





Local polynomial approximation and intersection of confidence intervals for removing noise of lightning electric field measurements

Herbert Enrique Rojas-Cubides^a, Camilo Andrés Cortés-Guerrero^b & María Carolina Forero-Mejía^c

^a Department of Electrical Engineering, Universidad Distrital Francisco José de Caldas, Bogotá DC, Colombia. herojasc@udistrital.edu.co ^b Department of Electrical and Electronic Engineering, Universidad Nacional de Colombia, Bogotá DC, Colombia. caacortesgu@unal.edu.co ^c ENEL Colombia., Bogotá DC, Colombia. maria.forero@enel.com

Received: April 20th, 2017. Received in revised form: May 3th, 2018. Accepted: May 30th, 2018.

Abstract

Lightning electric field (LEF) measurements are aperiodic signals characterized by inherent noise of different sources, i.e., it is not possible to register a noise-free signal. In the last decade, the denoising of LEF measurements has been achieved using some time-frequency representations such as short-time Fourier transform (STFT), wavelet transform (WT) and fractional Fourier transform (FRFT) without definitive results. In this paper, a denoising process applied on LEF measurements using the Local Polynomial Approximation (LPA) is proposed. The window size selection is made by combining the LPA with the intersection of confidence intervals (ICI) algorithm. Furthermore, a cross-validation criterion is used to select the optimal value of the threshold parameter in the LPA-ICI denoising method. It is shown that for different signal-to-noise ratio (SNR) values, the proposed method significantly reduces the noise present in the recorded signals. Finally, a discussion about the processed signatures in terms of some lightning electric field temporal features is performed.

Keywords: electric field; intersection of confidence intervals; lightning flash; local polynomial approximation; noise reduction.

Aproximación local polinomial e intersección de intervalos de confianza para la remoción de ruido en mediciones de campo eléctrico generado por rayos

Resumen

Las mediciones de campo eléctrico producido por rayos (LEF) son señales aperiódicas caracterizadas por ruido inherente de diferentes fuentes, es decir, no es posible registrar una señal libre de ruido. En la última década, se ha logrado filtrar mediciones de LEF utilizando algunas representaciones tiempo-frecuencia tales como la transformación de Fourier de corto tiempo (STFT), la transformación wavelet (WT) y la transformación de Fourier fraccionaria (FRFT) sin resultados definitivos. En este trabajo, se propone un proceso de eliminación de ruido aplicado a mediciones de LEF usando la aproximación polinomial local (LPA). La selección del tamaño de la ventana se realiza combinando la LPA con el algoritmo de intersección de intervalos de confianza (ICI). Además, un criterio de validación cruzada se utiliza para seleccionar el valor óptimo del umbral en el método de reducción de ruido LPA-ICI. Se muestra que para diferentes valores de relación señal-a-ruido (SNR), el método propuesto reduce significativamente el ruido presente en las señales registradas. Finalmente, se realiza una discusión sobre las señales procesadas en términos de algunas características temporales del campo eléctrico.

Palabras clave: aproximación local polinomial; campo eléctrico; intersección de intervalos de confianza; rayos; reducción de ruido.

1. Introduction

The measurement of electromagnetic fields generated by lightning has been carried out in the last four decades and consequently many features of the phenomenon have been revealed. A major part of these works was conducted using sensors to measure the radiated electric or magnetic fields generated by cloud-to-ground (CG) lightning flashes [1].

The knowledge of the electric fields produced by lightning (LEF) is important to: (a) identify some parameters

© The author; licensee Universidad Nacional de Colombia. Revista DYNA, 85(205), pp. 264-273, June, 2018, ISSN 0012-7353 DOI: https://doi.org/10.15446/dyna.v85n205.64151

How to cite: Rojas-Cubides, H.E., Cortés-Guerrero, C.A. and Forero-Mejía, M.C., Local polynomial approximation and intersection of confidence intervals for removing noise of lightning electric field measurements, DYNA, 85(205), pp. 264-273, June, 2018.

of lightning flashes (multiplicity, flash duration, interstroke intervals and currents); (b) understand the discharge mechanism and physical discharge processes; (c) design lightning protection systems; (d) lightning location networks; (e) determine the electromagnetic compatibility requirements of equipment and systems [1-4].

Due its importance, lightning characterization using LEF measurements has been developed in several regions of the world. In fact, lightning location networks, detection and warning systems, lightning flash density estimations and indirect determination of lightning current amplitudes are successful applications of indirect measurements of LEF. In addition, the measurement of these fields allows establishing relationships between lightning flash events and transients that could help in the classification of electromagnetic disturbances [5].

To develop and evaluate models of the lightning return stroke and determine the lightning current features, it is necessary to determine, in a reliable way, various temporal features of the electric field waveforms such as: the peak value, rise-time, zero-crossing time, slow-front and fast-front duration, and peak value of the electric field derivative [6–8] [6–8]. However, the extraction of lightning parameters from LEF or current measurements is not a simple task due to the presence of noise in the recorded signals, mainly caused by the measuring system itself and other undesired components present in the electromagnetic environment [7-9].

In the past, noise removal from LEF signals has been achieved using low-pass filters (hardware) and digital filters based on the discrete Fourier transform (software), both methods with cut-off frequencies ranging from 200 kHz to 30 MHz [9,10]. However, these methods may affect part of the original signal, causing loss of information. To overcome this problem, some denoising techniques based on time, frequency or hybrid time-frequency domains have been proposed in the last two decades.

The conventional Fourier transform (FT) and the shorttime Fourier transform (STFT) were the first approaches to reduce the noise on lightning electric field signatures [11]. Because of the limitations of FT and STFT (loss of resolution in time or frequency), new signal processing methods have been proposed. Wavelet transform (WT) has been used for different purposes in the study of lightning activity, including noise reduction and frequency anlaysis [3,8,12]. The major disadvantages of WT are the complexity of calculations, the sensitivity to some noise levels, and the dependence of its accuracy on the selected wavelet base (mother) function [8,12].

Recently, noise reduction methods on LEF signals have been extended to the application of the discrete fractional Fourier transform (DFRFT) [13,14]. However, it is not possible to have a desired signal because, by nature, the LEF have a random behavior and the waveforms change with respect to the distance from the lightning strike point. This disadvantage makes it difficult to minimize the error between the filtered and the original signal for any adaptive denoising process.

