Electric Power Systems Research xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Electric Power Systems Research

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

ELSEVIER

Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines

John A. Morales [a,b,*], E. Orduña [a], C. Rehtanz [c], R.J. Cabral [d], A.S. Bretas [d]

[a] Instituto de Energía Eléctrica, Universidad Nacional de San Juan, Av. Lib. Gral. San Martín Oeste 1109, (J5400ARL), San Juan, Argentina
[b] Facultad de Ingeniería Eléctrica, Universidad Politécnica Salesiana, Calle vieja 12-30 y Elia Liut, 010150 Cuenca, Ecuador
[c] Institute of Energy Systems, Energy Efficiency and Energy Economics of TU Dortmund University, Emil-Figge-Straße 70, 44227 Dortmund, Germany
[d] Electrical Engineering Department, Federal University of Rio Grande do Sul, Av. Osvaldo Aranha, 103, Room 116, Porto Alegre, Rio Grande do Sul 90035-190, Brazil

ARTICLE INFO

Article history:
Received 7 February 2014
Received in revised form 23 May 2014
Accepted 27 May 2014
Available online xxx

Keywords:
Orthogonal functions
Lightning disturbances
Comparative evaluation
Filters

ABSTRACT

This paper presents an assessment between Principal Component Analysis (PCA) and Wavelet Transform (WT) signal processing techniques applied for Transmission Lines (TLs) lightning stroke classification. In this work, the atmospherics discharges signals are analyzed in two steps. The first step objective is patterns extraction, which is developed through Principal Component Analysis and the Wavelet Transform. The second step objective is pattern classification, which is developed using three different techniques: Artificial Neural Network (ANN), k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM).

This work presents as assessment of lightning stroke classification, providing useful information, especially in extraction and selection of mother functions and the use of PCA. Both methodologies are assessed under different lightning stroke conditions. Features as extraction, speed, orthogonal functions and others are comparatively assess.

Results show that by using PCA, optimal mother functions can be extracted, presenting a new alternative for relaying protection.

1. Introduction

Electric Power Systems (SEP) are daily exposed to disturbances. Events as faults produced by trees, birds, weather and others affect the system continuity operation [1–3]. However, between all possible disturbances, lightning is the severe, affecting specially transmission line protection relays [4–10].

Lightning effects on system devices can be divided in two: first when their overvoltages do not exceed the Basic Insulator Level (BIL), a fault is not produced and therefore the protection relay must not send a trip order. Second, when the generated overvoltage exceeds the Basic Insulator Level, a fault is produced and the relay must send the trip signal [11]. In this context, it is clear that protection devices must correctly classify these phenomena, but currently, traditional relays performance under lightning conditions still has room for improvement [12–14]. Therefore, it is crucial for relaying performance, correct lightning stroke classification.

Analyzing lightning induced overvoltages, it becomes clear that there is a difference between lightning stroke signals that generate or not faults, especially in their high frequency transient waveforms. However, the adequate extraction of their features depends of the Signal Processing Technique applied. An essential requirement in protection relays is to accurately classify different patterns corresponding to lightning stroke. For this reason, a very effective feature extraction is the most important step to improve the accuracy of lightning patterns classification.

Based on the above stated, for many years wavelets transforms have been widely used for the lightning stroke classification, especially because of their performance and applications [13,17]. On the other side [18,19], presented a new approach for the lightning stroke classification. This proposed approach uses Principal Component Analysis to extract different patterns, which are useful for signal classification.

* Corresponding author at: Instituto de Energía Eléctrica, Universidad Nacional de San Juan</org>, Av. Lib. Gral. San Martín Oeste 1109, J5400ARL San Juan, Argentina. Tel.: +54 264 4226444; fax: +54 264 4210299.Universidad Politécnica Salesiana, Calle Vieja y Elia Liut, Cuenca, Ecuador.

E-mail addresses: moralesg@ups.edu.ec, jmorales@iee.unsj.edu.ar (J.A. Morales).

http://dx.doi.org/10.1016/j.epsr.2014.05.018
0378-7796/

Since the progress and development of new samplers, computational efficiency, new microprocessors technologies and others [15,16], new Signal Processing Techniques can be applied to develop more dependable protection algorithms. Thus, currently, several Signal Processing Techniques can be considered for new protection methodologies. However, in order to determine if a technique can be considered as an option in protection algorithms, it is necessary to assess this technique. In the work, a comparison between the Wavelet Transform and a new technique based on Principal Component Analysis is presented.

In this context, in previous authors publications, initial evaluation of the advantages and disadvantages of Principal Component Analysis and Wavelet Transform applied for the lightning stroke classification was presented [4]. Both Signal Processing Techniques were used for extract different features corresponding to atmospherics discharges, and an Artificial Neural Network (ANN) classifier was used for classify those signals.

Instead, in this paper a more detailed comparison by using other classification techniques as k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) is presented. On the other hand, in order to illustrate the potential of the methodology based on Principal Component Analysis, different Transmission Lines are analyzed.

In Ref. [4], different disadvantages and advantages useful for protection relays as methodology, percent of classification using ANN, pattern extraction, mother function and others were considered. However, in this research, not only these characteristics but also other crucial topics in protection relays are analyzed and discussed as follows:

- Viability using different patterns classification techniques (k-Nearest Neighbors, Support Vector Machine).
- Methodology for mother functions extraction.
- Optimization for selection of mother functions.

2. Application of Signal Processing Techniques for Lightning Stroke Classification

2.1. Wavelet Transform

By using the Wavelet Transform any signal can be analyzed as an infinite series of wavelets, expressing this signal as a linear combination of a set of function though of translations (τ) and dilations (s) of a mother wavelet [20]. The methodology for lightning stroke classification based on MRA is developed through the following processing:

- Data collection through simulation in ATP program.
- Processing through Multiresolution Analysis, the function is expressed as follows:

$$f(t) = \sum m\psi(t) \tag{1}$$

where m is an integer and ψ represents a set of functions, which is expressed as:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \tag{2}$$

where τ and s are translation and scale, ψ is the transformation function called mother wavelet.

The function is represented as:

$$f(t) = \sum_{p} a_{qp}\varsigma_{qp}(t) + \sum_{q=q_0}\sum_{p} d_{p,q}\upsilon_{p,q}(t) \tag{3}$$

where, a_{qp} and d_{pq} are the approximation and detail coefficients, respectively.

From Eq. (3) it is possible to see that the function is a linear combination of wavelet approximation and detail coefficients a_{qp} and d_{pq}. These coefficients are calculated as follows:

$$a^j(n) = \sum_{i=0}^{L-1} l(i).a^{j-1}(2n-i), \quad 0 \le n < N_j \tag{4}$$

$$d^j(n) = \sum_{i=0}^{L-1} h(i).d^{j-1}(2n-i), \quad 0 \le n < N_j \tag{5}$$

where j represents the decomposition level, $l(i)$ and $h(i)$ correspond to low-frequency and high-frequency filters, respectively. They divide the function $f(s)$ in two parts, the first contains highest frequencies than $fs/2$, and the second contains lower frequencies than $fs/2$. This last part is used in order to obtain a second level, thus the process is continually repeated to obtain different levels. In this context, in this paper, five mother wavelets are used, developing all their decomposition levels.

