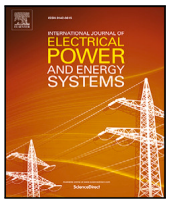
Electrical Power and Energy Systems 171 (2025) 110940

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

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



Parameter estimation of submarine power cables in offshore applications using machine learning-based methods

Felipe P. de Albuquerque [a], Rafael Nascimento [b], Gabriel de Castro Biage [b], Rooney R.A. Coelho [b], Ronaldo F. Ribeiro Pereira [c], Eduardo C. Marques da Costa [b,*], Mario L. Pereira Filho [b], Cassio G. Lopes [b], José R. Cardoso [b]

[a] UFMT - Universidade Federal do Mato Grosso, Cuiabá, Brazil
[b] USP - University of São Paulo, Polytechnic School, São Paulo, Brazil
[c] UFAC - Federal University of Acre, Rio Branco, Brazil

ARTICLE INFO

Keywords:
Submarine cables
Parameter estimation
Machine learning
Phasor measurements
Power transmission

ABSTRACT

Monitoring electrical parameters of power transmission systems is essential to ensure reliability and optimal operating conditions. This research presents an accurate methodology for estimation of the sequence parameters of submarine power cables using a data-driven approach based synchrophasor measurements. Contrarily of conventional techniques, the proposed methodology is based on supervised machine learning models trained on realistic simulations, which incorporate the physical and geometric characteristics of the power cable, with its seven propagation modes. In practical conditions, the training dataset takes into account noise patterns using well-established modeling methods for phasor measurements. These patterns include time-correlated and statistically coherent disturbances, which are representative of those typically encountered in systems employing Phasor Measurement Units (PMUs). A detailed statistical investigation was also conducted to characterize the empirical distribution of the input data, supporting model design and validation. Remarkably, the proposed algorithm achieves accurate parameter estimation even under elevated noise conditions, requiring as few as 200 training samples. The maximum observed estimation error was approximately 1%, underscoring the robustness, efficiency, and practical viability of the proposed framework for the electrical characterization of submarine transmission systems.

1. Introduction

The previous knowledge on the electrical parameters of power systems represents an important aspect for effective monitoring, forecasting, and operational analysis. The accuracy in which such parameters are identified is intrinsically related to the system's performance, including protection parameterization, fault detection and location in overhead and underground networks, insulation coordination, wave propagation analysis, operational state prediction, voltage stability assessment, loadability, and numerous other monitoring and diagnostic applications [1–3] (see Table 1).

The electrical parameters of overhead and submerse (underground or submarine) are conventionally calculated based on the physical and geometric characteristics of the transmission systems, and several approximations such as the earth-return current, as well as skin and proximity effects [4]. For example, the electrical permittivity of insulating materials is subject to variation due to factors such as ambient humidity and the progressive degradation of insulation over time. Conventionally, the electrical parameters, which are determined during the conception and design stages, are assumed to remain constant throughout the entire operational life of the transmission system. However, this assumption is not representative of actual conditions, as overhead lines and cables are continuously exposed to aging processes and environmental issues. In this context, the periodic and continuous estimation of electrical parameters emerges as a valuable strategy for monitoring power systems, enabling the detection of eventual deviations from nominal conditions and supporting preventive maintenance and operational reliability.

Basically, the technical literature presents two estimation procedures: in the time domain and in the frequency domain by means of synchrophasor measurements. The first is usually carried out based on time-domain measurements and oscillographies, usually obtained from digital relays during fault occurrences [3,5,6]. These methods cannot provide a periodic and continuous parameter estimation, since they

* Corresponding author.
E-mail address: educosta@usp.br (E.C.M. da Costa).

https://doi.org/10.1016/j.ijepes.2025.110940
Received 7 June 2025; Received in revised form 11 July 2025; Accepted 21 July 2025
Available online 18 August 2025

Table 1
List of symbols and their meanings.

