Memorias de SICEL 2023 | UNIVERSIDAD NACIONAL DE COLOMBIA

Comparison of transformations and feature extraction techniques to characterize fault-induced voltage sags

Comparación de transformadas y técnicas de extracción de características para el análisis de huecos de tensión inducidos por fallas

Joaquín E. Caicedo [1a], Edwin Rivas [1b], Jan Meyer [2]

[1] Grupo de Compatibilidad e Interferencia Electromagnética (GCEM), Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Colombia. Orcid: 0000-0003-2168-829X [a], 0000-0003-2372-8056 [b]. Emails: jecaicedon@udistrital.edu.co [a], erivas@udistrital.edu.co [b]

[2] Institute of Electrical Power Systems and High Voltage Engineering, Technische Universität Dresden, Germany. Orcid: 0000-0002-6884-5101. Email: jan.meyer@tu-dresden.de

Recibido: 17/07/2023. Aceptado: 22/08/2023. Versión final: 12/10/2023

Abstract

This paper presents a comparative study of domain transformations and feature extraction techniques to characterize fault-induced voltage sags. For this purpose, synthetic signals of fault-induced voltage sags are generated through extensive simulations in MATLAB/Simulink. Next, some relevant transformations are applied to the synthetic signals, namely, the space phasor model, discrete Fourier transform, and short-time Fourier transform. A set of statistical, time series, and spectral features are extracted from transformation outputs to obtain signal characterization useful, for instance, for classification of voltage sags employing artificial intelligence techniques. The comparison of the applied domain transformations and feature extraction techniques covers quantitative and qualitative aspects including computation time, storage requirement, linear separability and physical interpretation of features, and suitability for characterizing voltage sags. Finally, the main findings of the work are discussed, and conclusions are remarked.

Keywords: Domain transformation; Feature extraction; Power quality; Voltage sag.

Resumen

Este artículo presenta una comparación de transformadas y técnicas de extracción de características para analizar huecos de tensión inducidos por fallas. Para este propósito, se generan señales sintéticas de huecos de tensión inducidos por fallas mediante simulaciones en MATLAB/Simulink. Luego, se aplican transformadas a las señales sintéticas como el espacio del modelo fasorial, la transformada discreta de Fourier y la transformada de Fourier de tiempo corto. Se extraen características estadísticas, de series temporales y espectrales de las salidas de las transformadas para obtener una caracterización de la señal, útil, por ejemplo, para clasificar huecos de tensión usando inteligencia artificial. La comparación de las transformadas y las técnicas de extracción de características cubre aspectos cuantitativos y

Como citar J. E. Caicedo, E. Rivas and J. Meyer, “Comparison of transformations and feature extraction techniques to characterize fault-induced voltage sags,” in XI Simposio Internacional de Calidad de la Energía Eléctrica, Valledupar: Universidad Nacional de Colombia, Nov. 2023. doi: https://doi.org/10.15446/sicel.v11.110048

cualitativos como el tiempo de cómputo, requisitos de almacenamiento, separabilidad lineal e interpretación física de las características, e idoneidad para caracterizar huecos de tensión. Finalmente, se discuten los principales hallazgos y se presentan las conclusiones.

Palabras clave: Calidad de potencia; Extracción de características; Hueco de tensión; Transformada de dominio.

1. Introduction

Power Quality Disturbance (PQD) characterization, detection, and classification, is essential for applications in power systems including real-time monitoring and impact assessment, analysis of disturbance propagation, source location, and definition of mitigation measures [1], [2]. Among PQDs, voltage sags are of particular interest because of the huge economic losses caused during the operation of distribution networks [3], [4].

Voltage sags are mainly caused by the connection of heavy loads such as large induction motors, energization of transformers, and power system faults [5]. The latter is the most common root-cause of voltage sags and derive a classification of the disturbance according to the type of fault, i.e., three-phase, two-phase, two-phase to ground, and single phase to ground, which result in ten types considering the possible combinations [6]. The identification of the type of fault that originates the voltage sag, including the recognition of the faulted phase(s), and the characterization of the disturbance, facilitates the protection system to locate and isolate the affected area and thus, decreases voltage sag duration, provides more information on the severity of the event, and enables the proper operation of mitigation measures and devices such as dynamic voltage restorers [7]. Thus, the accurate and automatic detection, classification, and characterization of fault-induced voltage sags is important for the proper operation of protection systems, and the implementation of mitigation measures.

The main challenge in the process of detection, classification, and characterization of PQDs, including voltage sags, is to find efficient indices and metrics that provide reliable and suitable results according to the application. Those indices and metrics can be obtained using domain transformations and feature extraction techniques [1]. Transformations applied to characterize and identify types of voltage sags include techniques in the time domain such as the Phase Space Reconstruction (PSR) [8] and Space Phasor Model (SPM) [9], also referred to as the space vector [6], [7], frequency domain such as Discrete Fourier Transform (DFT) [10], and time-frequency domain such as Short-Time Fourier Transform (STFT) [5], Wavelet Transform (WT) [11]–[14], and Stockwell Transform (ST) [15]–[17]. Moreover, various types of features have been extracted for voltage sag characterization and classification such as statistical [18], time series [19], [20], image-based [6], [7], [9], and spectral features [21]. It is worth to mention that the transformation step can be omitted, which means that features are not only extracted from transformation coefficients but also from the original waveforms. For instance, higher-order statistics [22], Euclidean distance [23], and gray images [24], have been directly extracted from original waveforms for PQD classification. Each type of features provides specific advantages according to the application and tools to be used. For instance, image-based features are effective when image processing techniques are used for classification, e.g., convolutional neural networks [8], [9].

The success in the stages following transformation and feature extraction depends on several aspects according to the required application. Therefore, some works in the literature have performed comparative studies considering diverse aspects of transformations and feature extraction techniques. Most of those studies focused on analyzing qualitative aspects of the approaches, such as mathematical representation, complexity of implementation, and general descriptions [2], [25]–[27], while few also considered quantitative aspects, including performance and computational burden [28]. Moreover, comparative studies in the literature are mainly dedicated to general-purpose methodologies for the classification of multiple PQD and not to specific disturbances, such as voltage sags.