- The patterns calculated with every mother wavelet are used as input to the classifiers. They are trained and verified for every pattern, and the best result is chosen. This process is repeated with each classifier (Artificial Neural Network, k-Nearest Neighbors and Support Vector Machine). A more detailed explication is presented in Ref. [17].

2.2. Principal Component Analysis

By employing PCA, lightning strokes can be represented as a linear combination of original variables. These new variables are extracted through their eigenvectors, which have orthogonal directions among them [21]. The methodology for lightning stroke classification based on PCA is developed through the following processing:

- Data collection through simulation in ATP program.
- Patterns extraction through PCA, which are based on the variance-covariance matrix as follows:

$$S = \frac{\sum_{i=1}^{n}(f_n - \bar{f})(f_n - \bar{f})}{n-1} \tag{6}$$

where f represents the vector corresponding to lightning stroke signals, which is a signal of 3000 points, and $\bar{f}$ is the mean vector.

In order to extract patterns, their eigenvectors must be calculated. They are selected based on the value of their eigenvalues. For example, the first eigenvector correspond to the eigenvalue, which has the higher variance percentage. The second eigenvector correspond to the eigenvector with the second higher variance percentage. The matrix corresponding to eigenvectors and eigenvalues are presented as follows:

$$U = [eigvector_1 \quad eigvector_2 \quad eigvector_3 \quad eigvector_4 \ldots eigvector_p] \tag{7}$$

$$V = \begin{bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & \lambda_p \end{bmatrix} \tag{8}$$

where, λ represents the eigenvalues of the variance-covariance matrix. Finally, the new patterns are represented as vectors projected on the new base corresponding to principal components.

Please cite this article in press as: J.A. Morales, et al., Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines, Electr. Power Syst. Res. (2014), http://dx.doi.org/10.1016/j.epsr.2014.05.018

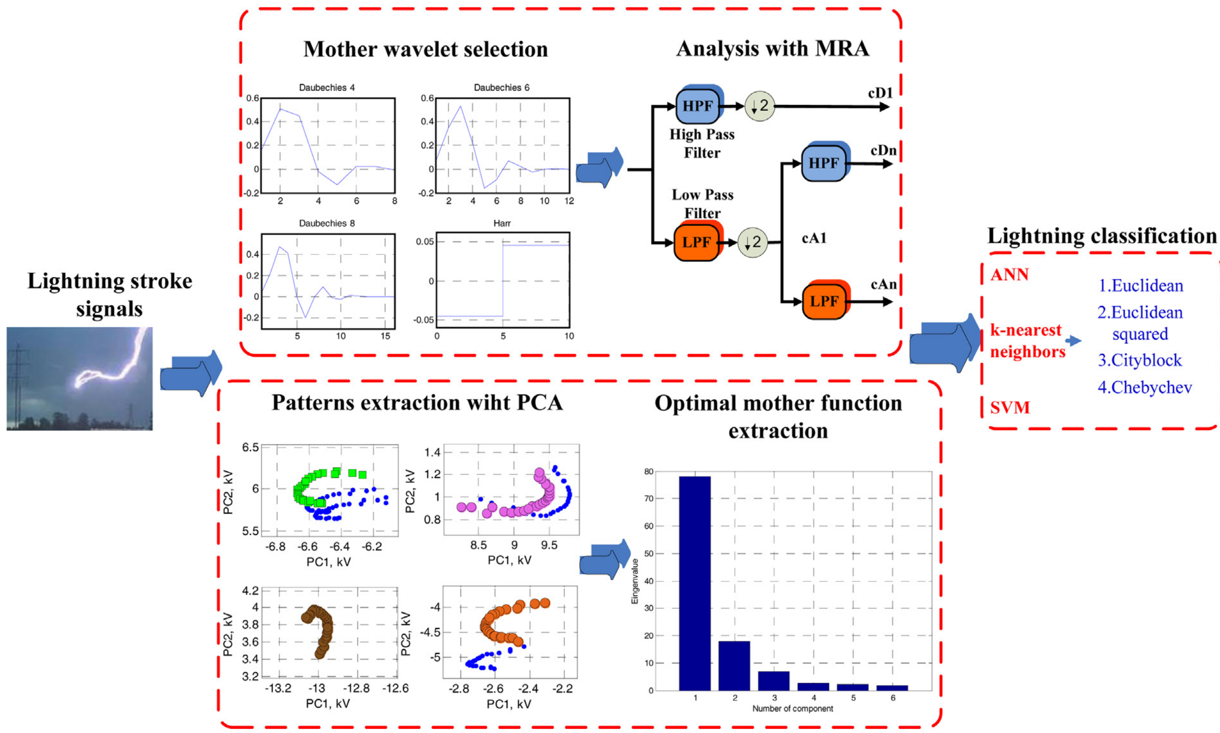


Fig. 1. Methodology based on MRA and PCA.

This representation is done as follows:

$$PC_{1,2,\ldots,p} = [eigvector_1 \;\; eigvector_2 \;\; eigvector_3 \ldots eigvector_p] \cdot [f - \bar{f}]^T \quad (9)$$

where PC represents the new variables, f and $\bar{f}$ is the original variable and the mean vector of original data, respectively.

- Projection of new signals, from previous analysis, it is clear that eigenvectors are functions useful to make the projection of original data to another new base. Thus, principal components corresponding to a test lightning stroke are calculated only projected this signal to the space previously established [18]. Therefore, the projection is done through the eigenvectors as follows:

$$f(pc) = (f_n(k) - \bar{f}) \times U \quad (10)$$

where $\boldsymbol{U}$ is the matrix of eigenvectors, f_n is the new signal, and $\bar{f}$ is the mean vector of the original data.

Similar to the Multiresolution Analysis methodology, Artificial Neural Network, k-Nearest Neighbors and Support Vector Machine are used for the lightning strokes classification. However, in the PCA methodology, their principal components are used as input to these classifiers.

3. Power System Simulation

Analysis Transient Program ATP/EMTP is considered and recognized to be one of the most important and the most widely used program for digital simulation of electromagnetic transient as well as electromechanical nature in Electric Power Systems [22]. Accordingly, in this research ATP software is used to simulate those different elements.

The Electric Power System used in this research contains 6 buses, 5 single transmission lines, 2 transformers and 6 generators. In this EPS, the 230 kV single-circuit transmission line was modeled with the frequency dependent model [23].

The lightning is represented by an impulsive current source (Heidler) and a parallel resistance of 400 Ω [24]. The flashover mechanism corresponding to insulators can be represented by the volt-time curve, which in this work is implemented using MODELS language [25,26]. Thus, in the model called Flashover the volt-time curve is implemented, being calculated in the INIT section.

More than 2500 scenarios are considered covering varieties of the parameters of lightning strokes. These test cases are selected covering the entire range of the transmission line with wave-fronts of voltage signals with 3 ms data windows. In this context, data bases composed with lightning strokes magnitude from 6 kA to 250 kA, positive and negative polarity, and ten different tower footing resistance values from 10 to 200 Ω are considered. The point of impact of lightning strokes is on tower, on phase directly, on mid-span corresponding to the shield wire or phase directly.

4. Results of the comparison

4.1. Methodology

In this research, as regards the Multiresolution Analysis, their coefficients are used as input to the classifiers; however this task is developed varying considerably the mother wavelet and their decomposition levels. Thus, in order to select their best parameters, a hard work of testing is developed.

