

Spectral Variation-Based Compression of Three-Phase Signals

Mercedes Diniz, Antonio Martins, Flávio Nunes, Adalbery Castro and Leonardo Ramalho

Abstract—This work adapts a spectral analysis-based compression algorithm, originally designed for single-phase leakage current signals, for compression of three-phase electrical signals with real disturbances. A PCA-based preprocessing step is introduced to reduce dimensionality and improve compression efficiency. The method is evaluated using real signals containing representative power quality events. The results show that the proposed approach achieves higher compression ratios, for instance, in some scenarios exceeding 118:1 and 76:1 for voltage and current signal, respectively. The reconstruction quality was comparable to the original algorithm, indicating its suitability for efficient storage and transmission in smart grid applications.

Keywords—Digital signal processing, compression, power systems.

I. INTRODUCTION

Emerging applications in Smart Grids, such as power quality analysis, fault detection, and high-resolution state estimation, require access to raw current signals sampled at high rates. While synchrophasors reduce data dimensionality by representing quasi steady-state conditions, they are inherently limited in capturing transient phenomena, harmonics, and non-linear distortions that are critical for modern grid analytics [1].

This shift toward waveform-level monitoring significantly increases data generation, as waveform acquisition typically operates at kHz-level sampling frequencies [2]. In this context, efficient compression of current signals becomes essential to enable scalable data transmission and storage. Although compressing raw current waveforms provides a more complete representation of system behavior, it introduces significant challenges due to high data rates and stringent fidelity requirements.

Moreover, power systems inherently operate with three-phase signals, which exhibit strong inter-phase correlations under both balanced and moderately unbalanced conditions. Treating each phase independently neglects this intrinsic structure and leads to suboptimal compression performance. Joint compression of three-phase current signals can exploit spatial redundancy across phases, improving compression efficiency while preserving essential electrical characteristics. This is particularly important for applications involving fault analysis, imbalance detection, and protection systems, where the relative behavior among phases carries critical information.

In this context, compression methods based on spectral analysis have been proposed, such as Spectral Variation Detection and Compression System (SVDCS) [3] and its variation

the Spectral Variation Detection Algorithm (SVDA) [4]. They have demonstrated an effective solution for monitoring single-phase current signals. Based on spectral variation detection and the exploitation of Hermitian symmetry to reduce memory usage, SVDA was specifically designed for resource-constrained Internet of Things (IoT) devices, overcoming limitations of more computationally demanding algorithms.

This work proposes an adaptation of the SVDA algorithm for three-phase signals, introducing a new preprocessing stage, the Principal Component Analysis (PCA) to reduce spatial redundancy in multi-phase signals. As validated in [5], this dimensionality reduction is essential to balance high compression ratios with the reconstruction fidelity.

This modification aims to enable efficient spectral novelty detection in multi-phase systems while preserving the reduced computational burden and high reconstruction fidelity that characterize the original approach.

II. MATERIALS AND METHODS

The proposed method was implemented and evaluated in MATLAB, providing greater flexibility for algorithm implementation and performance analysis, while maintaining conceptual consistency with the original approach. This differs from the implementations in [3] and [4], which target embedded hardware platforms, specifically an Field Programmable Gate Array (FPGA) and a microcontroller-based system, respectively. The following sections describe the investigated signal and scenarios, followed by the proposed compression method.

A. Analyzed Signals

The experimental evaluation of the proposed method is based on real oscillographic data obtained from the publicly available dataset¹ described in [6]. The original dataset contains three-phase electrical signals acquired from power systems under different operating conditions, including both normal operation and a variety of power quality disturbances. This dataset is organized using the Common Format for Transient Data Exchange (COMTRADE) standard, specifically following the IEEE C37.111-2013 ASCII format. In this structure, each record is composed of a pair of files: a `.cfg` file containing configuration and metadata, and a `.dat` file storing the corresponding raw signal samples.

Due to the presence of multiple channels from different buses within a single COMTRADE record, a preprocessing

LASSE – 6G & IoT Research Group, Federal University of Pará, Belém, Pará 66075-750, Brazil. E-mails: {mercedes.diniz, antonio.martins, flavio.nunes}@itec.ufpa.br, {adalbery, leonardolr}@ufpa.br.