In this context, this paper develops a comparative study of some relevant time-domain, frequency-domain, and time-frequency domain transformations applied specifically to fault-induced voltage sags, i.e., SPM, Discrete Fourier Transform (DFT), and STFT, respectively, and statistical, time series, and spectral feature extraction techniques, to demonstrate advantages and drawbacks and provide the following contributions:

- The analysis of quantitative and qualitative aspects of the techniques, such as computation time, storage requirement, linear separability of features, physical interpretation of features, and suitability for classification of fault-induced voltage sags.

- The proposal of feature engineering approaches for fault-induced voltage sags classification based on the combination of transformations and feature extraction techniques.

The remainder of the paper is organized as follows. Section 2 describes the background on domain transformations and feature extraction techniques. Section 3 presents the methodology for generation of extensive synthetic signals through simulations in MATLAB/Simulink, and the implementation of transformations and feature extraction techniques. Section 4 compares quantitative and qualitative aspects of transformations and features. Finally, Section 5 highlights the main conclusions of the paper.

2. Theoretical background

2.1. Domain transformations

There are different types of mathematical transformations for PQD detection, classification, and characterization, including techniques in the time domain, frequency domain, and time-frequency domain.

- **Time-domain** analysis involves observing the evolution of signal features over time, providing information on the signal's shape, duration, and amplitude. They are used to analyze events such as sags, and transients [1]. Examples of time-domain techniques are Kalman filters, PSR, and SPM.

- **Frequency-domain** analysis involves decomposing steady-state signals into their frequency components. This technique provides information on the spectral characteristics of the signal. Fourier Transform (FT) and its variants, including DFT, are the main frequency-domain techniques.

- **Time-frequency-domain** analysis involves tracking the evolution of the frequency content of a signal over time. This approach is useful for analyzing non-stationary signals. Examples of time-frequency-domain techniques include the WT, STFT, and ST.

In this paper, the application of SPM, DFT, and STFT to voltage sags is studied. SPM is analyzed because it has shown promising results in terms of accuracy and real-time implementation for classification of fault induced voltage sags [6], [7], [9]**.** Likewise, DFT is relevant because of its robustness to noise which is essential for accurate and reliable classification of sags in real-life environments [10]. Moreover, STFT has shown benefits of both time and frequency domains by representing spectral and time-series features, for instance, it allows the identification of the high-frequency components in the transition segments of voltage sags [5]. A brief theoretical description of the transformations mentioned above is presented in the following.

2.1.1. Space Phasor Model (SPM)

The SPM is directly derived from Clarke's transformation [9]. The resulting SPM from the three phase-to-neutral voltages, $u_a(t)$, $u_b(t)$, $u_c(t)$, is then given by Equation (1), [9].

$$\underline{s}(t) = \frac{2}{3}[u_a(t) + \alpha u_b(t) + \alpha^2 u_c(t)];\ t = 1, \cdots, n, \quad (1)$$

where $\alpha = e^{j2\pi/3}$ and $\underline{s}(t)$ is a complex function.

2.1.2. Discrete Fourier Transform (DFT)

The DFT of a finite sequence of equally spaced samples of a time series can be calculated with the fast Fourier transform algorithm as expressed in Equation (2) [17].

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-\frac{j2\pi kn}{N}\right);\ k = 0, \cdots, N-1, \quad (2)$$

where $X(k)$ is a set of complex numbers, and N is the length of the discrete signal $x(n)$.

2.1.3. Short-Time Fourier Transform (STFT)

The STFT is obtained by applying the DFT to the signal over short time intervals and a sliding window of length W_L, which can be overlapping [5]. The DFT of each segment is stored in a matrix as shown in Equation (3).

$$STFT(X) = [X_0(k), X_1(k), \cdots, X_T(k)], \quad (3)$$

with,

$$X_f(k) = \sum_{n=-\infty}^{\infty} x(n) \cdot g(n - sR) e^{-j2\pi kn}, \quad (4)$$

where $X_f(k)$ is the DFT of the windowed signal centered at time sR, R is the distance between two adjacent windows and $g(n)$ is the window function.

2.2. Feature extraction techniques

The step of feature extraction refers to the computation of numerical indices (from transformation outputs or directly from the original waveforms) that are usable for tools in the decision space (detection, classification, and characterization) [1]. These indices comprise statistical, time series, spectral, and/or image-based features. It should be noted that there are limitations in the combination of transformations and features that can be extracted. For instance, the Total Harmonic Distortion

(THD), and the individual harmonics, which are spectral features, can only be extracted from frequency-domain, or time-frequency domain transformations. In this paper, statistical, time series, and spectral features are studied and briefly described in the following.

2.2.1. Statistical features

Statistical features include measures of dispersion of data such as maximum and minimum and statistical moments of different orders. For instance, lower-order statistical moments such as variance, and higher-order statistics such as skewness and kurtosis. These central statistical moments of order r, μ_r, are expressed in Equation (5).

$$\mu_r = \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^r, \tag{5}$$

where N is the number of samples x_n, $\bar{x}$ is the arithmetic mean, and r is the moment order. Variance (σ^2), skewness, and kurtosis are given by Equations (6), (7), and (8), respectively.

$$\sigma^2 = \mu_2, \tag{6}$$

$$Skew = \frac{\mu_3}{\sigma^3}, \tag{7}$$

$$Kurt = \frac{\mu_4}{\sigma^4}. \tag{8}$$

2.2.2. Time series features

Sample-based or cycle-based features can be extracted from the time series data of a discrete signal, i.e., directly from the original waveform [1]. Euclidean distance, also referred to as momentary deviation [23], is a sample-based technique and is expressed in Equation (9).

$$u_{dev}(t) = u_s(t) - u_{ref}(t);\ t = 1,2,\cdots,N, \tag{9}$$

where $u_s(t)$ is an instant value of the original discrete voltage signal of length N, and $u_{ref}(t)$ is the corresponding instant value of a reference signal which is usually a rated power frequency (e.g., 60 Hz) undistorted sinusoidal voltage waveform.

