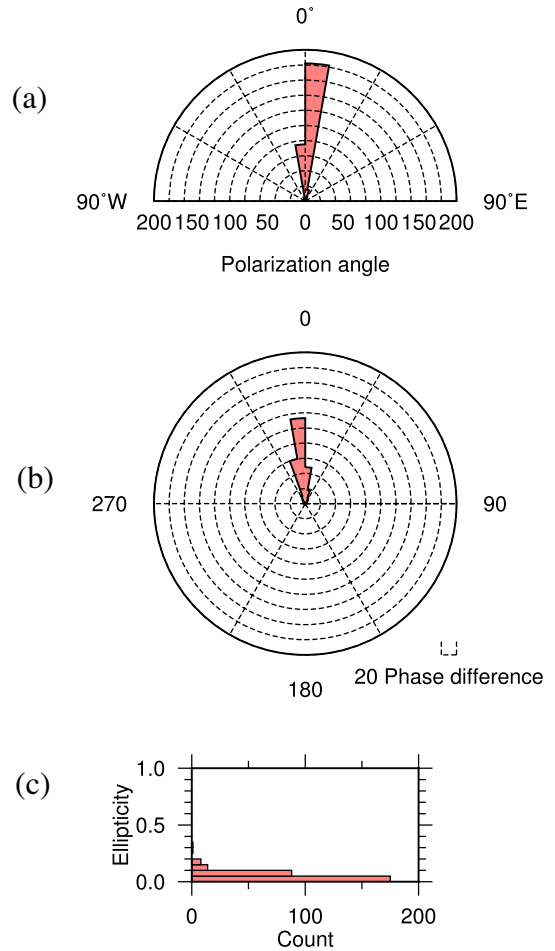
**Figure 11.** Flows of detected atmospheric events at Cratère Sud station depending on expected frequency content. Green: from 5 to 8 kHz; light blue: from 4 to 8 kHz; red: from 3 to 8 kHz; dark blue: from 2.5 to 8 kHz. Left: morning data. Right: afternoon data.



**Figure 12.** Distributions of slow tails polarization attributes obtained at Guenon and Youketi AMT stations. Red: Guenon. Blue: Youketi. (a) Polarization angle. (b) Phase difference. (c) Ellipticity.

**4.4 Discussion—AMT impedance tensor and atmospheric activity**

Because EM waves include narrow and broad frequency and low to high intensity contributions from naturally occurring sources, processing ground-recorded signals into AMT Earth transfer functions can be problematic. Electromagnetic induction methods, developed to constrain the electrical resistivity of the Earth’s subsurface, depend on reliable estimation of these Earth response functions. These EM waves induce electrical currents in the Earth, and a superposition of the source signals and ‘secondary’ EM waves arising

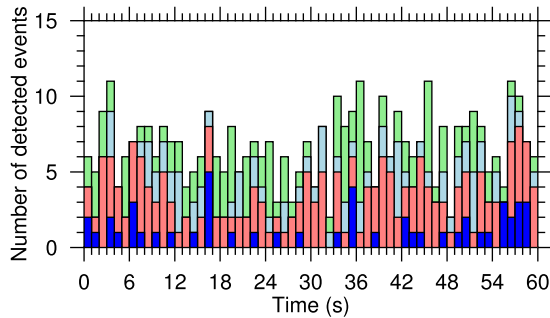


**Figure 13.** Difference of common slow tails polarization attributes obtained at Guenon and Youketi AMT stations. (a) Polarization angle. (b) Phase difference. (c) Ellipticity.

from these induced currents are measured as time-series using electrodes, magnetometers, amplifiers and filters, and digital recording equipment. Superposed EM fields thus measured are related to the particular tensor Earth response function called the impedance (Cagniard 1953):

$$\begin{aligned}
 E_x(\omega) &= Z_{xx}(\omega)H_x(\omega) + Z_{xy}(\omega)H_y(\omega) \\
 E_y(\omega) &= Z_{yx}(\omega)H_x(\omega) + Z_{yy}(\omega)H_y(\omega).
 \end{aligned}
 \tag{15}$$

One major advantage of EM waves such as slow tails and atmospheric is their broad frequency content. This broadness helps to constrain  $Z$  over a wide frequency band at once. Of critical