¹Dataset available at: <https://doi.org/10.6084/m9.figshare.28465427>

step was performed to extract only the signals associated with the disturbances of interest. These selected signals were then reorganized and stored in binary format to facilitate efficient processing and reproducibility of the experiments. The processed dataset, including both the filtered signals and the original COMTRADE files, has been made publicly available for reproducibility purposes².

From this dataset, a representative subset of signals was selected in order to cover distinct types of events relevant to power quality analysis. The selection includes voltage sags during motor start, current fluctuations, waveform distortions, phase sequence errors, fault conditions, and measurement-related noise. Table I summarizes the selected signals and their associated disturbance types. For some events, both current (I) and voltage (V) signals are considered, reflecting the multi-dimensional nature of three-phase systems. This selection defines the scope of the study, focusing on realistic and representative scenarios commonly encountered in modern power networks. In other words, the proposed method can be applied to voltage or current three-phase signals.

The selected records have a sampling frequency of 1600 Hz and a fundamental frequency of 50 Hz, resulting in file sizes ranging from approximately 122 kB to 438 kB. Although a single oscillographic record may occupy only a few hundred kilobytes, utilities routinely store thousands of disturbance records over long periods. For example, the public dataset adopted in this work contains more than 50.000 real oscillographic records, making efficient compression important for long-term storage, transmission, and cloud-based analysis.

TABLE I: Tested signals and their characteristics

#	Description of events	Type
1	Motor start voltage drop (<95%)	I
2		V
3	Motor start current fluctuations	I
4		V
5	Frequency-driven motor fluctuations	I
6		V
7	Voltage waveform distortion	I
8		V
9	Phase sequence error	I
10		V
11	Current phase sequence error	I
12	Two-phase-to-ground fault (3U0, voltage drop)	V
13	Sensor noise due to measurement fault	V

B. Proposed Compression Method

The proposed Spectral Variation Detection Algorithm for Three-Phase Signals (SVDA-3P) is shown in Fig. 1. It is an evolution of the previous SVDA [4] and SVDCS [3], adapted for three-phase signals. The proposed approach incorporates a PCA-based preprocessing stage to reduce the dimensionality of the three-phase signal prior to compression. Originally

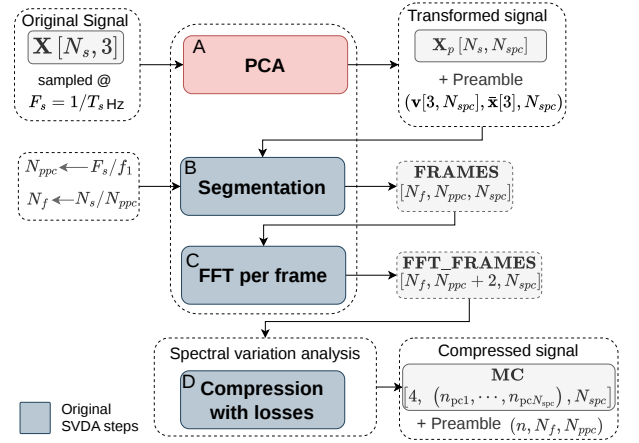


Fig. 1: Overview of the proposed algorithm with four steps, where the PCA is the main improvement in the previous algorithm.

introduced in [7], PCA is a multivariate statistical technique that transforms a set of possibly correlated variables into a new set of orthogonal components ordered according to their variance.

Although transformations such as the Clarke transform can provide a more compact representation of three-phase signals through a fixed projection matrix, they do not adapt to the statistical properties of the data. In contrast, PCA computes a data-driven orthogonal basis that maximizes the variance preserved in the retained principal components [8] while decorrelating the signal channels and reducing their linear redundancy [9].

The subsequent stages (highlighted in gray in Fig. 1) follow the original framework proposed in [4]: segmentation of the signal into smaller frames, calculation of the Fast Fourier Transform (FFT) of each frame, and removal of coefficients based on the energy of consecutive FFT frames. During signal reconstruction, all corresponding inverse operations are performed.