The top of Fig. 1 shows the procedure through the MRA, the first step is to select the mother wavelet, for example, the mother wavelet daubechies 8 can be selected, and then by using this function and two filters, lightning stroke signals are decomposed in different levels. Finally, these levels are used in the classification process, which are classified by using three classification techniques.

On the other hand, the bottom of Fig. 1 shows the methodology based on Principal Component Analysis. For instance, by using

Please cite this article in press as: J.A. Morales, et al., Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines, Electr. Power Syst. Res. (2014), http://dx.doi.org/10.1016/j.epsr.2014.05.018

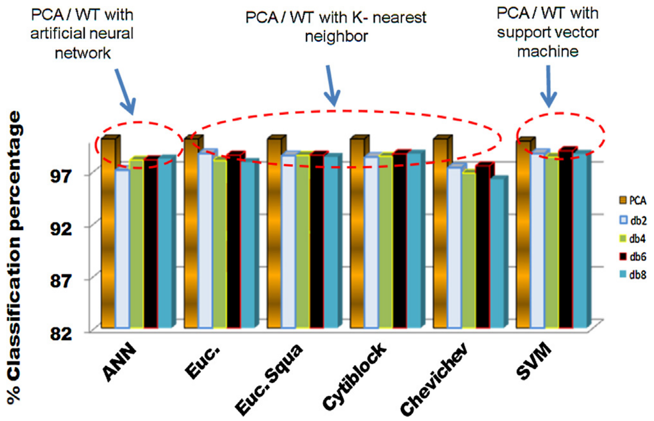


Fig. 2. Comparison of accuracy for lightning stroke classification based on PCA and WT.

the eigenvectors and eigenvalues corresponding to the variance-covariance matrix, their principal components are determined. It is clear that there are as much principal components (PCs) as original variables; however, there are different ways to select the best principal components.

In this context, using only the criterion called *the cumulative percent of variance accounted* (see Section 4.6) it is possible to select the appropriate principal components. Finally, these principal components are used as inputs to the classifiers.

4.2. Comparison of classification accuracy

After transient signals patterns are extracted according to Multiresolution Analysis and Principal Component Analysis, their viability by using three different patterns classification techniques based on k-Nearest Neighbors (k-NN) [27], Support Vector Machines (SVM) [28] and Artificial Neural Network [29], is presented.

Table 1 shows efficiency percentages assessment of lightning strokes classification, depending of the methodology used. From this table it is evident that by using PCA, lightning stroke signals are efficiently classified. On the other hand, by using the MRA methodology, a very acceptable performance also is achieved.

As regards the classification through k-NN, it can be noted that this research used four different distance metrics as: Euclidean, Euclidean squared, Cityblock and Chevyshev.

Based on the above said, the results demonstrate the potential of both methodologies. After an adequate analysis considering the most used mother wavelets, the best classification percentages with the Multiresolultion Analysis are *98.14% with Artificial Neural Network*, *98.62% with k-Nearest Neighbors* and *98.87% with Support Vector Machine*. Still, the results obtained through Principal Component Analysis are practically 100% in most cases. A percentage less than 100% is obtained with the classifier based on Support Vector Machine. However, these values are higher than those percentages obtained with MRA. The graphical representation of this issue is presented in Fig. 2.

4.3. Dimensional reduction

The Multiresolution Analysis methodology uses an empirical procedure i.e. the signals are tested with different mother wavelets and different decomposition levels whose size depends of the decomposition grade. In this context, it is interesting to note that the wavelet analysis reduced the dimensional of the lightning stroke signal by a factor of 62, approximately.

As regards the Principal Component Analysis methodology, it reduced the dimensional based on an optimization corresponding to their variances, obtaining an excellent dimensional reduction. By using PCA, lightning stroke signals are reduced by a factor of 500, approximately.

4.4. Patterns defined

The most important feature of an extraction-classification problem is related to the extraction of features. In this context, Figs. 3–6 show the extracted patterns through the PCA and MRA processing, respectively.

Fig. 3 shows the patterns corresponding to atmospherics discharges extracted through PCA. For example, in Fig. 3a the blue color pattern corresponds to lightning strokes on live wires without fault. Still, in Fig. 3b the red color pattern represents lightning strokes with fault.

Analyzing the figure, it is possible to see that well defined patterns are extracted, being lightning strokes signals clearly distinguished.

It is necessary to note that by using Eq. (10); new signals f_n can be projected. For instance, Fig. 4 shows how new lightning strokes are situated on the original patterns. Thus, new signals can be easily tested with the PCA methodology.

On the other hand, through the Multiresolution Analysis, the wavelets coefficients corresponding to atmospherics discharges are analyzed considering the maximum number of levels, obtaining 1×2 and 1×3 size coefficients (two and three dimensions).

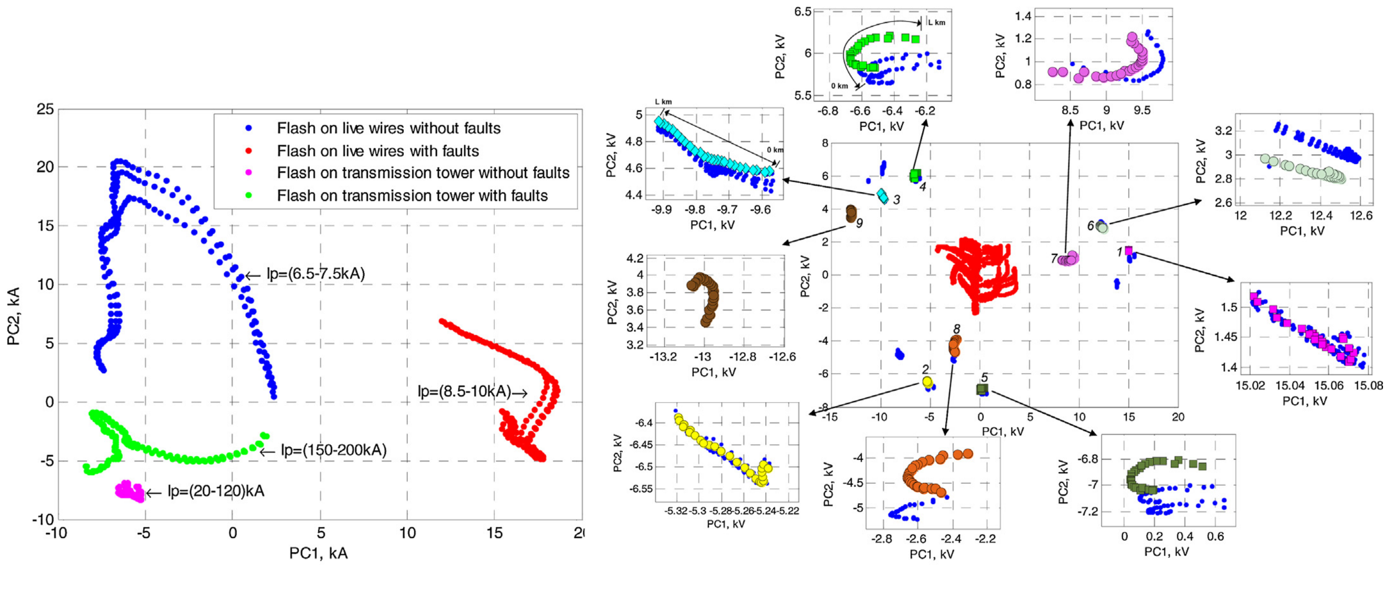
Figs. 5 and 6 show the coefficients projected to a 2D and 3D new base. From these figures, it is clear that lightning stroke signals with and without fault are not well distinguished. For example, Fig. 5 shows lightning stroke signals analyzed through the mother wavelet daubechies 4 and daubechies 6, in this figure the red and blue color patterns represent lightning strokes signals with and without fault, respectively. However, it is possible to see that these patterns are not well defined. Therefore, in order to classify these signals, it is necessary more coefficients than principal components. A similar analysis is possible to see in Fig. 6, where the signals daubechies 8 and haar are used as mother wavelets.