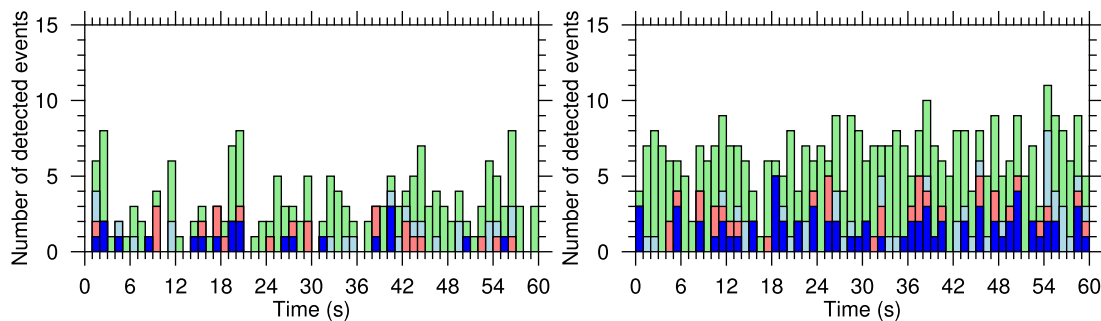


**Figure 14.** Flows of detected slow tails at Guenon station depending on expected frequency content. Green: from 128 to 500 Hz; light blue: from 128 Hz to 1 kHz; red: from 128 Hz to 1.5 kHz; dark blue: from 128 Hz to 2.5 kHz.

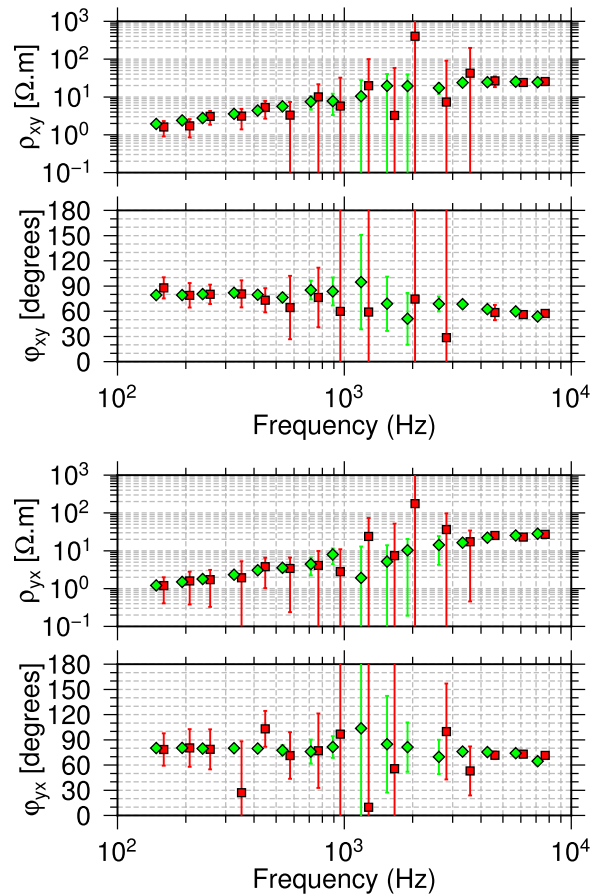
concern is the low signal-to-noise ratio in the so-called AMT dead band, from 1 to 5 kHz, where standard processing procedures often fail to properly determine MT response functions  $Z$  (Garcia & Jones 2002). Using the ADEM method to detect atmospheric and slow tails, and thus defined data segments with higher signal/noise ratios (Sections 4.1 and 4.2) serves to improve AMT impedances estimated for the dead band.

First let us consider the relationship between the relationship between inclusion of ADEM-derived knowledge of atmospheric activity and results from traditional ‘robust’ processing using Huber weighting schemes. We have shown that for Cratère Sud station, the occurrence rate of atmospheric events is lower in the morning than in the evening (Figs 11 and 15). Fig. 16 shows the difference between the processing of the corresponding morning and afternoon time-series using the BIRRP (Bounded Influence Remote Reference Processing) code from the work of Chave & Thomson (2004). Even though estimations of the response functions outside of the AMT dead band are similar, data quality, as evinced by larger uncertainties and scatter in estimates of the mean value, is lower for the morning versus afternoon data. Inside the AMT dead band, processing of morning time-series does not result in acceptable estimations of the response functions. Inversion of such data might lead to a loss of accuracy in the inverted resistivity distribution. Afternoon processing of time-series results in acceptable response functions for the whole frequency range, even in the AMT dead band.

However, Figs 11 and 15 demonstrate that atmospheric and slow tails with significant frequency content inside the AMT dead band occur in the morning time-series. Let us now propose to use the ADEM algorithm to select data segments exhibiting high lightning activity, particularly with substantive frequency content within the dead band, for further processing using standard schemes.