In this work, the three-phase signal is represented as a data matrix $\mathbf{X} \in \mathbb{R}^{N_s \times 3}$, where N_s is the number of samples and each column corresponds to one phase. In order to focus the analysis on variance, the data is first centered by subtracting the mean of each phase

$$x_{ij}^{\text{cent}} = x_{ij} - \bar{x}_j, \quad (1)$$

where x_{ij} denotes the sample at time index $i \in \{1, \dots, N_s\}$ and phase $j \in \{1, 2, 3\}$, and $\bar{x}_j = \frac{1}{N_s} \sum_{i=1}^{N_s} x_{ij}$.

From the centered data, the covariance matrix $\mathbf{C} \in \mathbb{R}^{3 \times 3}$ is computed, capturing the correlation between phases. The eigenvalue decomposition of $\mathbf{C}$ is then performed, yielding the eigenvalues λ_i and corresponding eigenvectors $\mathbf{v}_i \in \mathbb{R}^{3 \times 1}$ for i -th component. The eigenvalues represent the variance explained by each principal component, while the eigenvectors define the projection directions.

To reduce dimensionality, only the first N_{spc} components are retained according to an energy criterion. Specifically, the smallest number of components is selected such that the

²Dataset available at: <https://nextcloud.lasseufpa.org/s/iJtF9Zr9oZBsPbM>

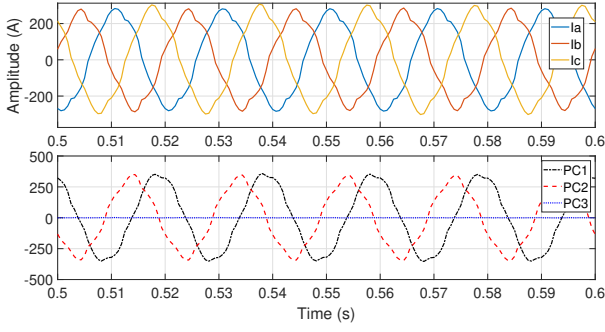


Fig. 2: Original current signal 11 from Table I and its principal components decomposition.

cumulative explained variance satisfies:

$$\frac{\sum_{i=1}^{N_{spc}} \lambda_i}{\sum_{i=1}^3 \lambda_i} \geq 0.9. \quad (2)$$

In practice, this results in retaining one or two eigenvectors, while the remaining eigenvectors associated with low variance are discarded.

The eigenvalues are sorted in descending order, and the centered signal is projected onto the new basis defined by the selected eigenvectors. Let $\tilde{\mathbf{V}} \in \mathbb{R}^{3 \times N_{spc}}$ denote the matrix formed by the first N_{spc} eigenvectors associated with the largest eigenvalues. The projection of the original three-phase signal onto the reduced subspace is then given by $\mathbf{X}_p = \mathbf{X}_c \cdot \tilde{\mathbf{V}}$, where the elements of $\mathbf{X}_c$ are given by (1).

This stage outputs a reduced data matrix $\mathbf{X}_p \in \mathbb{R}^{N_s \times N_{spc}}$, where $1 \leq N_{spc} \leq 3$ represents the number of retained principal components. This dimensionality reduction step decreases the number of channels to be processed in subsequent stages, reducing computational complexity and data redundancy while preserving most of the signal energy. Fig. 2 shows the original signal and its Principal Components (PCs) after PCA. The sinusoidal behavior of the signal is preserved in PC1 and PC2, whereas PC3 contains low energy and can therefore be discarded without significantly affecting the quality of the reconstructed signal.

The transformation parameters (mean vector and selected eigenvectors) are stored and transmitted as side information, enabling signal reconstruction at the decoder, where all inverse operations are performed. In Fig. 1, these parameters are represented in the preamble. Specifically, $\tilde{\mathbf{V}} \in \mathbb{R}^{3 \times N_{spc}}$ contains the eigenvectors that define the projection basis, and $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \bar{x}_3]$ stores the mean values of the original three-phase signal.