4.5. Methodology for mother functions extraction

As stated in Ref. [4], the daubechies mother wavelet family has been the most suitable function to analyze not only lightning stroke transients but also of other transients. Their best choice is a mother function that acceptably describes the studied signal. However, there are not criteria for the extraction or selection of mother wavelets. For example, from bibliographic review, it has been found that the daubechies mother functions order 2 (db2), order 4 (db4), order 6 (db6), order 8 (db8) and the mother wavelet Haar usually are a good choice for the lightning strokes classification on TLs [4,17].

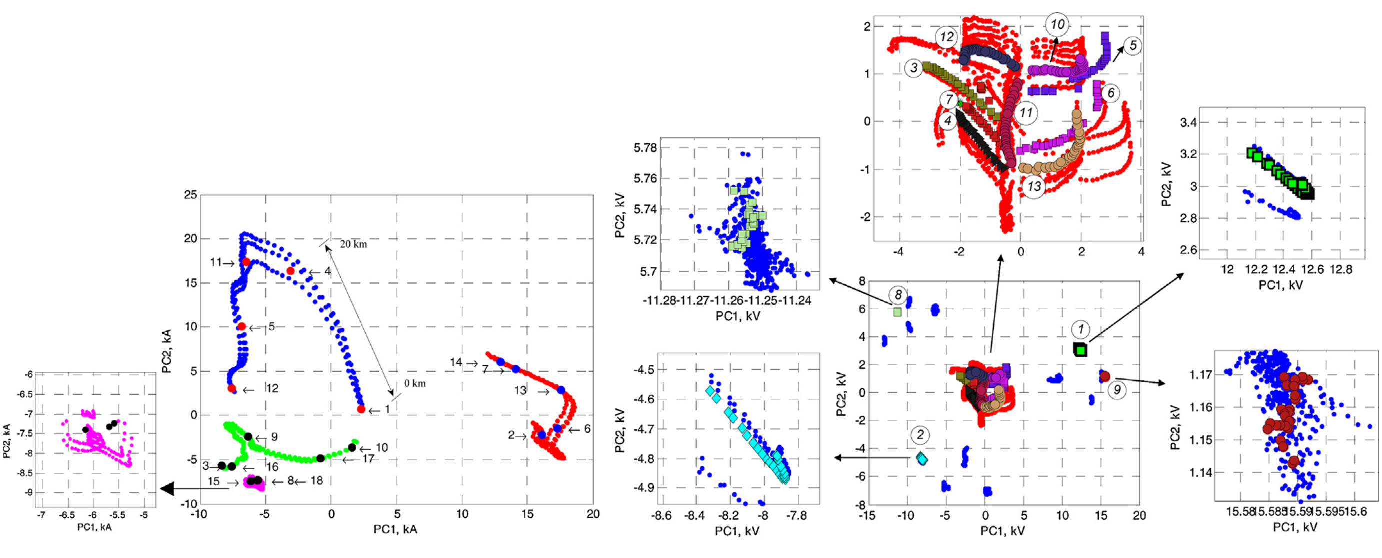
Thus, the best wavelet decomposition level useful for the classification process was determined in an empirical manner, i.e. by using different wavelets coefficients the classification percentage is verified, and the best classification percentage is used. Therefore, it is clear that the response of MRA depends of the mother wavelet selected.

On the other hand, by applying PCA, the eigenvectors corresponding to the first electric power system are used as mother functions. For instance, Fig. 7 shows their first four orthogonal eigenvectors. It is very important to note that these mother functions are considered the best selection not only to analyze the first electric power system but also others electric power systems. Therefore, mother functions standardized for different power



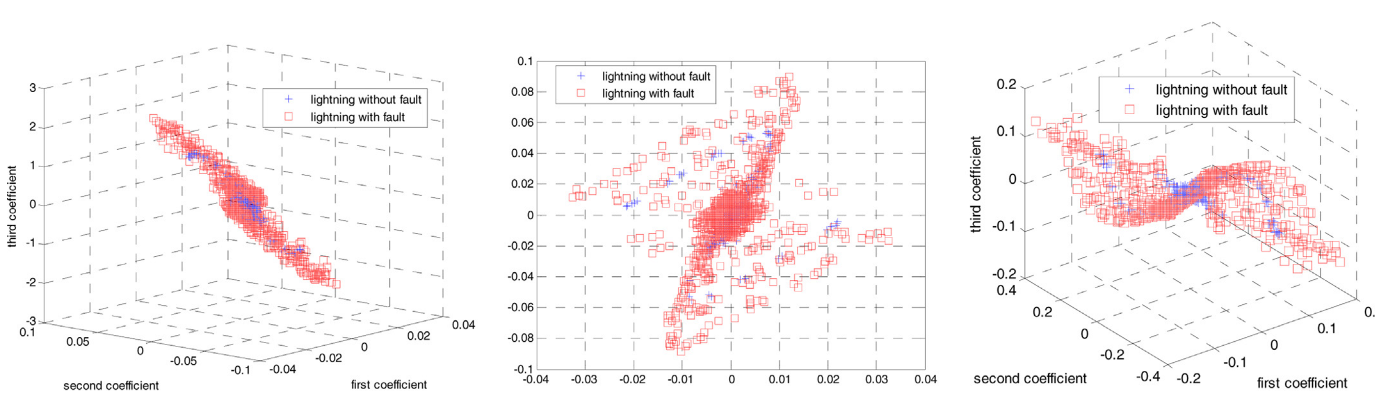
(a) *(b)*

Fig. 3. Representations of lightning strokes (a) flash peak current magnitude, (b) lightning with and without fault.



(a) *(b)*

Fig. 4. Signals projected on the first two PCs corresponding to (a) flash peak current magnitude, (b) lightning with and without fault.



(a) *(b)* *(c)*

Fig. 5. Patterns extracted in the MRA subspace, with mother wavelet, (a) daubechies 4 two dimensions, (b) daubechies 4 three dimensions, (c) daubechies 6.

Please cite this article in press as: J.A. Morales, et al., Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines, Electr. Power Syst. Res. (2014), http://dx.doi.org/10.1016/j.epsr.2014.05.018

Table 1
Total classification results based on PCA and WT.