The local polynomial approximation (LPA) is a method proposed and developed in Statistics for processing scalar and multidimensional noisy data [15]. However, several applications have shown that LPA is a powerful technique to deal with linear and non-linear problems [16], reduction of leakage errors [17], nonparametric modelling [18] and filter design [19]. In this way, this paper presents a method for denoising lightning electric field measurements based on the LPA application. Furthermore, to solve the problem of window size selection, the LPA and the intersection of confidence intervals (ICI) algorithm are combined.

This work continues with the research project launched by the authors oriented to develop signal processing tools to analyze LEF signals. Considering that the essential features of lightning flashes are affected by geographical and environmental conditions, Bogotá-Colombia is an interesting place due to its tropical location, its high altitude (2550 masl), and its high lightning activity, which allows a continuous study of lightning [20].

2. Local polynomial approximation

2.1. Local approximation theory

The basic signal model considered in this paper assumes that the measured signal z(t), defined in eq. (1), is composed by N observations of a noise-free signal y(t) with a sampling time T_S .

$$z(t_k) = y(t_k) + \eta(t_k) t_k = kT_S, \quad k = 1, 2, ..., N$$
⁽¹⁾

In eq. (1), $\eta(t_k)$ are independent additive noise components with $E[\eta(t_k)] = 0$, $E[\eta^2(t_k)] = \sigma^2$ and $E[\vartheta]$ is the expected value of ϑ .

From this signal model, the local approximation theory involves two general problems. The first of these arises when it is necessary to find a polynomial function that can be used to determine the approximate values of a function $y(t_k)$. The second problem is related to the use of fitting functions on a given data sequence and determine the best function to represent those data [17,21]. The first problem is associated with the parametric estimation, while the second one is linked with the non-parametric estimation.

2.3. Filtering in the LPA domain

The LPA is a locally adaptive non-parametric technique, which provides estimates of the signal $y(t) \approx y(t_k)$ defined in eq. (1) from the observations of z(t) in a pointwise manner and using a polynomial fitting in a sliding window. In this estimation, the error should be as small as possible [22].

In terms of the signal processing, LPA is a flexible tool to design linear transforms with respect to polynomial components of signals. For this reason, to obtain an estimation of y(t), the following criterion must be applied on the linear LPA [22,23]:

$$J_{h}(t_{0},C) = \sum_{k=1}^{N} \rho_{h}(t_{k}-t_{0}) \left(z(t_{k}) - C^{T} \phi(t_{k}-t_{0}) \right)^{2}$$

$$\phi(x) = [1, x, x^{2}/2, ..., x^{m-1}/(m-1)!]^{T}$$

$$C = (C_{0}, C_{1}, C_{2}, ..., C_{m-1})^{T}$$
(2)

where t_0 is the point of interest (center), *C* is a vector with the coefficients of the polynomial used during the estimation of y(t) and *m* is the order of the LPA. The window function $\rho_h(x)$, which formalizes the location of fitting with respect to the center point *x*, is a finite support function and must satisfy the conventional properties of the "kernel" used in non-parametric estimation, in particular: $\rho(x) \ge 0, \rho(0) = \max_x \rho(x), \rho(x) \to 0$ as $|x| \to \infty$ and $\int_{-\infty}^{\infty} \rho(u) du = 1$.

Now, if the function $J_h(t_0, C)$ (eq. (2)) is minimized with respect to C:

$$\hat{\mathcal{C}}(t_0, h) = \arg\min_{\mathcal{C} \in \mathbb{R}^m} J_h(t_0, \mathcal{C})$$
(3)

the coefficient $\hat{C}_0(t_0, h) = \hat{y}(t_0, h)$ is an estimate of $y(t_0)$ with respect to a window function of bandwidth h, while $\hat{C}_l(t_0, h) = \hat{y}_l(t_0, h)$ are related with the estimates of the derivatives $y^{(l)}(t_0)$ with l = 1, 2, ..., m - 1. In this way, the estimates $\hat{C}_l(t_0, h)$ can be represented in the form of a linear transform as follows [15,23]:

$$\hat{y}_{h}(t) = \sum_{k} g_{h}(t_{k}, t_{0}) \, z(t_{k}) \tag{4}$$

where g_h is the kernel of the estimator defined by [23]:

$$g_{h}(t_{k}, t_{0}) = \rho_{h}(t_{k} - t_{0}) \left[\phi_{h}(t_{k} - t_{0})\right]^{T} \Phi_{h}^{-1} \phi_{h}(0)$$

$$\Phi_{h} = \sum_{k} \rho_{h}(t_{k} - t_{0}) \left[\phi_{h}(t_{k} - t_{0})\right] \left[\phi_{h}(t_{k} - t_{0})\right]^{T}$$
(5)

Here, ϕ_h is a column vector of polynomials of the LPA whose length is equal to m + 1, and Φ_h is a symmetric matrix. It is necessary to emphasize that, for a parametric estimation, the order of the polynomial model defines the type of estimation curve (constant, linear, quadratic, etc.) and the coefficients *C* are fixed. On the other hand, in the nonparametric process, the estimate of y(t) is not a polynomial function and the sliding window estimation makes the coefficients $\hat{C}_l(t, h)$ different for each point of interest t_0 .

2.3. Influence of the window function

Mathematically, the window function used in the LPA is expressed by a scale parameter h > 0 that determines the window size or bandwidth. For this reason, the window function in eq. (2) and eq. (5) must be replaced by:

$$\rho_h(x) = \rho(x/h)/h \tag{6}$$

Smaller or larger values of *h* narrow or widen the function $\rho_h(x)$, respectively. The scale factor (1/h) in eq. (6) normalizes the window and the property $\int_{-\infty}^{\infty} [\rho(x/h)/h] dx = 1$ is satisfied. Thus, the polynomial approximation function ϕ_h also needs to be scaled by the parameter *h*:

$$\phi_h(x) = \phi_h(x/h) \tag{7}$$

From eq. (4) to eq. (7), it is possible to observe that the length of the window function h controls the smoothness of

the estimate $\hat{y}_h(t)$ [22,23]. If *h* is large, the difference between parametric and non-parametric approaches disappears and the estimation function is constant, linear, quadratic, etc., depending on the polynomial degree of the LPA. In this case, the smoothing of random noise in observations is the maximum. On the other hand, when *h* is relatively small, the estimate $\hat{y}_h(t)$ is near (or exactly the same) to z(t) and there is no smoothing of the data. For intermediate values of the windows size, the resulting curves demonstrate different levels of smoothing.