Euclidean distance can also be calculated for time-domain transformations such as the SPM. In such a case, it can be expressed as the distance to the origin in the complex plane as represented by Equation (10).

$$s_{dev}(t) = \left|\underline{s}(t)\right|. \tag{10}$$

2.2.3. Spectral features

Spectral features contain information on the frequency content of the signal, which is obtained with frequency- or time-frequency-domain transformations. For instance, magnitude and phase angle of individual harmonics, and THD as expressed in Equation (11), are spectral features.

$$THD = \frac{\sqrt{\sum_2^\infty U_h^2}}{U_1}, \tag{11}$$

where U_1 is the amplitude of the fundamental component of voltage, and U_h is the h^{th} harmonic amplitude.

3. Methodology and numerical simulations

According to [1], the process for PQD monitoring and analysis consist of eight steps, namely, (i) input data preparation, (ii) data preprocessing, (iii) transformation, (iv) feature extraction, (v) feature selection, (vi) detection, (vii) classification, and (viii) characterization. This paper deals with the first four stages as explained in the following. Furthermore, the study is limited to the analysis of fault-induced voltage sags and focused on the comparison of domain transformations in step (iii), feature extraction techniques in step (iv), and the performance of several combinations of transformations and feature extraction techniques.

In the first step, synthetic time-domain discrete signals are obtained from simulations of fault-induced voltage sags in MATLAB/Simulink using the scheme of Figure 1, adapted from [29]. The scheme presents a solidly grounded radial distribution network at the Medium-Voltage (MV) level of 13.2 kV. The network consists of two parallel feeders of 1 km and 10 km supplying passive aggregated residential loads of 100 kVA, and 1 MVA, respectively, with lagging power factor of 0.95.

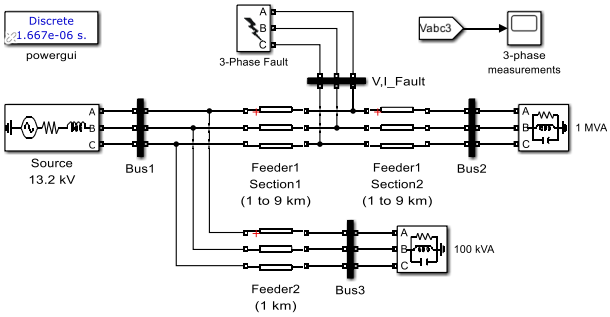

Figure 1. Schema in MATLAB/Simulink for the simulation of fault-induced voltage sags. Source: elaborated by the authors.

Ten types of sags are simulated, according to the type of fault as reported in Table 1. Similar identification performance is expected for the same groups of fault types, e.g., for all two-phase to ground faults. However, the distinction of the subtypes is made, i.e., abG, acG, and bcG, because the identification of the faulted phase(s) is useful for the proper operation of protection systems [7]. One hundred simulations are performed for each type varying randomly the location and duration of faults in Feeder1 for a total of 1,000 simulations. Thus, a database of 3,000 voltage signals (i.e., 1,000 for each phase) measured at Bus3 is generated. Only Bus3 is considered because end users at this bus would experience mainly voltage sags during faults in Feeder1, whereas end users at Bus2 would experience interruptions. A 1-μs time step is used for the simulation of ten cycles at 60 Hz of the voltage signals. Signals are resampled with a 10-μs time step because of storage issues. Therefore, each of the 3,000 voltage signals consists of 10,000 discrete values.

Table 1. Classification of sags according to fault type.

Fault type	Sag type
Three-phase	abc
Two-phase to ground (phases a, b)	abG
Two-phase to ground (phases a, c)	acG
Two-phase to ground (phases b, c)	bcG
Two-phase (phases a, b)	ab
Two-phase (phases a, c)	ac
Two-phase (phases b, c)	bc
Single-phase to ground (phase a)	aG
Single-phase to ground (phase b)	bG
Single-phase to ground (phase c)	cG

Source: elaborated by the authors.

Figure 2 shows examples of the simulated signals, illustrating the normalized three-phase voltages for a three-phase, and a two-phase to ground fault.

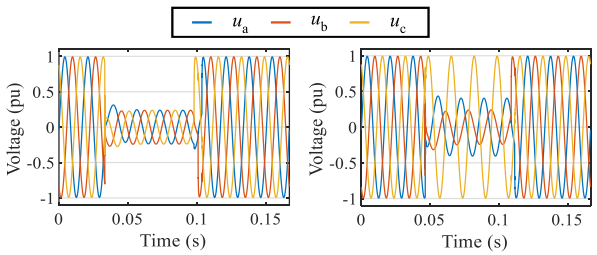

Figure 2. Synthetic sags caused by (a) a three-phase fault (abc), and (b) a two-phase to ground fault (abG). Source: elaborated by the authors.

The SPM, DFT, and STFT are applied to signals. Then, a set of statistical, time series, and spectral features are computed from transformation coefficients. A set of features is also extracted from the original time series signals to observe the results with no transformation. The combinations of applied transformations and extracted features are summarized in Figure 3.

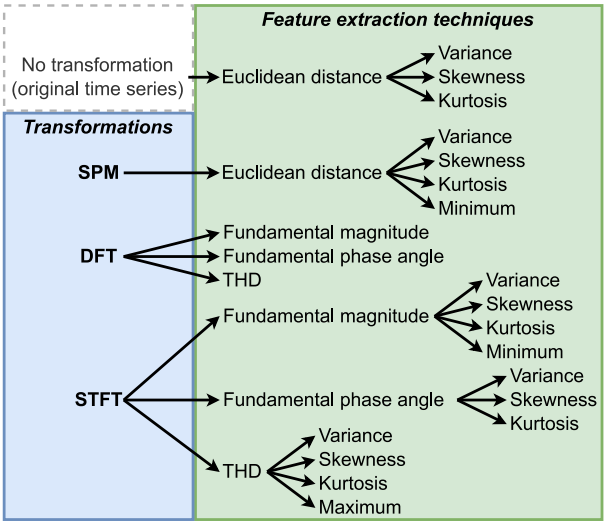

Figure 3. Applied transformations and features. Source: elaborated by the authors.