**Figure 15.** Flows of detected slow tails at Cratère Sud station depending on expected frequency content. Green: from 128 to 500 Hz; light blue: from 128 Hz to 1 kHz; red: from 128 Hz to 1.5 kHz; dark blue: from 128 Hz to 2.5 kHz. Left: morning data. Right: afternoon data.



**Figure 16.** Results from MT processing of ELF waves for Cratère Sud time-series using morning (starting at 10 hr 53 min 22 s in solar time) data in red and afternoon (starting at 14 hr 53 min 34 s in solar time) data in green.

We use the algorithm to detect, slow tails presenting maxima chains from 128 to 1024 Hz providing a catalogue of significant geomagnetic events. We use a confidence level of 0.9 for amplitude thresholding. Each segment of ‘active’ time-series starts 0.1 s before the occurrence of a selected event. While the difference between two successive events is below 0.5 s, the segment is still considered as active. When no major event is found for 0.5 s, the active segment stops 0.1 s after the last major event in the active segment. We consider as ‘quiet’ the remaining part of the time-series. Active and quiet segments are illustrated in Fig. 17.

In the following, we distinguish three distinct data sets. The first one corresponds to active time periods, the second one to quiet time periods, and the third one gathering the whole time-series. In

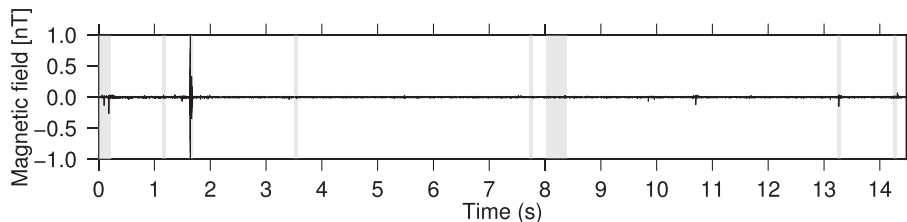


Figure 17. Illustration on east–west magnetic field of the selection of active segments (grey-shaded areas) for 16 s of Cratère Sud time-series.

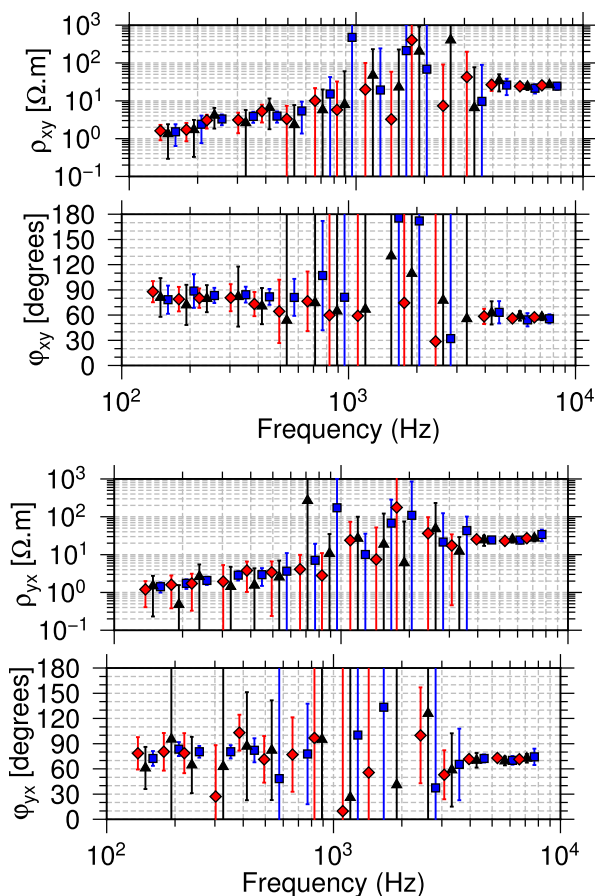


Figure 18. Results from MT processing of Cratère Sud morning data using the whole data series (red), quiet times (black) or active times (blue).

Fig. 18, we present the comparison between robust processing of the whole time-series and the selected active or quiet parts. Slight improvement is observed, both in the estimates and in the reduction of error bars between active and whole time-series. Still, it shows that even selecting active parts of the time-series before using standard robust procedures is still not enough to be able to properly process data scarce in atmospheric events. New procedures need to be adapted to such data sets but this is beyond the scope of this paper. Still, the comparison between active and quiet data sets demonstrates that processing results are mostly dependent on large signal-to-noise ratio events.