After dimensionality reduction, each column of $\mathbf{X}_p$ is segmented into N_f frames, as shown in Stage B of Fig. 1. Each frame contains N_{ppc} samples. In stage C, the FFT is applied to each frame. Since the input signals are real-valued, only $(N_{ppc}/2) + 1$ complex coefficients are retained due to the conjugate symmetry property, as illustrated in the output of block C. These coefficients are stored by separating their real and imaginary parts, resulting in $N_{ppc} + 2$ real-valued elements per frame, as detailed in Fig. 3.

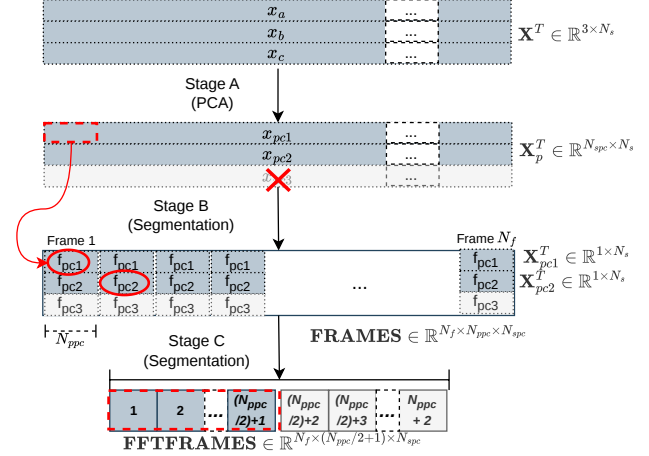


Fig. 3: Example of the data structure in the signal transformation stages for $N_{spc} = 2$.

The final step (D) of the algorithm consists of lossy compression applied to the FFT coefficients of the frames, referred to as *FFT_FRAMES*. Fig. 3 illustrates how the data structure evolves from the original three-phase signal to the input of this step.

With the data previously segmented and transformed into the frequency domain, spectral analysis is performed over all FFT frames for each principal component. The foundation of the SVDA-3P algorithm, illustrated in Fig. 4, remains conceptually similar to the original SVDA, with no significant structural modifications.

Thus, the procedure is independently executed for each FFT frame composed of $(N_{ppc}/2) + 1$ complex coefficients.

For each frame f and each principal component $\mathbf{X}_p$, the corresponding FFT coefficients are processed according to the flowchart shown in Fig. 4. The analysis is performed coefficient-wise for each frequency index q , from the DC component up to the Nyquist frequency.

To track spectral variations, four auxiliary vectors—*MIN*, *MAX*, *SUM*, and *CNT*—are employed. These structures monitor the temporal evolution of each spectral coefficient and count their occurrences across frames. A key element of the method is the use of a frequency-dependent threshold. Instead of a single global value, the algorithm adopts a tolerance vector $\gamma(q)$, defined for each frequency index q , as originally proposed in [3]. This threshold is given by $\gamma(q) = G q^{-\beta}$, where G is a gain factor and β is an exponent that controls the decay rate of the threshold across the frequency spectrum. In the experiments, β was set to 0.001 and kept constant for all components. The parameter G was defined individually for each component $\mathbf{X}_p$ as the maximum magnitude among its coefficients.

A spectral novelty is detected when the absolute variation of a coefficient exceeds this tolerance, i.e., when $|\text{MAX}(q) - \text{MIN}(q)| > \gamma(q)$. This formulation allows the algorithm to ignore small fluctuations and noise that fall within the tolerance band while capturing significant spectral changes associated with real electrical events.