		Classification percentage				
		k-NN				
	AzNN	Euclidean	Euc. Squared	Cityblock	Chebyshev	SVM
PCA	100	100	100	100	100	99.795
Daubechies 2	97.04	98.62	98.43	98.28	97.27	98.62
Daubechies 4	97.98	97.93	98.45	98.35	96.73	98.27
Daubechies 6	98.01	98.46	98.46	98.59	97.44	98.87
Daubechies 8 1	98.14	97.81	98.29	98.6	96.17	98.58

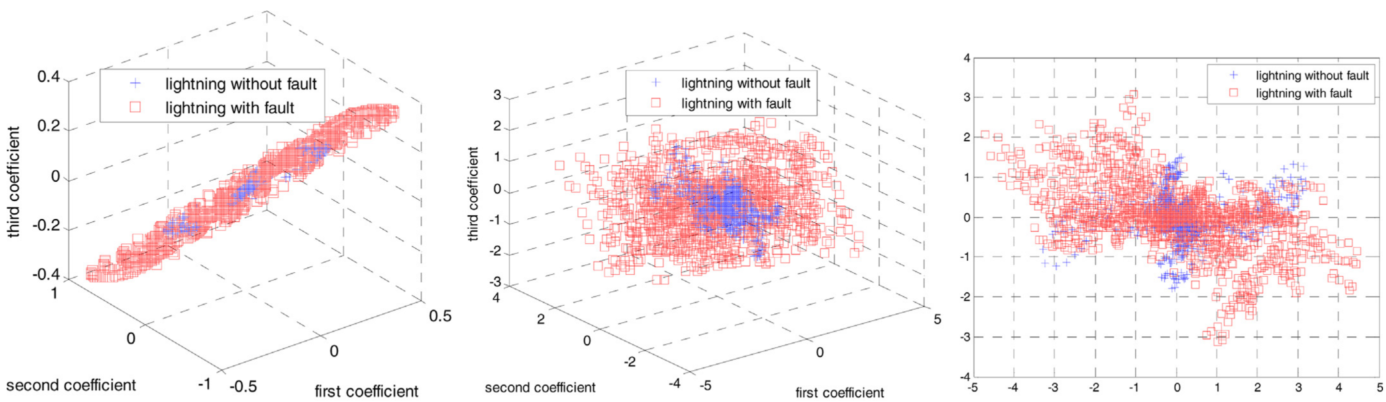


Fig. 6. Patterns extracted in the MRA subspace, (a) daubechies 8, (b) haar two dimensions, (c) haar three dimensions.

system considering different topologies of transmission lines and towers, can be analyzed (see Section 4.7).

4.6. Mother functions selection optimization

As stated in the previous section, there is not a criterion for mother wavelets selection. Instead, mother functions extracted through Principal Component Analysis corresponding to their eigenvectors can be extracted. However, PCA has a very interesting property useful to develop criteria, which can be used for an adequate choice of mother functions as follows:

First: The first two eigenvectors are usually used for testing signals analysis. Therefore, only the first two orthogonal functions are used. It is well acceptable if a visual representation of lightning strokes signals is obtained.

Second: A criterion called *the cumulative percent of variance accounted for* (*CPVA*) can be used for the number of mother functions choice, retaining the highest percentage of their variance. The criterion retains components that account a specified proportion (or percentage) of variance in the data set, i.e. the number of components is chosen based on their eigenvalues as follows:

$$\frac{\sum_{i=1}^{K} \lambda i}{\sum_{i=1}^{N} \lambda i} \geq \text{Threshold}\,(e.g.,\ 0.9\,\text{or}\,0.95) \tag{11}$$

where λ_K is the overall variance value and λ_N is the variance value corresponding to the N principal component.

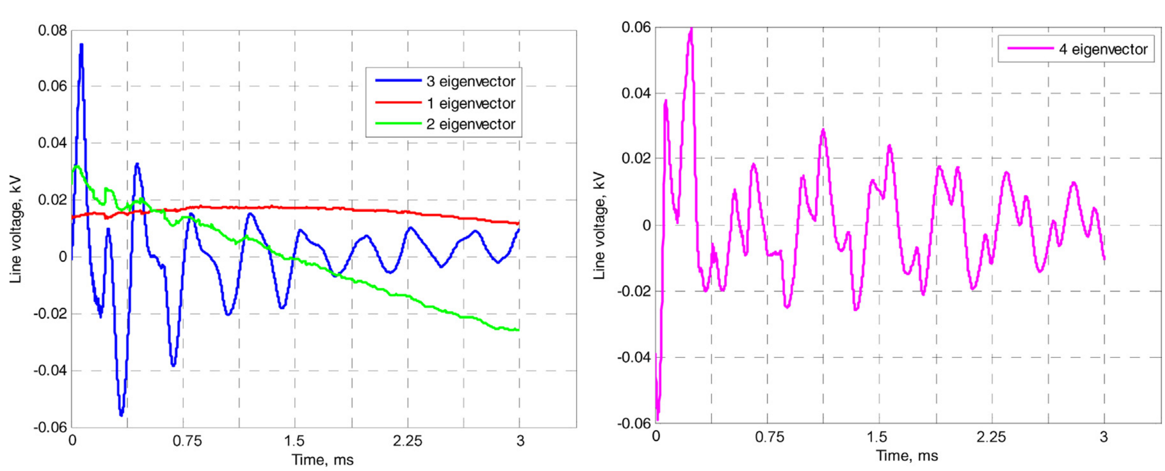


Fig. 7. Eigenvectors corresponding to PC, (a) first three eigenvectors, (b) fourth eigenvector.

Please cite this article in press as: J.A. Morales, et al., Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines, Electr. Power Syst. Res. (2014), http://dx.doi.org/10.1016/j.epsr.2014.05.018

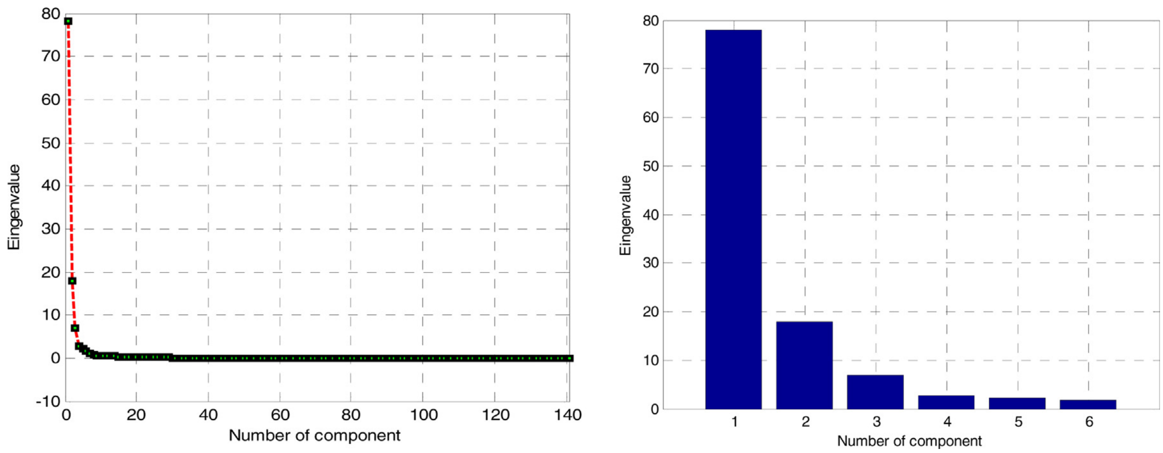


Fig. 8. Variance of the eigenvalues.

Table 2
Variance of the first six eigenvalues.

Component	1	2	3	4	5	6
% Variability	73.6	22.1	1.9	1.17	0.7	0.2

Usually, a CPVA value corresponding to 90–95% is acceptable. In this work, in order to make the comparison between both methodologies, the CPVA value selected was 98%. Results of the cumulative percent are detailed in Table 2. In this Table, it can be seen that those first six principal components account more than the 98% of the total variance. For example, the first component accounts the 73.6% of the total variance and the second component accounts the 22.1%. A visual representation of the variance is presented in Fig. 8.