According to Figure 3, the Euclidean distance of the original time series is computed using Equation (9). Then, normalized variance, skewness and kurtosis of the computed Euclidean distance are calculated with Equations (6), (7), and (8), respectively. This process is applied to all synthetic signals of each of the phases, i.e., 3,000 signals (1,000 per phase), resulting in single values of variance, skewness, and kurtosis per signal, i.e., 3,000 values for each one. Figure 4 shows results of the process applied to the simulation case of Figure 2b.

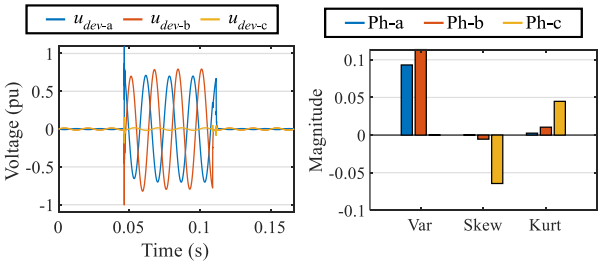

Figure 4. (a) Euclidean distance, (b) variance, skewness, and kurtosis for voltage signals of the fault in Figure 2b. Source: elaborated by the authors.

In the case of SPM, the signals of the three phases for each case simulated are converted to a single series of complex values using Equation (1), therefore, 1,000 series of complex values are obtained. The Euclidean distances for the 1,000 series are computed with Equation (10) resulting in 1,000 values for each dispersion measure (variance, skewness, kurtosis, and minimum). To illustrate the process, Figure 5 shows results of the two-phase to ground fault (abG) simulation case of Figure 2b.

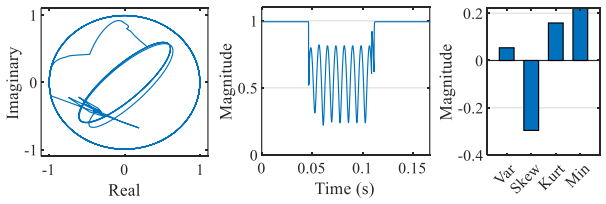

Figure 5. (a) SPM of the three phases, (b) corresponding Euclidean distance, and (c) variance, skewness, kurtosis, and minimum during the abG fault of Figure 2b. Source: elaborated by the authors.

Moreover, DFT is applied to signals of each of the phases with an interval of ten cycles and features are extracted as shown in Figure 3. A total of 3,000 values of fundamental magnitude, fundamental phase angle, and THD are calculated. Figure 6 shows an example with the computation of these features for the simulation case in Figure 2b. Fundamental magnitudes and phase angles of each phase are shown in Figure 6a.

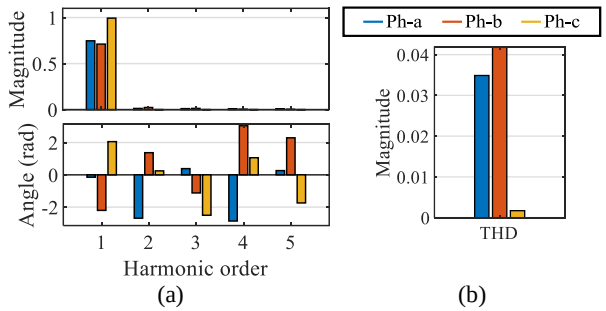

Figure 6. (a) DFT of each phase, magnitude (top), and phase angle (bottom), and (b) corresponding THD during the two-phase to ground fault (abG) of Figure 2b. Source: elaborated by the authors.

Finally, STFT is applied to signals of each of the phases, resulting in 3,000 matrices with information in the time and frequency domains using Equations (3), and (4). The window size is one cycle, i.e., 1,000 samples of the signal, and the overlap is half a cycle, i.e., the window hops over the original signal of ten cycles at intervals of R=500 samples, resulting in 19 computations of the DFT. The spectrograms for each of the phases in the simulation case of Figure 2b are shown in Figure 7, where changes in the spectrum are apparent during voltage sag initiation and ending. Figure 8 presents the evolution over time of the corresponding fundamental magnitude and phase angle, with the respective measures of dispersion. The evolution over time of the calculated THD is shown in Figure 9.

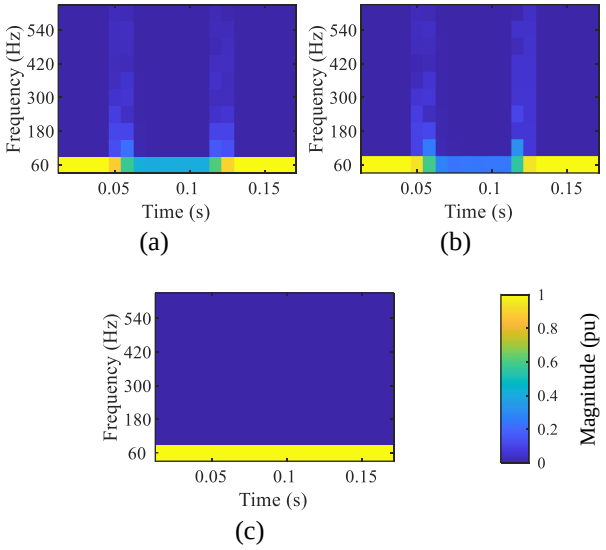

Figure 7. Spectrogram obtained with the STFT applied to synthetic signals of (a) phase a, (b) phase b, and (c) phase c during the abG fault of Figure 2b. Source: elaborated by the authors.