In eq. (2), if the window is a rectangular function, all observations have equal weights. Nonrectangular window functions such as Gaussian, Hamming, Welch, Hann, etc., usually provide higher weights for observations that are closer to the center point t_0 . In addition, there are three types of windows (with respect to the center point) that LPA can use: central, right and left windows.

3. Window size selection and the ICI algorithm

3.1. The ICI algorithm and the bandwidth selection

The intersection of confidence intervals (ICI) is a bandwidth selection algorithm that has been proposed and substantiated in several publications [19,22,24-26]. This approach does not require the bias estimation and presents an accuracy approximation better than the quality-of-fit statistics [19]. LPA combined with the ICI algorithm improves the windows size selection process and enables the algorithm to be adaptive. The basic idea of the ICI algorithm uses a finite set of bandwidths (*H*) expressed by:

$$H = \{h_1 < h_2 < \dots < h_M\}$$
(8)

Then, for each sample (t), the LPA-ICI algorithm introduces a sequence of M estimates accompanied with the following confidence intervals:

$$D_M(t) = [L_M(t), U_M(t)]$$
 (9)

The lower and upper limits of $D_M(t)$ are defined as:

$$L_M(t) = \hat{y}_{h_M}(t) - \Gamma \sigma_{h_M}(t)$$

$$U_M(t) = \hat{y}_{h_M}(t) + \Gamma \sigma_{h_M}(t)$$
(10)

where Γ is the threshold of the confidence interval and the standard deviation of the estimated signal $\hat{y}_{h_M}(t)$ is defined by [15]:

$$\sigma_{h_{M}}(t) = \sigma' \sqrt{\sum_{k} [g_{h_{M}}(t_{k}, t_{0})]^{2}}$$

$$\sigma' = \frac{\text{median} |z(t_{k}) - z(t_{k} - 1)|}{\sqrt{2} \cdot 0.6745}; \ k = 2, \dots, N$$
(11)

The ICI algorithm tracks the values of the largest lower \underline{U}_M and the smallest upper \overline{L}_M confidence intervals limits:

$$\overline{L}_M(t) = \max_{i=1,\dots,M} L_i(t)$$
(12)

$$\underline{U}_M(t) = \min_{i=1,\dots,M} U_i(t)$$

The chosen filter support (bandwidth) h^+ is the largest *i* support for which the following condition is still satisfied:

$$\overline{L}_i(t) \le \underline{U}_i(t) \tag{13}$$

It was shown in [15,22] that h^+ is close to the optimal support h^* , which results in the optimal trade-off between bias $(\omega_h(t))$ and variance $(\sigma_{h_M}^2(t))$. Besides, this solution minimizes the MSE [19].

The parameter Γ in eq. (10) plays an important role in the bandwidth selection. A large value of Γ gives $h^+ > h^*$, producing an over-smoothing in the signal. On the other hand, smaller values of Γ give $h^+ < h^*$, cause an under smoothing on the signal.

3.2. Threshold parameter Γ

To find the optimal value of the threshold parameter Γ from an theoretical analysis presents great difficulties [15]. For this reason, to compare the performance of the ICI algorithm with respect to Γ , a cross-validation (CV) criteria was selected to assess the statistical prediction. This validation technique has been used to compare the performance of different predictive models [27] and, from the experiments presented in [15,19,28], it can be used as a reasonable and efficient selector of Γ .

For the linear filter defined in eq. (4), the CV function with an estimator support h^+ is represented as a weighted sum of the squared residuals [15]:

$$I_{CV} = \sum_{k} \left(\frac{z(t_k) - \hat{y}_{h^+}(t_k)}{1 - g_{h^+}(t_k, t)} \right)^2$$
(14)

It is important to mention that the process explained in this section must be repeated for every value of $\Gamma \in R$, $R = {\Gamma_1, \Gamma_2, ..., \Gamma_R}$. In addition, the value of the threshold parameter used in the "best" estimation is defined for the following expression:

$$\widehat{\Gamma} = \arg\min_{\Gamma \in R} I_{CV} \tag{15}$$

4. The LPA-ICI denoising method

The main purpose of denoising LEF signatures is to reduce the noise components present in measurements without modifying the most important signal features. Although it is possible to carry out actions to reduce the emissions produced by the measuring system itself, all recorded signals are also distorted by undesired signals from the electromagnetic environment, whose sources cannot be controlled or eliminated.

For these reasons, the method based on LPA combined with the ICI algorithm for noise reduction on LEF signals consists of the following steps:

• Set
$$\Gamma = \Gamma_R$$
, for $R = \{\Gamma_1, \Gamma_2, \dots, \Gamma_R\}$

- Determine the sampling time (T_S) from the noisy signal observations z(t_k) and select a set of samples t_k = kT_S with k = 1,2,..., N
- For $H = h_i$, with i = 1, 2, ..., M, calculate the estimates $\hat{y}_h(t)$ and the confidence intervals defined by eq. (12)
- Select the largest of those *i* for which $\overline{L}_M(t) \leq \underline{U}_M(t)$ gives the bandwidth $i^+ = h^+$. From this result determine the adaptive window size $h^+(t)$ and the estimate $\hat{y}_{h^+}(t)$.
- Repeat the previous step for all t_k and each Γ_R
- Find $\hat{\Gamma}$ from eq. (15) and select the final estimation of $\hat{y}_{h+}(t_k)$ corresponding to the $\hat{\Gamma}$.