From previous analysis it is clear that by using the CPVA criterion, only the six eigenvectors (mother functions) can be used to make the analysis of the lightning stroke signals.

4.7. *Viability of PCA methodology*

Based on the above said, it is clear that Principal Component Analysis is useful to make the analysis and study of lightning stroke signals, improving some characteristic of Multiresolution Analysis. Thus, in order to investigate their potential and applicability for the lightning stroke classification, in this section a study is developed using another electric power system. Fig. 9 shows the EPS simplified topology simulated in the program ATP/EMTP [22].

The EPS contains 12 buses, 6 double transmission lines, 5 single transmission lines, 3 transformers and 7 generators. In this EPS, two 230 kV double-circuit transmission lines denoted as M-N and N-O with two-ground wires were modeled with the frequency dependent model.

The tower model shown in Fig. 10 is simulated along the transmission line M-N and N-O. As regards the insulator string and lightning stroke, they are simulated using a voltage-dependant flashover switch and an impulsive current source (Heidler), respectively.

4.7.1. *Verification of methodology*

Transmission Lines protection relays R_1 and R_2 are installed at both sides of the lines M-N and N-O, respectively. The data set is composed by different signals varying the flash peak current magnitude, positive and negative polarity, tower footing resistance (TFR), point of impact and others. From the previous database, their principal components are extracted, which are used for the classification process.

As regards the simulations, they are evaluated in three different forms at both terminals of the transmission lines as follows.

Case 1: Lightning hits along the transmission line M–N, and the signals are evaluated by relay R_1 at bus M.
Case 2: Lightning hits along the transmission line N–O, and the signals are evaluated by relay R_2 at bus N.
Case 3: Lightning hits along the transmission line N–O, and the signals are evaluated by relay R_1 at bus M.

More than 3300 scenarios are considered covering a lot of parameters of lightning strokes. These test cases are selected covering the entire range of the transmission line with wave-fronts

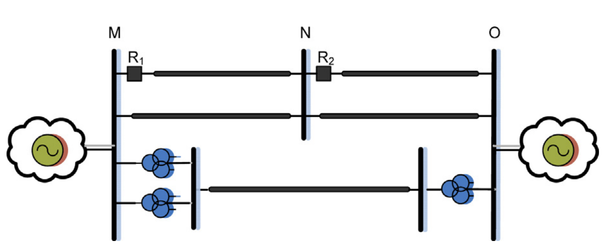


Fig. 9. Verification electric power system.

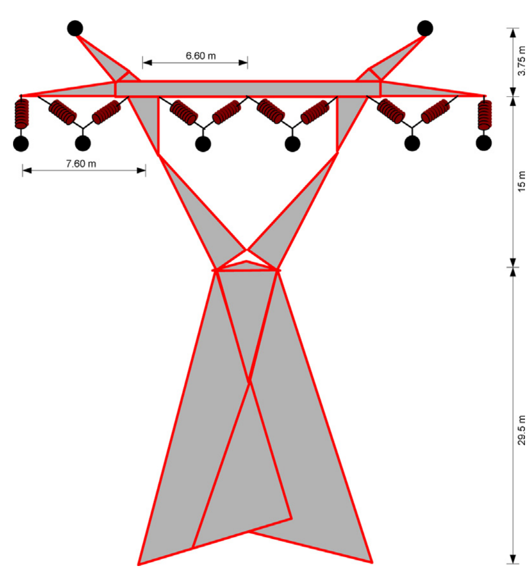


Fig. 10. Double tower model.

Please cite this article in press as: J.A. Morales, et al., Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines, Electr. Power Syst. Res. (2014), http://dx.doi.org/10.1016/j.epsr.2014.05.018

Table 3
Testing lightning stroke features.

Ip (kA)	Lightning features			Point of impact	
	Polarity		TFR (Ω)	On tower	On phase
	Positive	Negative			
3, 6, 9, 8, 12,15,	√	√	10		√
3.5, 7, 10.5, 14	√	√	10		√
3.3, 6.6, 9.9, 13.2	√	√	10		√
2.9, 5.8, 8.7, 11.6	√	√	10		√
3.15, 6.3, 9.45, 12.6	√	√	10		√
34, 68, 102, 136, 170, 204	√	√	40, 80, 120, 160, 200, 240	√	
40, 80, 120, 160, 200, 240	√	√	40, 80, 120, 160, 200, 240	√	
50, 100, 150, 200, 250, 300	√	√	42, 84, 126, 168, 216, 264	√	
30, 60, 90, 180	√	√	38, 76, 114, 152, 190, 228	√	
32, 64, 96, 128, 160, 192	√	√	45, 90, 135, 180, 225, 270	√	
33, 66, 99, 132, 165, 198	√	√	40, 80, 120, 160, 200, 240	√	
38, 76, 114, 152, 190, 228	√	√	40, 80, 120, 160, 200, 240	√	
Ip: flash peak current amplitude			TFR: tower footing resistance	√	

Table 4
Principal component values corresponding to the case 1.

Testing signals	Lightning features: Type	Ip (kA)	FTR (Ω)	Polarity	Principal components: PCI	PC2	PC3	PC4	PC5	PC 6
1	Without fault	3.150	10	Neg.	−57.00	−1.01	−1.13	0.74	−0.47	−0.95
2	With fault	9.45	10	Neg.	−31.76	1.02	2.24	2.54	−3.17	1.22
3	Without fault	34	40	Neg.	−57.95	−2.31	−1.37	0.83	1.30	−0.27
4	With fault	228	240	Pos.	26.80	−12.04	14.29	−17.55	−2.59	−2.92
5	With fault	102	120	Neg.	4.50	4.93	−2.76	3.93	3.59	−1.79

corresponding to voltage signals with 3 ms data windows. In this context, databases composed with lightning strokes of positive and negative polarity, twenty five different tower footing resistance values from 10 to 240 Ω are considered. The point of impact of lightning strokes is on tower, on phase directly (see Table 3).

4.7.2. Results

4.7.2.1. Case 1. It is clear that in Eq. (10) the matrix **U** is a projection matrix, thus in this case, the aim is transforms or project the new signals corresponding to the second electric power system onto the principal components axes corresponding to the first electric power systems. In this context, by using the eigenvectors corresponding to the first electric power system (see Fig. 7), the new signals are projected. It is necessary to note that is not necessary extract new eigenvectors.

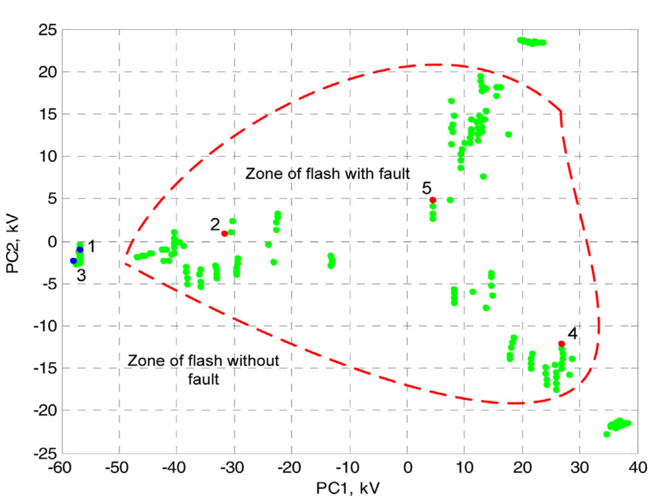


Fig. 11. Patterns corresponding to the case 1.