## 5 CONCLUSION

We develop a methodology for the automatic detection of lightning-based electromagnetic waves (ADEM). It is based on the characteristics of the EM waves in time–frequency plane. Two criteria are used: one relying on the amplitude of the expected waves; the

second one lies in the behaviour of local maxima in the time–frequency domain. Using both criteria, this procedure allows for direct detection, segmentation and characterization of three types of lightning source EM waves, namely atmospheric, slow tails and whistlers. This scheme can be extended to events such as q-Bursts or twecks. It allows either for a large detection where many events are recovered or a precise detection for events with wide frequency content. Using this technique, we have shown that large signal-to-noise ratio events have significant impact on MT data processing, specially for frequencies near the AMT dead band. Finally, the use of a remote station is critical for a proper detection of lightning-based EM waves. This does not change the framework of AMT acquisition where a remote station is almost always set up for processing. Only one site location is shown here but this methodology has been successfully applied to other locations at different times of day (including night-time) and at different latitudes.

However, we have shown that large events with significant content in AMT dead band have low occurrence frequency. This suggests that longer time-series may be necessary to acquire proper AMT data depending on the atmospheric activity or new methodologies need to be adapted to cope with the scarcity of events in AMT time-series. In our examples, 1 s of data may not be sufficient if no slow tails are present in the signal. Finally, this methodology allows to characterize the geomagnetic activity in events number, frequency content and polarization of geomagnetic waves in AMT time-series. This provides a way to assess the atmospheric activity before launching a new AMT campaign and optimize the acquisition schedule. This methodology also allows to control the quality of AMT responses obtained through MT processing of time-series by monitoring the actual content of atmospheric waves.

## ACKNOWLEDGEMENTS

We would like to thank Dean Livelybrooks and one anonymous reviewer who greatly helped to improve this manuscript. Data used in this manuscript have been collected using instrument from the CNRS INSU MT pool, financial support by CNRS INSU AO Aléas, Risques et Catastrophes Telluriques (AVMT project) and ANR RISKMAT (Domoscan project) and benefited from the implication of Sheldon Warden, Florence Nicollin, Fabrice Dufour and the OVSG observatory.

## REFERENCES

- Cagniard, L., 1953. Basic theory of the magnetotelluric method of geophysical prospecting, *Geophysics*, **18**, 605–635.
- Chand, R., Israil, M. & Rai, J., 2009. Schumann resonance frequency variations observed in magnetotelluric data recorded from Garhwal Himalayan region India, *Ann. Geophys.*, **27**(9), 3497–3507.
- Chave, A.D., 2014. Magnetotelluric data, stable distributions and impropriety: an existential combination, *Geophys. J. Int.*, **198**, 622–636.

Chave, A.D. & Jones, A.G., 2012. *The Magnetotelluric Method: Theory And Practice*, Cambridge Univ. Press.

Chave, A.D. & Thomson, D.J., 2004. Bounded influence magnetotelluric response function estimation, *Geophys. J. Int.*, **130**(2), 475–496.

Christian, H.J., Blakeslee, R.J., Boccippio, D.J., Boeck, W.L., Buechler, D.E., Driscoll, K.T. & Stewart, M.F., 2003. Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. geophys. Res.*, **108**(D1), ACL 4-1–ACL 4-15.

Cummer, S.A., 2000. Modeling electromagnetic propagation in the Earth-Ionosphere waveguide, *IEEE Trans. Antennas Propag.*, **48**, 1420–1429.

Cummer, S.A. & Inan, U.S., 2000. Modeling ELF radio atmospheric propagation and extracting lightning currents from ELF observations, *Radio Sci.*, **35**, 1437–1444.

Diallo, M.S., Kulesh, M., Holschneider, M., Scherbaum, F. & Adler, F., 2006. Characterization of polarization attributes of seismic waves using continuous wavelet transforms, *Geophysics*, **71**(3), 67–77.

Escalas, M., Queralt, P., Ledo, J. & Marcuello, A., 2013. Polarization analysis of magnetotelluric time series using a wavelet-based scheme: A method for detection and characterisation of cultural noise sources, *Phys. Earth planet. Inter.*, **218**, 31–50.

Egbert, G., 2002. Processing and interpretation of electromagnetic induction array data, *Surv. Geophys.*, **23**(2), 207–249.

Farge, M., 1992. Wavelet transforms and their applications to turbulence, *Annu. Rev. Fluid Mech.*, **24**, 395–457.

Fisher, N.I., 1995. *Statistical Analysis of Circular Data*, Cambridge Univ. Press, p. 277

Fowler, R.A., Kotick, B.J. & Elliott, R.D., 1967. Polarization analysis of natural and artificially induced geomagnetic micropulsations, *J. geophys. Res.*, **72**(11), 2871–2883.

Garcia, X. & Jones, A.G., 2002. Atmospheric sources for audio-magnetotelluric (AMT) sounding, *Geophysics*, **67**(2), 448–458.