In all cases, the cross-validation presented in eq. (14) and the maximum SNR of the signal (before and after noise reduction) were selected as appropriate indicators for comparing the obtained results. In this paper, the SNR is expressed as the relation between the power of the filtered signal (P_{signal}) and the power of noise removed from the measured signal (P_{noise}), so:

$$SNR_{dB} = 10 \cdot log_{10} \left(\frac{P_{signal}}{P_{noise}}\right)$$
(16)

5. Instrumentation and data

5.1. Typical waveforms of LEF signals

CG lightning flashes are transient discharges of high current that can be generated in the atmosphere by clouds, storms (rain, snow or dust) and volcanic eruptions. A typical CG lightning flash is composed by more than one discharge known as return stroke. These individual discharges have a duration of tens of microseconds and they are usually separated in time from 20 ms to 100 ms [29]. The typical waveforms for the electric fields produced by the first return stroke (FRS) and the subsequent return strokes (SRS) that compose a flash are shown in Fig. 1.

The most relevant characteristics of these waveforms include: (a) an initial peak with an acute shape (due to a high rate of growth) that decreases approximately with the inverse of the distance; (b) a slow descent ramp after the initial peak that may last more than 100 μ s in signals recorded in a nearby range (few tens of kilometers); (c) a zero crossing point that may be several tens of microseconds after the initial peak for the fields registered at distances greater than 50 km; and (d) an opposite-polarity overshoot typical of the electric fields recorded at distances greater than 100 km.

5.2. Measuring system

The typical system to measure LEF was proposed by Cooray and Lundquist in 1982 [30]. In this configuration, an electric field sensor is connected to an electronic circuit for the acquisition of the signal and it is used an equipment to record the data. This arrangement has been used during the last three decades in several regions (Sweden, Germany, Italy, Japan, Sri Lanka, Malaysia, Singapore, USA, Colombia and Brazil), providing valuable information about the behavior and characteristics of lightning flashes [10,31,32].

The scheme of the measuring station used in this work to record the LEF signals is shown in Fig. 2, and it is composed by the following parts:



Figure 1. Typical waveforms of the electric field signals produced by CG lightnings with respect to different distance ranges. First return stroke (solid lines), subsequent return stroke (dashed lines). Source: Adapted from [29]



Figure 2. Lightning electric field measuring system. (A) parallel-plate antenna, (B) electronic circuit, (C) oscilloscope, (D) coaxial cables. Source: The authors

- The electric field sensor is a 1.5 meters height parallel flat-plate antenna, which had two circular metallic plates with diameter of 0.45 m supported by insulating elements and 0.03 m air gap between them.
- The electronic circuit is based on the buffer-amplifier BUF-602 with 1000 MHz bandwidth and 8 kV/µs slew rate. The decaying time constant of the circuit is 38 ms, which is large enough to record signatures composed by electrostatic (near lightning electric fields) and radiated components (far lightning electric fields).
- Two coaxial cables were used to connect the devices. First, a 0.5 m length RG58/U cable was used to connect the antenna to the electronic circuit. Second, a 12 m

length RG58/U coaxial cable was used between the electronic circuit and the acquisition equipment.

• The acquisition equipment was an 8-bits, digital oscilloscope Agilent DSO6104A with a sampling rate of 10 MSa/s (sampling time of 100 ns).

The arrangement composed by the antenna, the short coaxial cable and the electronic circuit works as a passband filter with a cut-off frequency of 11.8 MHz, which is enough to study lightnings because this range of frequencies covers the bandwidth in which most of the spectrum of the return strokes is found (some hundreds of kHz) [3,12]. A complete description of the measuring system and the electronic circuit, including its calibration process, can be reviewed in [33].

During measurements, the digital oscilloscope was adjusted using a full observation window of 500 ms. This time was selected in order to record a complete CG lightning flash signature that includes the electric field waveforms produced by the first stroke and the subsequent strokes. The vertical trigger level was set at 100 mV to reduce the interference caused by the noise inherent to the measuring system (92 mV average) and to avoid the oscilloscope drive due to pulses caused by intra-cloud lightnings. In order to acquire signals before and after the trigger transient pulse, a 75 ms horizontal pre-trigger was adjusted. In addition, to reduce additional noise that could be produced at the output of the buffer-amplifier, the bandwidth of the oscilloscope was configured in 25 MHz.

Finally, during the pre-processing stage, each return stroke signal (first and subsequents) was extracted from the complete lightning flash signature and it was adjusted for an observation window of 400 μ s. Under this condition, using a sampling time of 100 ns, each return stroke signal analyzed in the following sections has 4000 samples.

5.3. Reference signatures

The measurements presented in this section were extracted from the records obtained during August-November 2016. The measuring system was installed in Bogotá, Colombia (4.641° N, 74.091° W) at an altitude of 2550 m above sea level. All the signatures analyzed belong to negative CG lightning flashes. Examples of the recorded events are shown in Fig. 3 and Fig. 4. These signals correspond to normalized signatures of two return strokes (FRS and SRS) and they agree with the waveforms shown in Fig. 1 (distance range between 50 and 200 km). The vertical axis in each plot represents the electric field magnitude in V/m whereas the horizontal axis represents the time in microseconds.

Reviewing Fig. 3 and Fig. 4 it can be seen the presence of noise components that makes it difficult to analyze and characterize the signal features. This is relevant especially in the estimation of temporal lightning parameters, such as: maximum value, signal rise-time (10–90%), zero-crossing time and the maximum electric field derivative, which is an important parameter to estimate the maximum variation of lightning current.



Figure 3. Electric field signature of a first return stroke (Bog3_st1) Source: The authors



Figure 4. Electric field signature of a subsequent return stroke (Bog3_st5) Source: The authors

6. Signal processing results

In this section, simulation results of the LPA-ICI denoising method over LEF measurements, the role of the threshold parameter in the ICI algorithm, and the performance of adaptive bandwidths are presented. In addition, the effects of different parameters of the LPA-ICI algorithm are analyzed.

6.1. Simulation parameters

To get a comparative idea about the temporal features and the magnitude of lightning electric field signatures, a set of six return stroke signatures (3 FRS and 3 SRS) were selected. These signals were extracted and normalized from different lightning flashes. In order to analyze the role of the threshold parameter, several values of $\Gamma = \{0.01: 0.01: 4\}$ are defined.