Fig. 11 shows the patters extracted corresponding to lightning strokes, which were tested by the relay R_1. This figure shows that those new signals were situated correctly in the different lightning stroke patterns. When lightning strokes with fault are presented, their principal components values are localized inside of the red line pattern. On the contrary, when lightning strokes without fault are presented, their principal components are localized outside of the red line pattern.

For example, Table 4 depicts the values corresponding to the first six principal components during some lightning strokes, which are extracted only by applying the first six eigenvectors corresponding to the variance–covariance. See Eq. (10).

On the other hand, Table 5 summarizes the classification results by using k-Nearest Neighbors and Support Vector Machine techniques. According to these results, by employing the eigenvectors and classifiers techniques, successfully classification of lightning stroke with and without fault is achieved.

4.7.2.2. Case 2. As regards the case 2, the lightning strokes signals evaluated by the relay R_2 at bus N are projected similar to the case 1, i.e. by applying Eq. (10) and the eigenvectors corresponding to the first electric power system.

Fig. 12 shows the patterns corresponding to this case, which show that their principal components are distinguished among them. Hence, the lightning strokes signals with fault are concentrated on the pattern center. On the contrary, the atmospheric discharges without fault are localized in the exterior pattern.

Based on the above said, Table 6 presents the principal components corresponding to some testing signals, in this table it is possible to verify that principal components values corresponding to lightning with and without fault are different.

Finally, the performance evaluation is summarized in Table 7, where by using the k-Nearest Neighbors classifier, a percentage of 100% was obtained. Regarding the Support Vector Machine classifier, some lightning strokes were not classified correctly, those

Table 5
Total classification results corresponding to the case 1.

	Lightning class	Testing lightnings	Correct	Incorrect	Correct (%)	Incorrect (%)
k-NN	Without fault	590	590	0	100.00	0.00
	With fault	970	960	10	98.97	1.03
SVM	Without fault	590	590	0	100.00	0.00
	With fault	970	970	0	100.00	0.00

Table 6
Principal component values corresponding to the case 2.

Testing	Lightning features				Principal components					
signals	Type	Ip (kA)	FTR (Ω)	Polarity	PCI	PC2	PC3	PC4	PC5	PC6
1	Without fault	6.300	10	Neg.	−15.63	−0.06	−1.03	1.53	−1.76	−1.06
2	Without fault	12	10	Pos.	9.72	6.54	−0.12	0.40	−0.40	0.43
3	With fault	150	126	Neg.	1.32	1.00	−1.50	2.91	−5.29	−1.37
4	With fault	14	10	Neg.	−7.22	0.14	0.24	0.16	−0.51	0.06

Table 7
Total classification results corresponding to the case 2.

	Lightning class	Testing lightning	Correct	Incorrect	Correct (%)	Incorrect (%)
k-NN	With fault	1080	1080	0	100.00	0.00
	Without fault	660	660	0	100.00	0.00
SVM	With fault	1080	1080	0	100.00	0.00
	Without fault	660	660	0	100.00	0.00

Table 8
Principal component values corresponding to the case 3.

Testing	Lightning features				Principal components					
Signals	Type	Ip (kA)	FTR (Ω)	Polarity	PCI	PC2	PC3	PC4	PC5	PC6
1	Without fault	2.900	10	Neg.	−56.74	−2.83	−0.27	0.19	−0.10	−0.49
2	With fault	160	160	Pos.	12.90	16.61	1.17	2.62	0.75	4.38
3	With fault	228	240	Pos.	11.51	18.39	−8.68	4.96	5.08	1.51
4	Without fault	3150	10	Neg.	35.88	−21.6	1.78	−0.37	0.29	−0.38

signals were classified as lightning with fault, but a success rate of 99% was obtained.

4.7.2.3. *Case 3.* Finally, this case is developed similar to the previous case. The simulations were repeated with all parameters corresponding to the case 2, except that the signals are evaluated by the relay R_1. The aim is to allow that the relay R_1 respond to lightning strokes, which hit the line N–O. Thus, it could be useful to develop remote backup protection.

In this context, Fig. 13 shows the patterns corresponding to atmospheric discharges with and without fault. From this figure, it is also possible to see that these signals are distinguished between them. On the other hand, the principal components values corresponding to some testing signals are presented in Table 8.

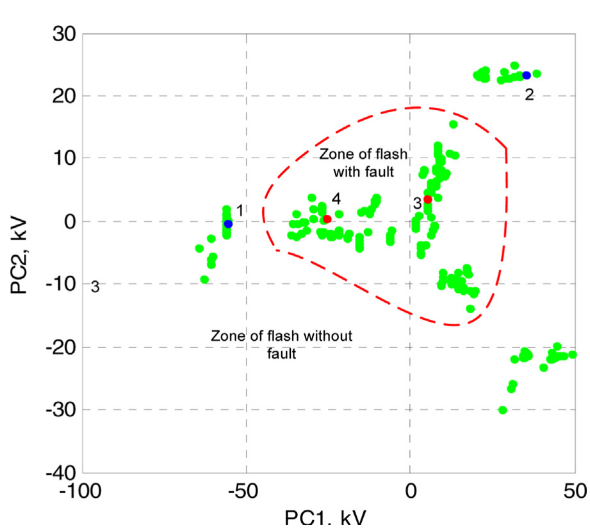


Fig. 12. Patterns corresponding to the case 2.

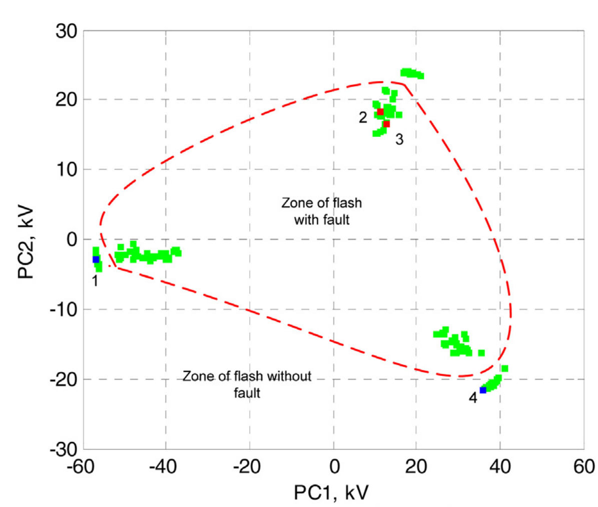


Fig. 13. Patterns corresponding to the case 3.

Table 9
Total classification results corresponding to the case 3.

	Lightning class	Testing lightning	Correct	Incorrect	Correct (%)	Incorrect (%)
k-NN	With fault	1080	1080	0	100.00	0.00
	Without fault	660	660	0	100.00	0.00
SVM	With fault	1080	1040	40	96.30	3.70
	Without fault	660	660	0	100.00	0.00

Classification results show that the performance of the proposed methodology, identifying if the lightning stroke produces or not faults is correct (see Table 9). In this context, the authors consider that the methodology could perform remote backup protection. However, such algorithm exceeds the scope of this research.