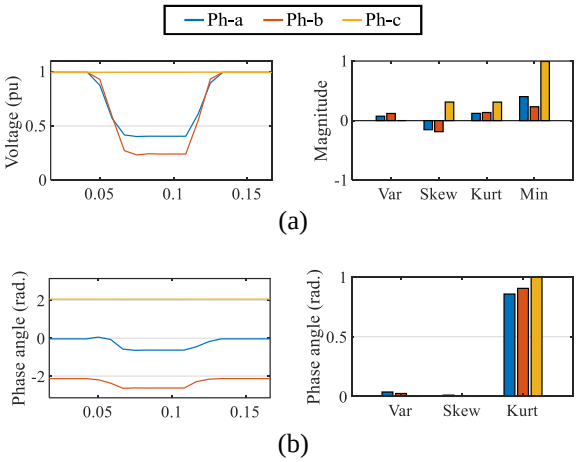

Figure 8. (a) Fundamental magnitude and (b) phase angle over time, obtained with the STFT applied to signals in Figure 2b (left), and dispersion measures (right). Source: elaborated by the authors.

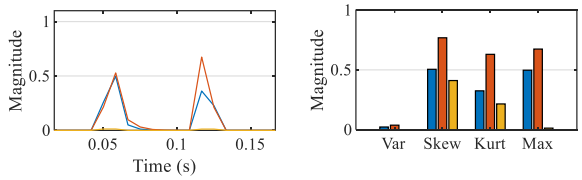

Figure 9. THD over time obtained with the STFT applied to signals in Figure 2b (left), and corresponding dispersion measures (right). Source: elaborated by the authors.

4. Comparison and results

The combinations of domain transformations and extracted features in Figure 3 are compared according to quantitative aspects including the computation time, storage requirement, and linear separability. Qualitative aspects such as physical interpretation, and suitability for fault-induced voltage sag characterization are also analyzed.

4.1. Computation time

The computation time reported in Table 2 is the median of 100 repeated calculations for each transformation and feature extraction technique, executed on a CPU with a six-core processor, 2.20 GHz, and 12 GB RAM. The computation time covers the application of techniques on the 1,000 fault simulation cases, i.e., 3,000 voltage signals (1,000 per phase). The table shows that SPM is the fastest, and STFT is the slowest of the applied transformations, because SPM uses simple operations in a sample-based manner, whereas STFT calculates and stores the DFT on a sliding window which requires more calculation steps.

Furthermore, Table 2 shows that the calculations of the Euclidean distances of the original time series and SPM, and THD from DFT have reduced durations of similar order. Computations of fundamental magnitudes and phase angles from DFT and STFT are performed in 1 ms or less, whereas THD from STFT takes 257 ms. Computations of dispersion measures take longer for Euclidean distances (up to 915 ms) because more samples must be evaluated, whereas spectral features from STFT are calculated only 19 times per signal. The total computation times demonstrates that transformation and features extraction is faster with SPM, followed by DFT and no transformation, whereas the STFT is the slowest.

Table 2. Computation time for transformation and feature extraction.

Transformation		**Feature extraction 1**		**Feature extraction 2**		**Total computation time (s)**	
Technique	**Computation time (s)**	**Feature**	**Computation time (s)**	**Feature**	**Computation time (s)**		
None	--	Euclidean distance	0.058	Variance	0.074	1.897	
				Skewness	0.850		
				Kurtosis	0.915		
SPM	0.071	Euclidean distance	0.033	Variance	0.033	0.654	
				Skewness	0.294		
				Kurtosis	0.290		
				Minimum	0.004		
DFT	0.936	Fundamental magnitude	~ 0	--	--	0.915	
		Fundamental angle	~ 0	--	--		
		THD	0.015	--	--		
STFT	1.986	Fundamental magnitude	0.001	Variance	~ 0	1.99	2.27
				Skewness	0.002		
				Kurtosis	0.002		
				Minimum	~ 0		
		Fundamental angle	0.001	Variance	0.001	2.00	
				Skewness	0.005		
				Kurtosis	0.006		
		THD	0.257	Variance	~ 0	2.25	
				Skewness	0.002		
				Kurtosis	0.002		
				Maximum	~ 0		

Source: elaborated by the authors.

4.2. Storage requirement

Table 3 reports the size of matrices/vectors with data obtained after transformations and feature extraction techniques for one fault simulation case and a signal length of ten power cycles. In terms of application, this storage capability would be the requirement for a device in the field to perform the algorithms applied in this paper. For instance, SPM results in a 1 x 10,000 vector per fault simulated, whereas STFT results in a 3 x 501 x 19 matrix per fault simulated, where dimensions correspond to phase, frequency data, and time data, respectively. After feature extraction with dispersion measures, SPM requires the smallest storage for four values (which would be the number of features used in a subsequent step, e.g., classification), whereas STFT requires the biggest for 11 vectors with size 3 x 1.

4.3. Linear separability

Linear separability of features is tested with the implementation of a linear support vector machine [23] in the classification learner app in MATLAB/Simulink. To that end, the dataset of synthetic signals from the 1,000 simulated cases is used as input. Then, a holdout validation dataset with 30% of the signals is defined randomly but considering an equal proportion to each type of sag. The linear SVM is automatically trained and tested in the classification learner app. If the features included are linearly separable for classification of fault-induced voltage sags according to Table 1, the results of accuracy tend to be 100%.

Results of the classification are reported in Table 4 where very good accuracy is observed in all cases except SPM. In that case, using only the Euclidean distance of SPM causes loss of information to distinguish between two-phase and single-phase faults. However, SPM provide more interesting image-based features [9] that are not analyzed in this paper. Spectral features provide useful information for sag classification, especially, THD. In the case of STFT, the individual performance of the dispersion measures applied only to the fundamental magnitudes is assessed resulting in a very satisfactory 99.9% accuracy. Similarly, performance of the dispersion measures of the fundamental phase angle and THD are also analyzed resulting in 96.3% and 100% accuracy. Finally, the combined performance of all dispersion measures obtained with STFT is assessed resulting also in 100% accuracy.

Table 3. Storage requirement of transformations and features.