The parameters G and β play a fundamental role in

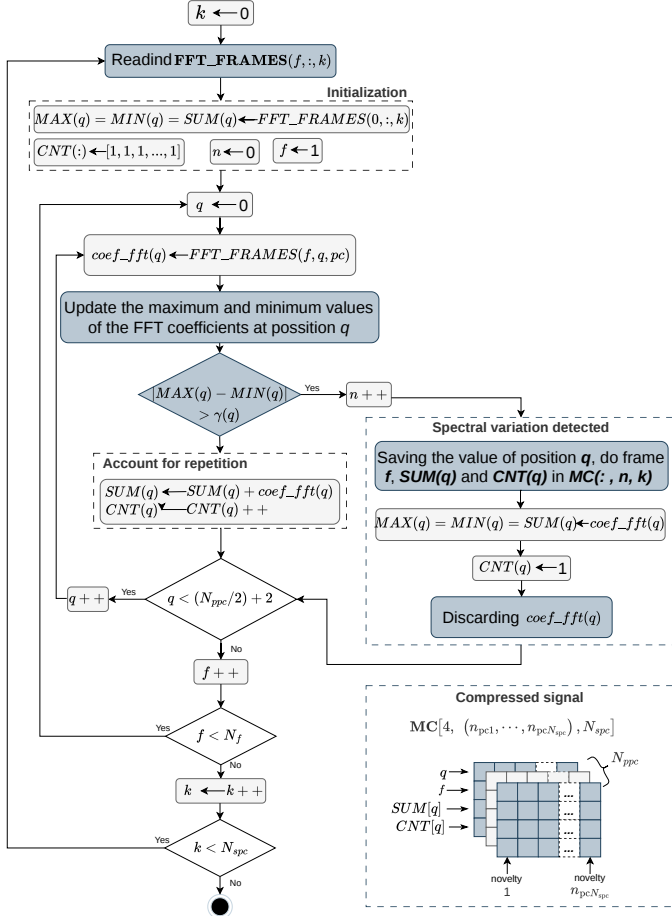


Fig. 4: Compression method based on spectral variation analysis proposed (step D of Fig. 1). Adapted from [4].

controlling the trade-off between compression efficiency and reconstruction fidelity. Specifically, G acts as a global scaling factor: decreasing G reduces the threshold values, making the algorithm more sensitive to small variations and improving reconstruction fidelity at the cost of lower compression ratios. Conversely, increasing G raises the thresholds, causing more coefficients to be treated as redundant, which increases the compression ratio but may lead to the loss of fine spectral details.

The parameter β , in turn, controls how sensitivity is distributed across the spectrum. Proper tuning of β ensures that the algorithm remains less sensitive near the fundamental frequency, where signal energy is dominant, and more sensitive at higher frequencies, where relevant low-energy spectral components might otherwise be neglected.

A detailed sensitivity analysis of the parameters G and β , including their influence on the trade-off between compression ratio and reconstruction error, was previously presented in [4].

The output of stage D corresponds to the compressed representation of the signal, illustrated in the lower-right corner of Fig. 4. This compressed data is represented by N_{spc} matrices in the form $\mathbf{MC}_p \in \mathbb{R}^{4 \times n_p}$ where n_p denotes the number of spectral novelties detected for the p -th PCA component, and $p = 1, \dots, N_{spc}$. Note that n_p is different for each principal component since the number of spectral novelties detected

TABLE II: Compression performance comparison.

#	SVDA				SVDA3P				CR gain
	CR	NMSE (dB)			CR	NMSE (dB)			
		A	B	C		A	B	C	
1	7.4	-26.3	-25.7	-26.2	11.2	-26.1	-25.7	-25.7	1.50
2	32.6	-32.1	-33.3	-32.7	47.0	-32.8	-32.7	-32.7	1.44
3	7.5	-26.8	-25.5	-26.2	11.4	-25.5	-25.7	-26.4	1.52
4	33.4	-32.6	-32.3	-32.4	49.2	-32.0	-32.4	-32.7	1.47
5	9.0	-27.0	-25.3	-26.9	11.4	-27.2	-27.6	-28.0	1.27
6	31.9	-31.5	-31.3	-33.0	46.5	-31.6	-31.8	-31.6	1.46
7	0.6	-8.2	-11.4	-7.5	0.6	-7.7	-11.6	-8.1	0.98
8	81.2	-31.8	-32.8	-31.7	118.5	-32.9	-32.7	-33.0	1.46
9	0.4	-16.0	-11.5	-12.2	0.4	-12.1	-14.1	-13.3	1.00
10	63.1	-32.3	-32.1	-31.8	92.5	-32.4	-32.3	-32.4	1.47
11	54.3	-27.4	-26.7	-27.3	76.2	-27.4	-26.7	-27.3	1.40
12	30.5	-28.0	-26.5	-30.3	64.8	-18.4	-19.4	-18.7	2.13
13	3.4	-14.9	-12.3	-10.0	4.6	-14.1	-11.3	-8.8	1.39

depends on the characteristics of each principal component.