Garcia, X. & Jones, A.G., 2008. Robust processing of magnetotelluric data in the AMT dead band using the continuous wavelet transform, *Geophysics*, **73**(6), 223–234.

Goupillaud, P., Grossmann, A. & Morlet, J., 1984. Cycle-octave and related transforms in seismic signal analysis, *Geoexploration*, **23**, 85–102.

Helliwell, R.A., 1965. *Whistlers and Related Ionospheric Phenomena*, Stanford Univ. Press.

Heydt, G., 1982. Instrumentation, in *Handbook of Atmospheric*, Vol. 2, pp. 203–256, ed. Volland, H., CRC Press.

Holschneider, M., 1995. *Wavelets, An Analysis Tool*, Oxford Science Publications.

Huang, G.L., Wang, D.Y. & Song, Q.W., 2004. Whistler waves in Freja observations, *J. geophys. Res.*, **109**, A02307 doi:10.1029/2003JA010137.

Hughes, H.G. & Papert, R.A., 1975. Propagation prediction model selection using VLF atmospherics, *Geophys. Res. Lett.*, **2**, 96–98.

Kulesh, M., Nosé, M., Holschneider, M. & Yumoto, K., 2007. Polarization analysis of a Pi2 pulsation using continuous wavelet transform, *Earth Planets Space*, **59**, 961–970.

Laby, T.H., McNeill, J.J., Nicholls, F.G. & Nickson, A.F.B., 1940. Waveform, energy and reflexion by the ionosphere, of atmospherics, *Proc. R. Soc. A*, **174**, 145–163.

Larnier, H., Sailhac, P. & Chambodut, A., 2016. New application of wavelets in magnetotelluric data processing: reducing impedance bias, *Earth Planets Space*, **68**(70), doi:10.1186/s40623-016-0446-9

Mackay, C. & Fraser-Smith, A.C., 2010. Lightning location using the slow tails of sferics, *Radio Sci.*, **45**(5), doi:10.1029/2010RS004405

Maraun, D., Kurths, J. & Holschneider, M., 2007. Nonstationary gaussian processes in wavelet domain: Synthesis, estimation, and significance testing, *Phys. Rev. E*, **75**, 016707, doi:10.1103/PhysRevE.75.016707

Rakov, V.A. & Uman, M.A., 2003. *Lightning: Physics and Effects*, Cambridge Univ. Press.

Sailhac, P., Larnier, H., Warden, S., Nicollin, F., Zlotnicki, J. & Bruère, F., 2016. Processing and modeling magnetotelluric data at La Soufrière of Guadeloupe lava dome (France), EMIW 2016 Chiang Mai Extended Abstract.

Schumann, W.O. & Koenig, H., 1954. Über die beobachtung von atmosphärischen bei geringsten frequenzen, *Naturwissenschaften*, **41**, 183–184.

Strangway, D.W., Swift, C.M. & Holmer, R.C., 1973. The application of audio frequency magnetotellurics (AMT) to mineral exploration, *Geophysics*, **38**, 1159–1175.

Torrence, C. & Compo, G.P., 1998. A practical guide to wavelet analysis, *Bull. Am. Meteorol. Soc.*, **79**(1), 61–78.

Wait, J.R., 1960. On the theory of the slow tail portion of atmospheric waveforms, *J. geophys. Res.*, **65**(7), 1939–1946.

Wait, J.R., 1962. *Electromagnetic Waves in Stratified Media*, Pergamon Press.

Williams, E.R., 1992. The Schumann resonance: a global tropical thermometer, *Science*, **256**, 1184–1187.

Zhang, Y. & Paulson, K.V., 1997. Enhancement of signal-to-noise ratio in natural-source transient MT data with wavelet transform, *Pure appl. Geophys.*, **149**, 405–419.

## APPENDIX A: REPRODUCING KERNEL OF MORLET WAVELET

The reproducing kernel  $K_\psi$  of the wavelet  $\psi$  characterizes the correlation of the wavelet transform between two different points in the scale-time domain (Farge 1992; Holschneider 1995). It is defined by the correlation of the wavelet function dilated and translated using different parameters  $(a_1, b_1)$  and  $(a_2, b_2)$ :

$$K_\psi(a_1, a_2, b_1, b_2) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \frac{1}{a_1} \psi\left(\frac{t-b_1}{a_1}\right) \frac{1}{a_2} \psi\left(\frac{t-b_2}{a_2}\right) dt \quad (\text{A1})$$

where  $C_\psi$  is the following normalizing factor defined from the Fourier transform of the wavelet  $\psi$ :

$$C_\psi = \int_{-\infty}^{\infty} |\hat{\psi}_0(\omega)| \frac{d\omega}{|\omega|}. \quad (\text{A2})$$