Simulations were performed using a symmetric Gaussian window $[\rho_L(u) = \rho_L(-u)]$. This function was selected because it is well known, from numerical analysis, that symmetric windows result in smaller approximation errors than asymmetric windows [21,34]. The adaptive bandwidths (sample size) for the window function are given by $h = \{3, 5, 6, 9, 12, 18, 25, 35, 50, 71, 100, 143, 203, 290\}$.

In order to select an adequate polynomial order for LPA, a linear algorithm was employed by increasing the order of LPA from m = 2 to m = 9. From simulations, it was observed that polynomial orders less than m = 4 provide results with ripples and remarkable noise components, while high polynomial orders $(m \ge 4)$ present similar results but increase the computational costs. In this way, a polynomial order of m = 5 for the LPA-ICI algorithm was selected for

all the cases. Signal processing results were obtained from routines developed in MATLAB® by the authors.

6.2. High SNR environment – first return strokes

For the example given in Fig. 5(a), the best estimate using a symmetric Gaussian window is presented in Fig. 5(b). In this case, the graphical results illustrate a substantial improvement in the signal features when $\Gamma = 1.23$. This result is reinforced by the behavior of the threshold parameter with respect to the CV criterion shown in Fig. 5(c).

In addition, the adaptive bandwidths for the adjusted Γ are shown in Fig. 5(d). It can be seen from this curve that at the peak value of y(t) and at the points where the signal slope



Figure 5. Results for FRS signature (Bog3_st1). (a) measured signal; (b) LPA-ICI denoising signal for symmetric Gaussian window with $\Gamma = 1.23$; (c) CV criterion for LPA-ICI with m = 5; (d) adaptive bandwidth variation. Source: The authors

changes, the adaptive bandwidths decrease. This means that the LPA-ICI method is sensitive with respect to fast variations of the signal and the symmetrical window presents a good filter response.

6.3. Low SNR environment – subsequent return strokes

Fig. 6(a) and Fig. 6(b) show the measured signal and the filtered signal of a subsequent return stroke using the LPA-ICI denoising method with polynomial order m = 5. In this case, the denoising result is better because the subsequent stroke signal has a low SNR.

The cross-validation criterion as a function of the threshold parameter is illustrated in Fig. 6(c). For the subsequent stroke



Figure 6. Results for SRS signature (Bog3_st5). (a) measured signal; (b) LPA-ICI denoising signal for symmetric Gaussian window with $\Gamma = 1.38$; (c) CV criterion for LPA-ICI with m = 5; (d) adaptive bandwidth variation. Source: The authors



Figure 7. LPA-ICI estimates for electric field signature of a subsequent return stroke (Bog3_st5) to $\Gamma = \{0.1: 0.25: 4.1\}$ Source: The authors

signal, the minimal value of CV is 4.505×10^{-4} for the adjusted value $\Gamma = 1.38$. Fig. 6(d) shows the adaptive varying bandwidths for the symmetric window filter. These bandwidths present faster changes in time, which means that the LPA approximation is enough for denoising the measured signal using the size of windows that are available.

It is important to note that, in this case, the threshold parameter changed from $\Gamma = 1.23$ (FRS) to $\Gamma = 1.38$ (SRS), which demonstrates the adaptive response of the ICI algorithm. An advantage of the LPA-ICI denoising method is that the best estimate does not present high noise oscillations and ripples from the measured signal that certainly represent difficulties in the interpretation of lightning parameters.

From the best estimate obtained for the SRS signal it was noticed that the adaptive window size, especially for a small value of Γ , could be corrupted by spikes and ripples that erroneously isolate small values of the window sizes. Fig. 7 shows the estimate when the threshold parameter is adjusted from the initial conditions (see section 7.1) to $\Gamma = \{0.1: 0.25: 4.1\}$.

Note that when the change in the step-size of Γ is made, some ripples in the initial stage of the signal rise and the spikes on the crossing-to-zero region (between $t = 1.2 \times 10^{-4} s$ and $t = 3 \times 10^{-4} s$) are slightly reduced. In addition, the threshold parameter of the signal estimate changes from 1.38 to 1.35 with a larger step-size.

7. Discussion

In this work, the parameters selected to analyze LEF are 10–90% signal rise-time (Tr), zero crossing time (ZC) and maximum electric field (Ep). In addition, it is important to know the maximum electric field derivative (∂ Ep/dt), which is a useful parameter in the estimation of the maximum variation of lightning current. Finally, the SNR improvements after the denoising process are included.

Besides the results obtained from a visual comparison, Table 1 presents the results obtained from characterizing the parameters (Tr, ZC, Ep, ∂ Ep/dt and SNR) of the signatures before and after noise removal. In addition, the right column in Table 1 (% DIF) presents the percentage difference (for each parameter) between the measured signal and the estimation obtained with the LPA-ICI (processed signal). The maximum values of the electric field and the electric field derivative were normalized taking the maximum of the respective flash as the reference value.

Table 1.
Parameters of the LEF signatures before and after the denoising process.

Signature		Paramotor	Signal		% DIF	
		1 al alletel	ME	PR	70 DIF	
		Tr (µs)	8.4	6.4	23.8%	
Bog1_st1	ZC (µs)	78.2	78.9	0.9%		
	Ep (V/m)	1	0.987	1.3%		
		∂Ep/dt (V/m/µs)	1	0.66	34.0%	
		SNR (dB)	17.9	30.6	70.9 % *	
FRS Bog3_st1		Tr (µs)	11.4	10.4	8.8%	
	ZC (µs)	43.2	42.3	2.1%		
	Ep (V/m)	1	0.989	1.1%		
	(11g. 5)	$\partial Ep/dt (V/m/\mu s)$	1	0.588	41.2 %	
		SNR (dB)	15.6	33.8	116.7 % *	
Bog4_st1		Tr (µs)	12.4	13	4.8%	
	ZC (µs)	65.2	66.4	1.8%		
	Bog4_st1	Ep (V/m)	1	0.988	1.2%	
		∂Ep/dt (V/m/µs)	1	0.635	36.5%	
		SNR (dB)	20.5	31.1	51.7% *	
		Tr (µs)	7.8	7.6	2.6%	
		ZC (µs)	47.5	51.3	8.0%	
	Bog1_st4	Ep (V/m)	0.5	0.491	1.9%	
		∂Ep/dt (V/m/µs)	1	0.535	46.5%	
		SNR (dB)	11.7	23.3	99.1% *	
		Tr (µs)	8.6	7.1	17.4%	
SRS _		ZC (µs)	73.2	79.8	9.0%	
	Bog2_st3	Ep (V/m)	0.375	0.341	9.1%	
		$\partial Ep/dt (V/m/\mu s)$ 1		0.452	54.8%	
		SNR (dB)	9.4	22.52	139.6% *	
		Tr (µs)	5.8	4.1	29.3%	
	Dog2 at5	ZC (µs)	40.6	41.6	2.5%	
	(Fig. 6)	Ep (V/m)	0.262	0.236	9.9%	
	(11g. 0)	$\partial Ep/dt (V/m/\mu s)$	1	0.306	69.4%	
		SNR (dB)	6.5	19.8	204.6% *	