Each column of $\mathbf{MC}_p$ stores four fundamental pieces of information associated with a detected novelty: the frequency index q , the frame index f , the accumulated sum $SUM(q)$, and the repetition counter $CNT(q)$. The vectors $SUM(q)$ and $CNT(q)$ store, respectively, the accumulated sum of the coefficients at frequency position q and the number of consecutive frames in which the coefficient remained stable, i.e., its variation was lower than the threshold $\gamma(q)$. During reconstruction, the decoder uses the ratio $SUM(q)/CNT(q)$ together with the frame index f to recover the average coefficient values and correctly distribute them over the corresponding frames.

III. RESULTS AND DISCUSSION

To evaluate the impact of the dimensionality reduction stage, the signals listed in Table I were compressed using both the proposed SVDA-3P and the original SVDA for comparison.

Since SVDA was originally designed for single-phase signals, its application to three-phase data was performed independently for each phase, without any preprocessing step to exploit the correlation between them.

Table II presents the obtained Compression Ratios (CRs) and the Normalized Mean Squared Errors (NMSEs) values, which are used to assess reconstruction quality by comparing the original and reconstructed signals. The compression ratio is defined as the ratio between the size (in bytes) of the original signal and its compressed representation. As discussed in Section II-A, the selected disturbance signals were extracted and stored in binary format. The reconstruction error was evaluated using the NMSE, computed independently for each phase as described in [4].

Based on the results presented in Table II, it can be observed that, in general, the proposed SVDA-3P improves the compression ratio in cases where the original SVDA already achieves satisfactory performance. This indicates that the dimensionality reduction stage effectively exploits the correlation between phases, reducing redundancy without significantly degrading reconstruction quality.

In contrast, for cases such as signals 7 and 9, where the original SVDA already exhibits limited compression perfor-

mance, the proposed method provides little or no improvement in the compression ratio. In these scenarios, the intrinsic characteristics of the signals impose fundamental limitations on the effectiveness of the spectral variation-based compression approach.

A more detailed analysis of specific cases highlights the impact of the dimensionality reduction stage on compression performance. For instance, the current signal 1 achieved an increase of approximately 50% in compression ratio, improving from 7.4 with the original SVDA to 11.2 with the proposed SVDA-3P. In this case, the PCA stage retained only $N_{spc} = 2$ principal components, while the third PC remained close to zero, similarly to the behavior illustrated for signal 11 in Fig. 2. Consequently, discarding this component had little impact on the reconstructed signal quality.

On the other hand, signal 12 also produced $N_{spc} = 2$, but the discarded third PC contained relevant information for reconstruction. This effect can be observed in the obtained NMSE, which remained around -18 dB. In contrast, signals 7 and 9 produced $N_{spc} = 3$, indicating that no dimensionality reduction was achieved after the PCA stage. As a result, the compression ratios obtained with SVDA and SVDA-3P were essentially the same for these cases.

An example of waveform reconstruction is shown in Fig. 5, which presents the original and reconstructed waveforms, as well as the residual error for Signal 8. In this case, the compression ratio increases significantly from 81.2 to 118.5 when using SVDA-3P. Despite this improvement, the reconstructed signal closely follows the original waveform, with NMSE values remaining around -32 dB across all phases.

In scenarios involving noisy signals or disturbances with strong spectral variability, the compression performance tends to degrade, as discussed in [3], [4]. Under such conditions, the algorithm interprets these variations as relevant information, reducing its ability to identify redundancy in consecutive FFT frames.

Finally, the proposed SVDA-3P significantly enhances performance in most scenarios by effectively leveraging strong inter-phase correlations. Its main limitations in achieving high compression ratios are found in scenarios characterized by significant noise or highly complex spectral content.