By simple change of variables, one can write the reproducing kernel as the wavelet transform of the wavelet  $\psi$ :

$$K_\psi(a_1, a_2, b_1, b_2) = \frac{1}{C_\psi} \mathcal{W}_\psi\left(\frac{a_2}{a_1}, \frac{b_2 - b_1}{a_1}\right). \quad (\text{A3})$$

Calculation for the Morlet wavelet yields the following formula (Maraun *et al.* 2007):

$$K_\psi(a_1, a_2, b_1, b_2) = \frac{2a_1a_2}{a_1^2 + a_2^2} \exp\left(i\omega_0 \frac{a_1 + a_2}{a_1^2 + a_2^2} (b_2 - b_1)\right) \times \exp\left(-\frac{1}{2} \frac{(b_2 - b_1)^2 + \omega_0^2 (a_2 - a_1)^2}{a_1^2 + a_2^2}\right). \quad (\text{A4})$$

The correlation length of the Morlet reproducing kernel is used to define the resolution in time. By choosing a critical correlation level  $c$ , we define the resolution time limit at scale  $a$  by

$$l_{\psi,c}(a_1) = \max(b_2 |K_\psi(a_1, a_2, b_1, b_2)| > c, \forall a_2, \forall b_2) - \min(b_2 |K_\psi(a_1, a_2, b_1, b_2)| > c, \forall a_2, \forall b_2). \quad (\text{A5})$$

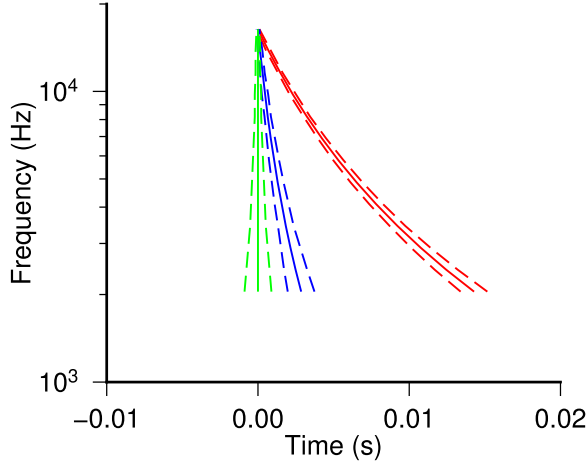
## APPENDIX B: DEFINITION AND RULES OF MAXIMA CHAINS

At scale  $a_j$  a local maximum at index  $i$  is defined by

$$|\mathcal{W}_\psi(a_j, b_i)| > |\mathcal{W}_\psi(a_j, b_{i+1})| \\ |\mathcal{W}_\psi(a_j, b_i)| > |\mathcal{W}_\psi(a_j, b_{i-1})| \quad (\text{B1})$$

Helliwell (1965) has shown that the group delay for whistlers (for frequencies below the nose frequency) can be approximated by

$$\Delta t = Df^{-1/2}, \quad (\text{B2})$$



**Figure B1.** Illustration of possible maxima chains with three various dispersion parameters  $D$ . Solid lines illustrate eq. (B2). Dashed lines illustrate maxima broadening due to temporal resolution of the Morlet wavelet.

with  $D$  being a constant factor for every frequency called the dispersion (in  $s^{1/2}$ ). This parameter is related to the density of electron on the path taken by the whistler wave (Helliwell 1965). The time difference is characterized by being proportional to the inverse square root of the frequency. In the following, we use this dispersion assumption to chain local maxima in the time frequency plane for the three types of waves namely the atmospherics, the slow tails and the whistlers. The factor  $D$  is chosen as the parameter responsible for dispersion.

For slow tails, no dispersion is expected, so  $D$  is set at 0. For atmospherics and whistlers, the dispersion depends on the path taken from the source. The  $D$  is therefore variable and has to be assigned to each type of events. In the paper, we have used  $D = 0.012 t^{1/2}$  for atmospherics and  $D = 0.12 t^{1/2}$  for whistlers waves. Fig. B1 illustrate the influence of  $D$  on sought maxima chains. Besides, because of noise in the time-series, we allow a slight variation along the time axis around the function defined by eq.(B2). This time interval where the local maxima is sought is defined as proportional to the half width  $l_{\psi,c}(a)/2$  of the Morlet kernel  $K$  above a critical value  $c$  (as detailed in Appendix A).