ME: measured signal --- PR: processed signal --- % DIF: percentage difference between measured signal (ME) and processed signal (PR) * Improved

Source: The authors

From the results presented in Table 1, it can be observed that the rise-time presents a difference up to 23.8% and 29.3% in the FRS and SRS cases, respectively. These differences may produce inadequate interpretation of lightning parameters. For the zero-crossing time, the differences were less than 3% for high SNR environments (FRS), while for the low SNR signals (SRS) a maximum difference of about 9% was found (signature Bog2_st3).

On the other hand, the maximum normalized electric field (Ep) shows a variation up to 2% for the FRS signatures, whereas for the SRS cases the variation in this parameter reached 10%. These results indicate that the reduction in the peak value of the signatures depends on the signal noise level. However, it is important to clarify that the removal of noise is directly affected by the selected filtering method.

Regarding the maximum electric field derivative ($\partial Ep/\partial t$), remarkable differences were observed between the measured signatures and the filtered ones. The minimum difference was 34% for Bog1_st1 signature (FRS case) and the maximum difference was 70% for Bog3_st5 signature

(SRS case). These results show that the estimation of electric field derivative from the measured signal could be erroneous due to the presence of high frequency noise.

By comparing the SNR of the measured signal with the best result obtained with the LPA-ICI method, it is possible to notice that for the signal Bog3_st1 (shown in Fig. 3 and Fig. 5) the change in SNR ranged from 15.6 dB to 33.8 dB after noise removal. For the signal Bog3_st5 (shown in Fig. 4 and Fig. 6), the SNR exhibits a change from 6.5 dB to 19.8 dB. In general terms, for the whole group of signatures, the improvement in the SNR varied from 51% up to 117% for the FRS cases, while for the SRS cases this parameter was improved between 99% and 205%.

In order to verify the usefulness of LPA-ICI denoising method, a comparison with an alternative method using adaptive filters in the discrete fractional Fourier domain (DFRFd) was conducted. This filtering process is described with detail in [14]. In this work, the adaptive filter in DFRFd is configured with the following parameters: fractional order a = 0.25; normalized leakage LMS adaptive (NL-LMS) algorithm; normalized step-size $\mu_{NL-LMS} = 0.95$; leakage factor $\gamma = 0.997$; stabilization factor $\beta = 2 * 10^{-14}$; number of coefficients is 6. Fig. 8 shows the comparison between the signatures filtered with the LPA-ICI method (black dotted line) and using adaptive filters in DFRFd (red continuous line).

In this case, it is possible to observed that the LPA-ICI provides similar results than those obtained with the technique based on DFRFT. However, a slight reduction is observed in the peak value of the FRS signature filtered with LPA-ICI algorithm (see Fig. 8a). In addition, adaptive filters in DFRFd introduce oscillations, especially in the SRS signature (see Fig. 8b). These oscillations are typical of Fourier rotated methods, which produce difficulties in the interpretation of some temporal parameters.



Figure 8. Comparison between LPA-ICI and adaptive filters in DFRFd. (a) FRS signature (Bog3_st1); (b) SRS signature (Bog3_st5) Source: The authors

Table 2. Parameters comparison of LEF signatures (Bog3_st1 and Bog3_st5) filtered with LPA-ICI method and adaptive filters in DEREd

Signature		Tr	CZ	Ер	∂Ep/dt	SNR
		(µs)	(µs)	(V/m)	(V/m/µs)	(dB)
D = -2 -+1	Measured	11.4	43.2	1	1	15.6
$(Eig S_0)$	LPA-ICI	10.4	42.3	0.989	0.59	33.8
(Fig. 8a)	FRFT	10.3	42.1	0.991	0.60	31.4
Dog2 at5	Measured	5.8	40.6	0.262	1	6.5
(Eig Ob)	LPA-ICI	4.1	41.3	0.236	0.31	19.8
(Fig. 80)	FRFT	3.7	42	0.253	0.33	22.1

Source: The authors

Table 2 presents the comparison of parameters for signals presented in Fig. 8, which were obtained using both denoising methods. With respect to rise-time, the differences between denoising methods are less than 2% for the FRS signature (shown in Fig. 8a), while for the SRS signature (show in Fig. 8b) the difference reaches 10%. This result in the SRS case is also produced because of the oscillations present in the filtered signal using DFRFT, which do not allow to clearly identify the initial zero-crossing point of the processed signal.

For zero-crossing time and maximum electric field, the differences for both examples were less than 2% and 7%, respectively. The main difference (up to 9%) is presented in the estimation of the maximum value of the electric field derivative for SRS signature (Fig. 8b). This is due to the difference between the maximum value of the filtered signal using LPA-ICI and the output signal obtained with the method based on DFRFT.

8. Conclusions

In this paper, the local polynomial approximation (LPA) combined with the adaptive size window algorithm (ICI) is applied to achieve a successful noise removal process on lightning electric field (LEF) signatures. The advantage of the proposed method compared to other filter methods, as adaptive filters, is that the LPA-ICI does not require a preliminary desired signal that represents the phenomenon.

The proposed method for denoising LEF signals is simple to implement and requires only the calculation of the estimates and their standard deviations for a set of bandwidth values. It is important to note that the LPA-ICI algorithm does not require the estimation of the bias. In addition, the proposed algorithm could be considered an adaptive method because its performance depends on an adequate selection of the window size (bandwidths).