Transformation		**Feature extraction 1**		**Feature extraction 2**	
Technique	**Matrix size**	**Feature**	**Matrix size**	**Feature**	**Matrix size**
None	3 x 10,000	Euclidean distance	3 x 10,000	Variance	3 x 1
				Skewness	3 x 1
				Kurtosis	3 x 1
SPM	1 x 10,000	Euclidean distance	1 x 10,000	Variance	1 x 1
				Skewness	1 x 1
				Kurtosis	1 x 1
				Minimum	1 x 1
DFT	3 x 5,001	Fundamental magnitude	3 x 1	--	--
		Fundamental angle	3 x 1	--	--
		THD	3 x 1	--	--
STFT	3 x 501 x 19	Fundamental magnitude	3 x 19	Variance	3 x 1
				Skewness	3 x 1
				Kurtosis	3 x 1
				Minimum	3 x 1
		Fundamental angle	3 x 19	Variance	3 x 1
				Skewness	3 x 1
				Kurtosis	3 x 1
		THD	3 x 19	Variance	3 x 1
				Skewness	3 x 1
				Kurtosis	3 x 1
				Maximum	3 x 1

Source: elaborated by the authors.

Table 4. Linear separability of the extracted features.

Transformation	Feature extraction 1	Feature extraction 2	Accuracy (%)	
None	Euclidean distance	Variance, Skewness, Kurtosis	97.4	
SPM	Euclidean distance	Variance, Skewness, Kurtosis, Minimum	27.6	
DFT	Fundamental magnitude, Fundamental angle, THD	--	100	
STFT	Fundamental magnitude	Variance, Skewness, Kurtosis, Minimum	99.9	100
	Fundamental angle	Variance, Skewness, Kurtosis	96.3	
	THD	Variance, Skewness, Kurtosis, Maximum	100	

Source: elaborated by the authors.

4.4. Physical interpretation

Interpretability of features is very useful to obtain a better understanding of the voltage sag phenomenon to provide better applications of the analysis. For instance, duration and magnitude are relevant features for sag impact assessment. In that sense, STFT provide very detailed information as can be seen in spectrograms of Figure 7 where sag initiation and ending can be identified with changes in the spectral content and the fundamental magnitude. Moreover, SPM shows potentialities in image-based features (see Figure 5a).

It should be noted that physical features such as duration and magnitude in voltage sags can be obtained through well-stablished and widely accepted methods indicated in standards, e.g., IEC 61000-4-30. Therefore, the implementation of this capability would be very valuable in proposed detection, classification, and characterization methodologies to comply with standards. Moreover, to analyze these capabilities in different transformation and feature extraction techniques is relevant because they can provide enhanced performance compared to standard algorithms. For instance, DFT and STFT are robust against noise, and SPM has demonstrated improved capabilities for the identification of voltage sag characteristics such as phase angle jump [30], and shape in multistage voltage sags [31].

4.5. Suitability for voltage sag characterization

Results of dispersion measures applied in the Euclidean distance of the original time series show very good performance for sag characterization and classification. However, the interpretation of features is not straightforward, and the storage requirements can hinder the real-time application in embedded systems. Despite the Euclidean distance of the SPM shows inaccurate classification results, potentialities are observed in the use of the image-based features that are interpretable and seemed to provide better classification performance [9].

Furthermore, DFT shows that frequency-domain information provides a very good performance in terms of classification but limited interpretable features. That drawback is overcome with the implementation of time-frequency domain transformations such as STFT, where very useful information in both time and frequency domains is obtained at the cost of higher computational time and storage requirements.

4.6. Summary and comparative tables

To summarize the results of the comparison discussed above, Table 5 presents advantages and drawbacks of each of the analyzed transformations in the time domain, i.e., none, SPM; the frequency domain, i.e., DFT; and the time-frequency domain, i.e., STFT. In contrast to the comparison presented in the previous tables, Table 5 intends to compare qualitative aspects of the individual domain transformations and not the combined performance. It should also be noted that the comparison is made in the context of the work developed in this paper, therefore, the observations are obtained from the simulations and algorithms implemented, as well as for the specific application of detection, classification, and characterization of fault-induced voltage sags.

Similarly, Table 6 reports advantages and drawbacks of the time-series, spectral, and statistical feature extraction techniques studied in this work.

Some examples in the literature illustrate the specific advantages of each domain transformation and feature extraction technique. For instance, SPM is useful for real-time applications including the operation of protection systems and the implementation of mitigation measures such as dynamic voltage restorer operation [6], [7]**.** Moreover, robustness of DFT and STFT, and spectral and statistical features is relevant in the application of detection, classification, and characterization of fault-induced voltage sags in noisy environments [5], [10].

Table 5. Comparison of domain transformations.

Domain	Transformation	Advantages	Drawbacks
Time	None	- No computational burden - Simplicity - Domain transformation knowledge is not required - No loss of information	- Involve more complex subsequent decision space - Lack of spectral information - High storage requirement
	SPM	- Low computational burden - Simplicity - Low loss of information - Capability of representing three-phase voltages in one signal - It provides useful image-based features	- High/medium storage requirement - Lack of spectral information - Lack of single-phase information
Frequency	DFT	- Reduced storage requirement - Higher resolution of spectral information (compared to the application of one-cycle-based STFT)	- Loss of time information - Inaccurate for non-periodic signals - Medium/high computational burden - It is prone to spectral leakage and windowing
Time-frequency	STFT	- Flexibility in window selection - It provides information in both time and frequency domains - Robust to noise	- Complexity - Computational burden - Trade-off between time and frequency resolution - It is prone to spectral leakage and windowing

Source: elaborated by the authors.

5. Conclusion

This paper provides a comparative study of domain transformations and feature extraction techniques applied to synthetic signals of fault-induced voltage sags aimed at providing input data for subsequent steps in the process of detection, classification, and characterization of the disturbance. Domain transformations including the Space Phasor Model (SPM), Discrete Fourier Transform (DFT), and Short-Time Fourier Transform (STFT) are implemented, and feature extraction techniques such as Euclidean distance, variance, skewness, kurtosis, minimum, and maximum are applied. The results show that Euclidean distance of the original signal provides very good and simple performance at the cost of relatively high storage requirement, and computation time. Similarly, STFT provides high linear separability and physical interpretability of features using more storage and computation time. DFT allows accurate classification results with reduced storage requirement and low/medium computational burden, however, physical interpretability of features is very limited. Moreover, Euclidean distance of the SPM shows loss of information for classification, but image-based features, that are not analyzed in detail in this paper, seem promising for characterization and classification of fault-induced voltage sags.