In the following, we call  $M_i$  the set of all local maxima at scale  $a_i$ . The chain rule is as follows:

- (i) At starting scale  $a_\alpha$ , we seek every local maxima and determine  $M_\alpha$ , the set of all local maxima at scale  $a_\alpha$ .
- (ii) For every  $b_{i_\alpha} \in M_\alpha$ , we connect the local maxima ( $a_\alpha, b_{i_\alpha}$ ) to a maximum on larger scale when

$$\exists b_{i_{\alpha+1}} \begin{cases} b_{i_{\alpha+1}} \in M_{\alpha+1} \\ b_{i_{\alpha+1}} \in [b_{i_\alpha} + \Delta_t - l_{\psi,c}(a_{\alpha+1})/2 : b_{i_\alpha} + \Delta_t + l_{\psi,c}(a_{\alpha+1})/2] \end{cases} \quad (\text{B3})$$

- (iii) To connect it to a maximum on smaller scale:

$$\exists b_{i_{\alpha-1}} \begin{cases} b_{i_{\alpha-1}} \in M_{\alpha-1} \\ b_{i_{\alpha-1}} \in [b_{i_\alpha} - \Delta_t - l_{\psi,c}(a_{\alpha-1})/2 : b_{i_\alpha} + \Delta_t + l_{\psi,c}(a_{\alpha-1})/2] \end{cases} \quad (\text{B4})$$

## APPENDIX C: GENERATION OF SYNTHETIC ELF WAVES

To create synthetic magnetic field time-series, we generate magnetic field responses of far lightning pulses following Wait (1960, 1962). The propagation medium is defined as the atmosphere comprised in a spherical shell from the ground at  $r = a$  and the ionosphere at a distance  $r = a + h$  where  $h$  the ionosphere height. Both layers act as waveguide boundary where the field propagates in modes. We assume that the lightning source is equivalent to a vertical electric dipole source on the Earth's surface. The current source is assumed to be an impulse described by a Dirac  $\delta$  function. The frequency domain response is then given by

$$P_s(\omega) = |p_s|. \quad (\text{C1})$$

The vertical electric field at the receiver location is given by

$$E_0(\omega) = i \left( \frac{\eta\omega}{2\pi c\rho} \right) (e^{-i\omega\rho/c}) P_s(\omega), \quad (\text{C2})$$

where  $\rho$  is the distance source-receiver,  $c$  the speed of light,  $\eta$  the impedance of free space and  $\omega = 2\pi f$  the wave pulsation.

Wait (1962) derived the expression of the EM field in frequency domain at receiver location. This field is given by a sum over all the waveguide modes:

$$H_r(\omega) \cong \left( \frac{\rho/R_E}{\sin(\rho/R_E)} \right)^{1/2} \frac{(2\pi c\rho)^{1/2}}{h\eta} \frac{E_0\omega}{(i\omega)^{1/2}} \cdot \sum_{n=0}^{\infty} \Re(S_n(\omega))^{1/2} \delta_n e^{-i\omega\rho S_n(\omega)/c}, \quad (\text{C3})$$

where  $R_E$  is the Earth radius,  $h$  the height of the ionosphere,  $\eta$  the intrinsic impedance of free space.  $\delta_n$  is related to the excitation of each mode and depends on reflection coefficients on the ground and the ionosphere.  $S_n$  is the sinus of the complex angle of the incident wave on the ionosphere. The magnetic field in time domain at receiver location is then given by

$$h_r(t) = \int_{-\infty}^{+\infty} H_r(\omega) e^{i\omega t} d\omega. \quad (\text{C4})$$

The slow tails propagate for frequencies below 2 kHz. The first cut-off frequency  $n = 0$  is about 1.8 kHz (Mackay & Fraser-Smith 2010), so only the zeroth mode is kept in the sum.

Following these assumptions, the magnetic field equation becomes

$$H_r(\omega) \cong \left( \frac{\rho/R_E}{\sin(\rho/R_E)} \right)^{1/2} \frac{(2\pi c\rho)^{1/2}}{h\eta} \frac{E_0\omega}{(i\omega)^{1/2}} \cdot \Re(S_0(\omega))^{1/2} \delta_0 e^{-i\omega\rho S_0(\omega)/c}, \quad (\text{C5})$$