The obtained results show that the LPA-ICI method is efficient to work with non-stationary signals that have jumps, ripples and fast slope changes, such as LEF signatures. In fact, the use of adaptive bandwidths provides better estimates because the most suitable bandwidth controls the smoothness of the filtered signal. In the particular case of LEF signals, the results provided by the LPA-ICI algorithm were found using symmetric Gaussian windows. However, other kinds of symmetric window functions can be tested to analyze the response of the algorithm.

Simulation results, including the effect of window

bandwidth *h* and threshold parameter Γ , were analyzed for both high and low SNR environments. From the tests, it was observed that the change in the step-size of Γ improved the performance of LPA-ICI algorithm. In the low SNR cases, the use of a larger Γ reduces ripples and spikes on several zones of the output signal which are presented when the stepsize of Γ is small.

The performance of the proposed method was evaluated in terms of a temporal parameter comparison and the SNR improvement factor for electric field signals produced by first return strokes (FRS) and subsequent return strokes (SRS). Remarkable differences between the measured signal and the filtered signal were found in terms of the rise-time (up to 29.3%) and the maximum electric field derivative (up to 70%). The maximum improvement achieved with respect to SNR was 118% in the FRS cases and 205% in the SRS cases.

Acknowlegdements

Prof. H. E. Rojas would like to acknowledge Universidad Distrital Francisco José de Caldas for support this work through the postgraduate commission contract N° 00002-2016. Eng. Carlos Rivera M.Sc. and Eng. Johanna Chaves from the EMC-UNC Research Group are also acknowledged for their collaboration in the construction of the remote lightning measuring station (EMIR-UN).

References

- Sonnadara, U., Cooray, V. and Götschl, T., Characteristics of cloudto-ground lightning flashes over Sweden. Phys. Scr., 74(5), pp. 541-548, 2006. DOI: 10.1088/0031-8949/74/5/010.
- [2] Cooray, V., The mechanism of the lightning flash, in: The lightning flash, London: IEE Power Engineering Series, 2003, pp. 127-240.
- [3] Sharma, S.R., Cooray, V., Fernando, M. and Miranda, F.J., Temporal features of different lightning events revealed from wavelet transform. J. Atmos. Solar-Terrestrial Phys., 73(4), pp. 507-515, 2011. DOI: 10.1016/j.jastp.2010.11.009
- [4] Bennett, A.J., Identification and ranging of lightning flashes using colocated antennas of different geometry. Meas. Sci. Technol., 24(12), pp. 1-8, 2013. DOI: 10.1088/0957-0233/24/12/125801.
- [5] Peretto, L., Rinaldi, P., Sasdelli, R., Tinarelli, R. and Fioravanti, A., Implementation and characterization of a system for the evaluation of the starting instant of lightning-induced transients. IEEE Trans. Instrum. Meas., 56(5), pp. 1955-1960, 2007. DOI: 10.1109/TIM.2007.895669.
- [6] Hussein, A.M., Milewski, M. and Janischewskyj, W. Correlating the characteristics of the CN Tower lightning return-stroke current with those of its generated electromagnetic pulse. IEEE Trans. Electromagn. Compat., 50(3), pp. 642-650, 2008. DOI: 10.1109/TEMC.2008.924398.
- [7] Gou, X., Chen, M., Zhang, Y., Dong, W. and Qie, X., Wavelet multiresolution based multifractal analysis of electric fields by lightning return strokes. Atmos. Res., 91(2–4), pp. 410-415, 2009. DOI: 10.1016/j.atmosres.2008.04.012.
- [8] Santamaria, F., Cortés, C. and Román., F., Use of the Wavelet transform in noise reduction of the electric field signals produced by lightning. Inf. Tecnológica, 23(1), pp. 65-78, 2012. DOI: 10.4067/S0718-07642012000100008.
- [9] Nedjah, O., Hussein, A.M., Krishnan, S. and Sotudeh, R., Comparative study of adaptive techniques for denoising CN tower lightning current derivative signals. Digit. Signal Process, 20, pp. 607-618, 2010. DOI: 10.1016/j.dsp.2009.08.002.
- [10] Sonnadara, U., Cooray, V. and Fernando, M., The lightning radiation field spectra of cloud flashes in the interval from 20 kHz to 20 MHz.

IEEE Trans. Electromagn. Compat., 48(1), pp. 234-239, 2006. DOI: 10.1109/TEMC.2006.870692.