Based on the conclusions mentioned above, accurate and simple classifiers, i.e., based on Euclidean distance of the original waveform + dispersion measures, and DFT + spectral features are suitable for real-time applications where only classification is required (limited interpretability of features). In addition, a more complex approach, e.g., STFT + spectral features + dispersion measures, is useful for robust and reliable classification (required, e.g., in noisy environments caused by different disturbances), and characterization when physical interpretability is needed.

Finally, the application of SPM in combination with complementary transformation steps and additional feature extraction techniques applied to the resulting image-based features in the complex plane (see Figure 5a) seem to be a promising future work. Research in this direction has been found in the literature related to sags [6], [7], [9], but a more comprehensive methodology is still required to provide results in all steps of the decision space, i.e., detection, classification, and characterization.

Funding

This work was supported by Centro de Investigaciones y Desarrollo Científico (CIDC) of the Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.

Table 6. Comparison of feature extraction techniques.

Type	Feature	Advantages	Drawbacks
Time-series	Euclidean distance	- Low computational burden - Simple and intuitive - Robust to noise - Suitable for low-dimensional spaces	- Sensitive to scale and normalization. - Limited in high-dimensional spaces. - It may not capture complex relationships
Spectral	Fundamental magnitude	- No computational burden - Robust to noise - High capability for analysing sags	- No time information - Lack of location information
	Fundamental phase angle	- No computational burden - Robust to noise - Phase information during events	- Lack of time information - No information on event severity
	THD	- Low computational burden - Spectral information on a wide range of frequencies - Robust to noise	- No time information - Sensitive to signal bandwidth - It may not capture transient behaviour (sag start/end) - Limited dynamic range
Statistical	Variance	- It captures signal dispersion - Easy to interpret - Sensitive to signal amplitude changes - Robust to outliers	- Limited information - May not capture shape information - Limited dynamic range
	Skewness	- Captures signal asymmetry - Sensitive to changes in signal shape - Easy to interpret	- Limited information - Sensitive to outliers
	Kurtosis	- Captures signal shape information - Can differentiate between signal types - Sensitive to changes in signal shape - Provides complementary information to other measures	- Limited information - Sensitive to outliers
	Minimum / Maximum	- Provide information about the lowest and highest points - Easy to compute - Can be used to detect outliers - Complement other measures	- Limited information - Sensitive to noise - May not capture information about signal dynamics

Source: elaborated by the authors.

Declaration of competing interests

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

References

[1] J. E. Caicedo, D. Agudelo-Martínez, E. Rivas-Trujillo, and J. Meyer, "A systematic review of real-time detection and classification of power quality disturbances," *Prot. Control Mod. Power Syst.*, vol. 8, no. 1, p. 3, 2023, doi: 10.1186/s41601-023-00277-y.

[2] R. Igual and C. Medrano, "Research challenges in real-time classification of power quality disturbances applicable to microgrids: A systematic review," *Renew. Sustain. Energy Rev.*, vol. 132, p. 110050, Oct. 2020, doi: 10.1016/j.rser.2020.110050.

[3] J. Blanco-Solano, N. Kagan, C. F. M. Almeida, J. F. Petit-Suárez, and G. Ordóñez-Plata, "Voltage sag state estimation based on -normminimization methods in radial electric power distribution systems," *Rev. UIS Ing.*, vol. 17, no. 2, pp. 309–318, Feb. 2018, doi: 10.18273/revuin.v17n2-2018026.

[4] R. Bhujade, S. Maharjan, A. M. Khambadkone, and D. Srinivasan, "Economic analysis of annual load loss due to voltage sags in industrial distribution networks with distributed PVs," *Sol. Energy*, vol. 252, pp. 363–372, 2023, doi: 10.1016/j.solener.2023.01.041.

[5] M. Veizaga, C. Delpha, D. Diallo, S. Bercu, and L. Bertin, "Classification of voltage sags causes in industrial power networks using multivariate time-series," *IET Gener. Transm. Distrib.*, 2023, doi: 10.1049/gtd2.12765.

[6] M. R. Alam, K. M. Muttaqi, and T. K. Saha, "Classification and Localization of Fault-Initiated Voltage Sags Using 3-D Polarization Ellipse Parameters," *IEEE Trans. Power Deliv.*, vol. 35, no. 4, pp. 1812–1822, 2020, doi: 10.1109/TPWRD.2019.2954857.

[7] M. R. Alam, F. Bai, R. Yan, and T. K. Saha, "Classification and Visualization of Power Quality Disturbance-Events Using Space Vector Ellipse in Complex Plane," *IEEE Trans. Power Deliv.*, vol. 36, no. 3, pp. 1380–1389, 2021, doi: 10.1109/TPWRD.2020.3008003.

[8] Y. Pu, H. Yang, X. Ma, and X. Sun, "Recognition of voltage sag sources based on phase space reconstruction and improved VGG transfer learning," *Entropy*, vol. 21, no. 10, 2019, doi: 10.3390/e21100999.

[9] A. Bagheri, I. Y. H. Gu, M. H. J. Bollen, and E. Balouji, "A Robust Transform-Domain Deep Convolutional Network for Voltage Dip Classification," *IEEE Trans. Power Deliv.*, vol. 33, no. 6, pp. 2794–2802, 2018, doi: 10.1109/TPWRD.2018.2854677.

[10] V. A. Katic and A. M. Stanisavljevic, "Smart Detection of Voltage Dips Using Voltage Harmonics Footprint," *IEEE Trans. Ind. Appl.*, vol. 54, no. 5, pp. 5331–5342, 2018, doi: 10.1109/TIA.2018.2819621.

[11] M. Upadhya, A. K. Singh, P. Thakur, E. A. Nagata, and D. D. Ferreira, "Mother wavelet selection method for voltage sag characterization and detection," *Electr. Power Syst. Res.*, vol. 211, 2022, doi: 10.1016/j.epsr.2022.108246.

[12] M. K. Saini and A. Aggarwal, "Fractionally delayed Legendre wavelet transform based detection and optimal features based classification of voltage sag causes," *J. Renew. Sustain. Energy*, vol. 11, no. 1, 2019, doi: 10.1063/1.5049189.