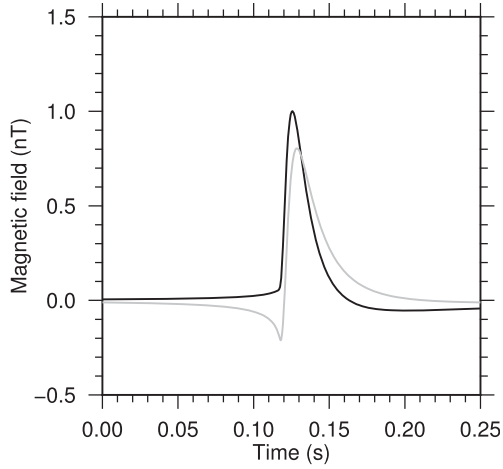
The Earth's radius is set to  $R_E = 6371$  km, the speed of light is  $c = 3 \cdot 10^5$  km/s, free space impedance is equal to  $376.730 \Omega$ , the ionospheric conductivity is  $\sigma = 10^{-6}$  S  $m^{-1}$ , and the free space permittivity  $\epsilon_0 = 9 \cdot 10^{-12}$  F  $m^{-1}$ .  $\delta_0$  is approximately equal to 1 for  $n = 0$  (Wait 1960) and  $S_0$  can be approximated to

$$S_0(\omega) = \left( 1 + \frac{c}{h(i\sigma_i\omega/\epsilon_0)^{1/2}} \right)^{1/2}, \quad (\text{C6})$$

where  $\sigma_i$  is the ionosphere conductivity (assumed to be homogeneous here with a value of  $10^{-6}$  S  $m^{-1}$ ).

From this magnetic field, we build a polarized signal following the monochromatic definition of polarization attributes (Fowler *et al.* 1967). For each event, we set  $\Delta\phi$  the phase difference between horizontal components and the relative amplitude  $A_x$  and  $A_y$  of both





**Figure C1.** Synthetic slow tail with parameters  $e = 0.262$ ,  $\theta = 0.826$ ,  $\Delta\phi = 150.5$ .

orthogonal components. The ellipticity  $e$  and the polarization angle  $\theta$  are then given by

$$\sin(2 \arctan(e)) = \frac{2A_x A_y}{A_x^2 + A_y^2} \sin(\Delta\phi) \quad (\text{C7})$$

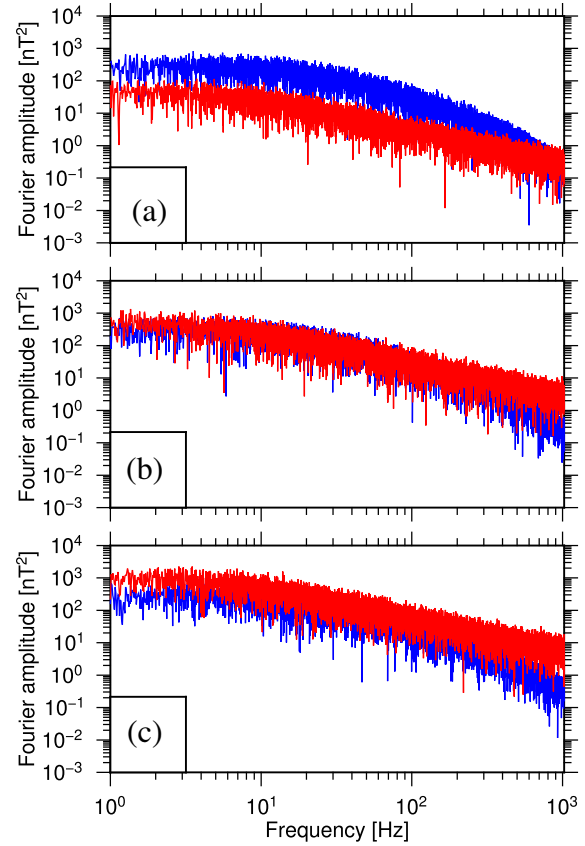
$$\tan(2\theta) = \frac{2A_x A_y}{A_x^2 - A_y^2} \cos(\Delta\phi). \quad (\text{C8})$$

In Fig. C1, we illustrate an example of a polarized synthetic slow tail. Each event was then randomly assigned to a time location along the time-series.

Finally, we add order 1 auto regressive noise to the time-series using the following model:

$$x_{i+1} = c + \alpha x_i + \sigma \varepsilon_i, \quad (\text{C9})$$

with  $\alpha$  the parameter of the AR-1 model,  $\varepsilon_i$  a standard white noise model,  $\sigma$  the standard deviation applied to the white noise to increase its amount in the modelled time-series and  $c$  a constant.  $c$  is set to zero in this experiment.



**Figure C2.** Power spectra for synthetic time-series. Red: added noise. Blue: signal. (a) Low noise level. (b) Moderate noise level. (c) Large noise level.

To create the associated remote time-series, we assume that the wave is quasi-uniform on the area where both stations are set. The polarization attributes and the time index of synthetic events are then the same in both local and remote time-series. Only the added noise is different in the two sets of time-series. Fig. C2 shows the three noise levels used in this paper.