IV. CONCLUSIONS

This work proposed an extension of the SVDA algorithm [4] for three-phase signals through the introduction of a PCA-based preprocessing stage. The results demonstrate that exploiting inter-phase correlation effectively increases compression ratios in scenarios where the original method already performs well, while preserving reconstruction quality, as indicated by comparable NMSE values. The method maintains the low-complexity characteristics of SVDA, however, its performance remains limited in signals with high noise levels or significant inter-harmonic content, where spectral variability reduces redundancy and hinders compression efficiency. These limitations are inherent to the spectral variation detection approach and highlight the need for future strategies to better handle non-stationary and spectrally complex signals.

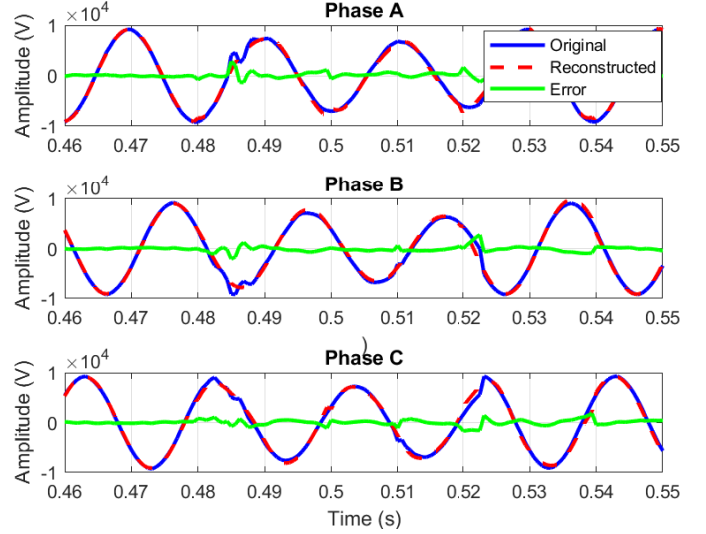


Fig. 5: Comparison of signal 8 with its reconstructed version after compression.

ACKNOWLEDGMENTS

The authors would like to thank Norte Energia S.A. and Agência Nacional de Energia Elétrica (ANEEL) for their support in carrying out this work within the framework of the R&D project ANEEL PD-07427-0223/2023 entitled "Embarcação Fluvial Elétrica de Pequeno Porte com uma Rede de Eletropostos Para Recarga de Baterias Visando ao Atendimento das Populações Ribeirinhas na Amazônia", as well as the support of CAPES and PCT Guamá.

REFERENCES

- [1] G. Cheng, Y. Lin, A. Abur, A. Gómez-Expósito, and W. Wu, "A survey of power system state estimation using multiple data sources: Pmus, scada, ami, and beyond," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 1129–1151, 2024.
- [2] S. He, X. Geng, W. Tian, W. Yao, Y. Dai, and L. You, "Online compression of multichannel power waveform data in distribution grid with novel tensor method," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–11, 2024.
- [3] E. B. Kapisch, V. V. de Moraes, L. R. M. Silva, L. M. A. Filho, and C. A. Duque, "Spectral variation-based signal compression technique for gapless power quality waveform recording in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4488–4498, 2022.
- [4] M. Diniz, P. Anjos, P. Rendeiro, C. Araujo, U. Lebre, and L. Ramalho, "Spectral variation-based compression: A case study with leakage current monitoring," in *2024 Workshop on Communication Networks and Power Systems (WCNPS)*, 2024, pp. 1–7.
- [5] L. Chandran, D. I. Karuppasamy, M. Nair, H. Sun, and P. Krishnakumari, "Compressive sensing in power engineering: A comprehensive survey of theory and applications, and a case study," *Journal of Sensor and Actuator Networks*, vol. 14, p. 28, 03 2025.
- [6] A. Evdakov, G. Filatova, A. Yablokov, A. Kovalenko, E. Skachkov, and I. Makarov, "A dataset of real-world oscillograms from electrical power grids," *Scientific Data*, vol. 13, p. 262, 2026.
- [7] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.
- [8] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2002.
- [9] L. Cheng, X. Ji, F. Zhang, H. Huang, and S. Gao, "Wavelet-based data compression for wide-area measurement data of oscillations," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 6, pp. 1128–1140, 2018.