[13] M. K. Saini, R. Kapoor, R. K. Beniwal, and A. Aggarwal, "Recognition of voltage sag causes using fractionally delayed biorthogonal wavelet," *Trans. Inst. Meas. Control*, vol. 41, no. 10, pp. 2851–2863, 2019, doi: 10.1177/0142331218814292.

[14] H. Ramírez-Murillo, C. A. Torres-Pinzón, E. Forero-García, and A. Álzate-Gómez, "Metodología de diagnóstico de perturbaciones en la calidad de potencia usando la Transformada S," *Rev. UIS Ing.*, vol. 20, no. 2, Feb. 2021, doi: 10.18273/revuin.v20n2-2021010.

[15] M. Mishra and R. R. Panigrahi, "Advanced signal processing and machine learning techniques for voltage sag causes detection in an electric power system," *Int. Trans. Electr. Energy Syst.*, vol. 30, no. 1, pp. 1–16, Jan. 2020, doi: 10.1002/2050-7038.12167.

[16] B. Patnaik, R. R. Panigrahi, M. Mishra, R. K. Jena, and M. kumar Swain, "Detection and Classification of Voltage Sag Causes Based on S-Transform and Extreme Learning Machine," in *Lecture Notes in Electrical Engineering*, vol. 630, 2020, pp. 277–289.

[17] J. Li, Y. Yang, H. Lin, Z. Teng, F. Zhang, and Y. Xu, "A voltage sag detection method based on modified s transform with digital prolate spheroidal window," *IEEE Trans. Power Deliv.*, vol. 36, no. 2, pp. 997–1006, 2021, doi: 10.1109/TPWRD.2020.2999693.

[18] Ç. Kocaman and M. Özdemir, "Determining five kinds of power quality disturbances by using statistical methods and wavelet energy coefficients," *Renew. Energy Power Qual. J.*, vol. 1, no. 15, pp. 745–750, 2017, doi: 10.24084/repqj15.455.

[19] W. L. Rodrigues Junior, F. A. S. Borges, R. de A. L. Rabelo, J. J. P. C. Rodrigues, R. A. S. Fernandes, and I. N. Silva, "A methodology for detection and classification of power quality disturbances using a real-time operating system in the context of home energy management systems," *Int. J. Energy Res.*, vol. 45, no. 1, pp. 203–219, Jan. 2021, doi: 10.1002/er.5183.

[20] J. Xiu, X. Guangye, M. Xiangping, and D. Guilin, "Voltage sag detection method based on dq transform and complex wavelet transform," in *2021 IEEE International Conference on Electrical Engineering and Mechatronics Technology, ICEEMT 2021*, 2021, pp. 429–434, doi: 10.1109/ICEEMT52412.2021.9602691.

[21] N. M. Khoa and D. D. Tung, "An extended Kalman filter for detecting voltage sag events in power systems," *J. Electr. Syst.*, vol. 14, no. 2, pp. 192–204, 2018, [Online]. Available: https://www.proquest.com/openview/17dc530e46e9e2bc3226e3dc476e9119/1?pq-origsite=gscholar&cbl=4433095.

[22] E. A. Nagata *et al.*, "Real-time voltage sag detection and classification for power quality

diagnostics," *Meas. J. Int. Meas. Confed.*, vol. 164, 2020, doi: 10.1016/j.measurement.2020.108097.

[23] R. Qi, O. Zyabkina, D. Agudelo-Martínez, and J. Meyer, "Automatic detection of voltage notches using support vector machine," *Renew. Energy Power Qual. J.*, vol. 19, pp. 528–533, 2021, doi: 10.24084/repqj19.337.

[24] L. Lin, D. Wang, S. Zhao, L. Chen, and N. Huang, "Power quality disturbance feature selection and pattern recognition based on image enhancement techniques," *IEEE Access*, vol. 7, pp. 67889–67904, 2019, doi: 10.1109/ACCESS.2019.2917886.

[25] A. M. Stanisavljević, V. A. Katić, B. P. Dumnić, and B. P. Popadić, "A Comprehensive Overview of Digital Signal Processing Methods for Voltage Disturbance Detection and Analysis in Modern Distribution Grids with Distributed Generation," *Acta Polytech. Hungarica*, vol. 16, no. 5, pp. 125–149, Aug. 2019, doi: 10.12700/APH.16.5.2019.5.8.

[26] P. Ray, G. K. Budumuru, and B. K. Mohanty, "A comprehensive review on soft computing and signal processing techniques in feature extraction and classification of power quality problems," *J. Renew. Sustain. Energy*, vol. 10, no. 2, Mar. 2018, doi: 10.1063/1.5006772.

[27] Rahul, "Review of Signal Processing Techniques and Machine Learning Algorithms for Power Quality Analysis," *Adv. Theory Simulations*, vol. 3, no. 10, Oct. 2020, doi: 10.1002/adts.202000118.

[28] M. Mishra, "Power quality disturbance detection and classification using signal processing and soft computing techniques: A comprehensive review," *Int. Trans. Electr. Energy Syst.*, vol. 29, no. 8, Aug. 2019, doi: 10.1002/2050-7038.12008.

[29] J. Caicedo, F. Navarro, E. Rivas, and F. Santamaria, "Voltage sag characterization with Matlab/Simulink," in *2012 Workshop on Engineering Applications*, May 2012, pp. 1–6, doi: 10.1109/WEA.2012.6220073.

[30] J. R. Camarillo-Penaranda and G. Ramos, "Fault classification and voltage sag parameter computation using voltage ellipses," *IEEE Trans. Ind. Appl.*, vol. 55, no. 1, pp. 92–97, 2019, doi: 10.1109/TIA.2018.2864108.

[31] A. Bagheri, M. H. J. Bollen, and I. Y. H. Gu, "Improved characterization of multi-stage voltage dips based on the space phasor model," *Electr. Power Syst. Res.*, vol. 154, pp. 319–328, 2018, doi: 10.1016/j.epsr.2017.09.004.