Previous analysis shows that there is very little difference among the extracted patterns. This similarity among these patterns demonstrates why this methodology can be useful for protection relays. Based on the previous tables, the results show a 100% overall accuracy and close to 99% in the classification of lightning strokes is achieved. It is necessary to note that excellent results are achieved by using another electric power system, which is completely different to those used in previous sections and publications [4].

5. Conclusions

- This paper presents an assessment of different signal processing techniques (SPT) applied to lightning stroke classification considering different signal conditions.
- Test results show that the both methodologies performance for lightning stroke classification considering the principal features of the lightning are acceptable. However, Principal Component Analysis technique has better results than Wavelet Transform, improving several of their characteristics.
- Comparisons considering different classification techniques as Artificial Neural Network, k-Nearest Neighbors and Support Vector Machine are developed. Results illustrate that not only PCA, but also these techniques can easily be adapted for lightning strokes classification.
- As regards the PCA, an easy mathematical analysis provides an optimal system to determine orthogonal functions for relaying purposes. Therefore, only the mean vector and the first eigenvectors are used to apply the methodology, reducing considerably the processing time.
- The PCA methodology can facilitate the selection of mother functions for their application and response to lightning phenomena. Thus, PCA could be the most recent solution to overcome the shortcomings of Wavelet Transform in the selection of mother functions.
- This comparison demonstrates how PCA can be employed to make the analysis and study of protection relays considering different electric power systems. Double transmission lines have been widely employed in EPS. These topologies can have severe mutual coupling effect, and the protection of these transmission lines is a serious problem. However, by using the PCA methodology, the lightning stroke classification of parallel transmission lines presents an acceptable performance.

Acknowledgments

The authors gratefully acknowledge: German Academic Interchange Program (DAAD), to the Institute of Electric Energy at the San Juan National University (IEE, UNSJ), to the Electrical Engineering Faculty at the Salesian Polytechnic University, to the Department of Electrical Engineering at the Federal University of Rio Grande do Sul and the Institute of Energy Systems, Energy Efficiency and Energy Economics (ie3) at the TU Dortmund University by the assistance and financial support.

References

[1] A.A. Mohd, S.P. Abdul Karim, Protection system analysis using fault signatures in Malaysia, Int. J. Electr. Power Energy Syst. 45 (2013) 194–205.
[2] E. Bompard, T. Huang, Y. Wu, M. Cremenescu, Classification and trend analysis of threats origins to the security of power systems, Int. J. Electr. Power Energy Syst. 50 (2013) 50–64.
[3] G. Moreno, R. Hernandez, J. Montoya, Failures in outdoor insulation caused by bird excrement, Electr. Power Syst. Res. 80 (2010) 716–722.
[4] J.A. Morales, E.A. Orduña, C. Rehtanz, R.J. Cabral, A.S. Bretas, Comparison Between PCA and WT Filtering Methods for Lightning Stroke Classification on Transmission Lines, in: Proc. SIPDA, XII International Symposium on Lightning Protection, Belo Horizonte, Brazil, October, 2013.
[5] I. Ramirez, R. Hernandez, G. Montoya, C. Romualdo, Analysis of the Mexican lightning activity monitored by NASA satellites, Electr. Power Syst. Res. 72 (2) (2004) 187–193.
[6] J.O. Adepitan, E.O. Oladiran, Analysis of the dependence of power outages on lightning events within the ijebu province, Nigeria, J. Environ. Earth Sci. (2012) 850–856.
[7] U.J. Minnaar, C.T. Gaunt, F. Nicolis, Characterization of power system events on South African transmission power lines, Electr. Power Syst. Res. 88 (2012) 25–32.
[8] S. Shimada, et al., Estimation of Transmission Line Faults from Observation of Winter Lightning Channel in Hokuriku Area. IEE Japan. ED-01-180, SP-01-25, HV-01-79, 2001, pp. 7–11.
[9] G. Andersson, et al., Causes of the 2003 major grid blackouts in North America and Europe and recommended means to improve system dynamic performance, IEEE Trans. Power Syst. 20 (4) (2005) 1922–1928.
[10] Y.V. Makarov, V.I. Reshetov, A. Stroev, I. Voropai, Blackout prevention in the United States, Europe, and Russia, Proc IEEE 93 (11) (2005) 1942–1955.
[11] EPRI AC Transmission Line Reference Book – 200 kV and Above, third ed., Electric Power Research Institute, 2005.
[12] W. Fan, On Power Quality and Protection, Chalmers University of Technology, Goteborg, Sweden, 2001.
[13] J.A. Morales, R.P. Aguilar, E.A. Orduña, F.E. Pérez, Advances in lightning identification for transient based protection of power transmission lines, in: Clagtee, Ninth Latin-American Congress on Electricity Generation and Transmission, Argentina, 2011.
[14] ANSI/IEEE Standard 100, IEEE Standard Dictionary of Electrical and Electronics Terms, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854, 2014.
[15] N. Kehtarnavaz, M. Keramat, DSP System Design: Using the Tms320c6000, first ed., Prentice Hall, Saddle River, USA, 2000.
[16] Texas Instruments, Floating-Point Digital Signal Processor Tms320c6713, SPRS 186L, Dallas, Texas, 2005.
[17] J. Morales, E.A. Orduña, C. Rehtanz, Classification of lightning stroke on transmission line using multi-resolution analysis and machine learning, Int. J. Electr. Power Energy Syst. 58 (2014) 19–31.
[18] J. Morales, E. Orduña, Patterns extraction for lightning transmission lines protection based on principal component analysis, IEEE Latin America Trans. 11 (1) (2013).
[19] J. A. Morales, E. A. Orduña, C. Rehtanz, "Identification of Lightning Stroke due Shielding Failure and Back flashover", submitted to IEEE Transaction Power Delivery, in press.
[20] S.G. Mallt, A theory for multiresolution signal decomposirion: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (July (7)) (1989) 674–693.
[21] D. Peña, Análisis de Datos Multivariantes, McGRAW-HILL, Madrid-España, 2002.
[22] ATP DRAW version 3.5 for Windows 9x/NT/2000/XP Users'Manual.
[23] J.R. Marti, Accurate modeling of frequency-dependent transmission lines in electromagnetic transient simulations, IEEE Trans. Power Appl. Syst. PAS-101 (1) (1982) 147–157.
[24] F. Heidler, M. Cvetic, B.V. Stanic, Calculation of lightning current parameters, IEEE Trans. Power Deliv. 14 (2) (1999) 399–404.
[25] L. Dubè, I. Bonfanti, MODELS: a new simulation tool in the EMTP, European Trans. Electr. Power 2 (January/February (1)) (1992) 45–50.
[26] L. Dubé, Users Guide to Models in ATP, 1996.
[27] J. Han, M. Kamber, Data Mining: Concepts and Techniques, second ed., Elsevier, Illinois, 2006.
[28] C. Cortes, V. Vapnik, Support-vector network, Mach. Learn. 20 (1995) 273–297.
[29] A. Abraham, Artificial Neural Networks, Oklahoma State University, Stillwater, 2005.

Please cite this article in press as: J.A. Morales, et al., Comparison between Principal Component Analysis and Wavelet Transform 'Filtering Methods for Lightning Stroke Classification on Transmission Lines, Electr. Power Syst. Res. (2014), http://dx.doi.org/10.1016/j.epsr.2014.05.018