- [11] Islam M.J. and Hussein, A.M., Frequency domain approach to denoise the CN tower lightning derivative signal and its parameters calculations. J. Comput. Electr. Eng., 1(3), pp. 328-333, 2009. DOI: 10.7763/IJCEE.2009.V1.52.
- [12] Miranda, F.J., Wavelet analysis of lightning return stroke. J. Atmos. Solar-Terrestrial Phys., 70(11–12), pp. 1401-1407, 2008. DOI: 10.1016/j.jastp.2008.04.008.
- [13] Escobar, O.F., Rojas, H.E., Román, F. and Cortes, C.A., Lightning magnetic field measuring system in Bogotá — Colombia: signal processing method, Proceedings of 2013 International Symposium on Lightning Protection (XII SIPDA), 1, 2013, pp. 162-166. DOI: 10.1109/SIPDA.2013.6729210.
- [14] Rojas, H.E. and Cortés, C.A., Denoising of measured lightning electric field signals using adaptive filters in the fractional Fourier domain. Measurement, 55, pp. 616-626, 2014. DOI: 10.1016/j.measurement.2014.05.026.
- [15] Katkovnik, V., Egiazarian, K. and Astola, J., Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule. J. Math. Imaging Vis., 16, pp. 223-235, 2002. DOI: 10.1023/A:1020329726980.
- [16] Li, X., Bi, G., Stankovic, S. and Zoubir, A.M., Local polynomial Fourier transform: a review on recent developments and applications. Signal Processing, 91(6), pp. 1370-1393, 2011. DOI: 10.1016/j.sigpro.2010.09.003.
- [17] Schoukens, J., Vandersteen, G., Pintelon, R., Emedi, Z. and Rolain, Y., Bounding the polynomial approximation errors of frequency response functions. IEEE Trans. Instrum. Meas., 62(5), pp. 1346-1353, 2013. DOI: 10.1109/TIM.2012.2232451.
- [18] Agrež, D. and Lušin, T., Non-parametric estimation of the amplitude ratio of sinusoidal signals with common frequency. Measurement, 55, pp. 176-185, 2014. DOI: 10.1016/j.measurement. 2014.04.039.
- [19] Lerga, J., Vrankic, M. and Sucic, V., A signal denoising method based on the improved ICI rule. IEEE Signal Process. Lett., 15(8), pp. 601-604, 2008. DOI: 10.1109/LSP.2008.2001817.
- [20] Rojas, H.E., Santamaría, F., Escobar, O.F. and Román, F.J., Lightning research in Colombia: lightning parameters, protection systems, risk assessment and warning systems. Ing. y Desarro., 35(1), pp. 240-261, 2017. DOI: 10.14482/inde.35.1.8951.
- [21] Richard, B. and Faires, D., Numerical Analysis, 9th Ed. Boston, USA: Cengage Learning, 2010.
- [22] Katkovnik, V., A new method for varying adaptive bandwidth selection. IEEE Trans. Signal Process., 47(9), pp. 2567-2571, 1999. DOI: 10.1109/78.782208.
- [23] Katkovnik, V., Egiazarian, K. and Astola, J., Local approximation techniques in signal and image processing, 1st ed. Washington, USA: SPIE Publications, 2006.
- [24] Goldenshluger, A. and Nemirovski, A., On spatially adaptive estimation of nonparametric regression. Math. methods Stat., 6(2), pp. 135-170, 1997.
- [25] Tomic, M., Sersic, D. and Vrankic, M., Edge-preserving adaptive wavelet denoising using ICI rule. Electron. Lett., 44(11), pp. 698-699, 2008. DOI: 10.1049/el:20080227.
- [26] Song, Y., Han, Y. and Lee, S., Structure based noise reduction using LPA-ICI, Proceedings of 2012 IEEE International Conference on Consumer Electronics, 2012, pp. 498-499. DOI: 10.1109/ICCE.2012.6161992.
- [27] Stone, M., Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Ser. B, 36(2), pp. 111-147, 1974.
- [28] Katkovnik, V., Egiazarian, K. and Astola, J., A spatially adaptive nonparametric regression image deblurring. IEEE Trans. image Process., 14(10), pp. 1469-1478, 2005. DOI: 10.1109/TIP.2005.851705
- [29] Lin, Y.T., Uman, M.A., Tiller, J.A., Brantley, R.D., Beasley, W.H., Krider, E.P. and Weidman, C.D., Characterization of lightning return stroke electric and magnetic fields from simultaneous two-station measurements. J. Geophys. Res., 84(C10), p. 6307-6314, 1979. DOI: 10.1029/JC084iC10p06307.
- [30] Cooray, V. and Lundquist, S., On the characteristics of some radiation fields from lightning and their possible origin in positive ground

flashes. J. Geophys. Res., 87(C13), pp. 11203-11214, 1982. DOI: 10.1029/JC087iC13p11203.

- [31] Galvan, A. and Fernando, M., Operative characteristics of a parallelplate antenna to measure vertical electric fields from lightning flashes, 1st ed. Uppsala, Sweden: Uppsala University, 2000.
- [32] Salimi, B., Mehranzamir, K. and Abdul-Malek, Z., Statistical analysis of lightning electric field measured under Malaysian condition. Asia-Pacific J. Atmos. Sci., 50(2), pp. 133-137, 2014. DOI: 10.1007/s13143-014-0002-0.
- [33] Rojas, H.E., Rivera, C.A., Chaves, J., Cortés, C.A., Roman, F.J. and Fernando, M., New circuit for the measurement of lightning generated electric fields, Proceedings of 2017 International Symposium on Lightning Protection (XIV SIPDA), 2017, pp. 1-7. DOI: 10.1109/SIPDA.2017.8116922.
- [34] Ralston, A. and Rabinowitz, P., A first course in numerical analysis, 2nd Ed. New Jersey, USA: Dover Publications, 2003.

H.E. Rojas- Cubides, received the BSc. degree in Electrical Engineer (with honors), the MSc. degree in High Voltage and the PhD. degree from the Universidad Nacional de Colombia in 2003, 2007 and 2018, respectively. He was doctoral visiting student in 2016 at the Atmospheric and Lightning Physics Group, University of Colombo, Sri Lanka. During 2017, he was doctoral visiting student at the Institute of Energy Systems, Energy Efficiency and Energy Economics (IE3), Technical University of Dortmund, Germany. Since 2012, he has been with the Universidad Distrial Francisco José de Caldas, Bogotá, Colombia, where he is currently an Associate professor and researcher at the Electromagnetic Compatibility and Interference Group (GCEM-UD). His current research interests include signal processing, lightning measurements and characterization, high voltage techniques, grounding systems, power quality and electromagnetic compatibility.

ORCID: 0000-0003-1253-6964

C.A. Cortes- Guerrero, received the BSc. degree in Eng. from the Universidad Nacional de Colombia in 2000 and the PhD. degree (with honors) from the Universidad Nacional de San Juan, Argentina, in 2005. He was doctoral visiting student in 2002 at the FH Giessen-Friedberg and the NLÖ, Hannover, Germany. From 2005 to 2007, he was professor of the Universidad de la Salle, Bogotá, Colombia. In 2006, he was postdoctoral visiting scholar at the Katholieke Universiteit Leuven, Belgium. Between 2015 and 2016, he was a visiting researcher at the Galvin Center for Electricity Innovation, Illinois Institute of Technology, Chicago, IL, USA. He is an associate professor of the Universidad Nacional de Colombia since 2008, and a DAAD Research Ambassador since 2017. ORCID: 0000-0002-0986-3975

M.C. Forero- Mejía, received the BSc. degree Eng. (with honors) in Universidad Distrital Francisco José de Caldas, Bogotá, Colombia in 2016. From 2012 to 2014, she was undergraduate researcher at the GCEM-UD research group. Since 2015, she is with Grupo Enel–Colombia as junior auditor. Her research interests include distribution system analysis, energy management, signal processing and power quality disturbances. ORCID: 0000-0002-1312-6417