Symbol	Description
$\boldsymbol{R}'$	Matrix of lumped resistances
$\boldsymbol{L}'$	Matrix of lumped inductances
$\boldsymbol{C}'$	Matrix of lumped capacitances
p	Generic parameter of the power cable $p \in \{R, L, C\}$
$p*$	Generic actual parameter of the power cable
x	Random variable x
$f_p(\mathrm{x})$	Probability density function of a random variable x
$\{\Omega, \mathcal{F}, \mathcal{P}\}$	Probability space
v	Propagation velocity of a electromagnetic wave
ω	Angular frequency of the signal
γ	Complex propagation constant
(k)	Sequence measurement $k \in \{0, 1, 2\}$
$\hat{X}_{s,(k)}$	Phasor X of sequence k from sender end
$\hat{X}_{r,(k)}$	Phasor X of sequence k from receiver end
$R_{(k)}, L_{(k)}, C_{(k)}$	Electrical sequence parameters of the cable
$\mathbb{R}^d$	d-dimensional real space
$\mathbf{x}^{(i)}$	i-sample of the vector of features
$\mathbf{y}^{(i)}$	i-sample of the target vector
$\boldsymbol{D}$	Matrix of data
$\mathcal{L}(f)$	Loss function of the function f
$\mathcal{H}$	Hypothesis class of machine learning algorithms
$\mathcal{N}(\mu, \sigma)$	normal distribution with mean μ and standard-deviation σ
$\mathcal{U}(a, b)$	Uniform distribution between interval $[a, b]$

depend on eventual fault occurrences to register oscillographies. In contrast, frequency-domain methods require a phasorial measurement series, which can be obtained by Phasor Measurement Units PMUs synchronously at both terminals of the transmission system [7]. Several techniques have been proposed to solve the estimation problem with synchrophasor measurements, such as optimal filtering, nonlinear optimization, metaheuristics, and others [8–10].

The research in [7] proposes a methodology for parameter estimation of transmission lines and simultaneously modeling errors by using a modified weighted least squares approach. Although this method shows to be versatile, it requires synchronized measurements from both ends of the transmission system and depends on a detailed understanding of the underlying complex algebraic equations. In contrast, our proposed method builds upon this foundation by adopting a data-centric approach that explicitly incorporates rigorous noise modeling, thereby improving robustness and practicality in real-world scenarios.

In [8], the authors developed a methodology based on a Kalman filter solution combined with the noise propagation model introduced in [11] to estimate transmission line parameters. However, the results reported in the paper exhibit errors of approximately 10%. The technical literature describes that the performance of Kalman filters is highly dependent on precise knowledge of both the system model and the statistical properties of the noise, which are usually unknown in practical terms [12]. In contrast, our proposed approach uses regression methods that do not require an explicit transmission line model or prior assumptions about the noise characteristics.

Several optimization and estimation methods, including those based on metaheuristic strategies, have been applied to parameter estimation problems. Among these, notable examples include the Multi-objective Geometric Mean Optimizer (MOGMO), the Multi-objective Exponential Distribution Optimizer (MOEDO), the Symbiotic Organism Search Algorithm, Multi-objective Moth–Flame Optimizer as well as various other bio-inspired optimization techniques [13–16]. These methodologies have demonstrated promising results and competitive performance in tackling complex estimation problems across various research domains. Practical examples are given in [17,18], in which different machine learning models are used for air-quality level forecasting in urban areas.

In addition to traditional model-based approaches, recent studies have explored the application of advanced machine learning techniques for parameter estimation in electrochemical and power systems. For instance, in [19] proposed hybrid frameworks that combine adaptive filtering, multivariable decoupling, and real-time correction to estimate state-of-charge and internal parameters of lithium-ion batteries under severe temperature and noise conditions. Other relevant contributions involve deep neural networks—such as LSTM and BiLSTM—combined with Bayesian optimization and Kalman filtering strategies, aiming to reduce the impact of noise and input uncertainty while improving remaining useful life (RUL) prediction [20,21].

The solution presented in [3] provides an insightful analysis of the relationship between errors in transmission line parameter estimation and fault location. The proposed methodology is formulated in the time domain and relies on numerical differentiation in the presence of noise [22]. Furthermore, the study overlooks the impact of systematic measurement errors, where the estimation errors exceed 4%, limiting its applicability to high-precision processes.

An interesting estimation method, based on machine learning techniques, is described in [23]. In this research, the parameters of a three-phase distribution feeder are estimated using general regression neural networks (GRNN). However, the approach is tailored to distribution line models that neglect the effect of shunt capacitance, limiting its applicability to transmission systems. Furthermore, the method is considerably more computationally intensive than the one proposed in this paper. It is also important to highlight an average mean absolute percentage error (MAPE) of 1% using 10000 samples, significantly more data than required by the proposed approach to achieve lower error rates.

In this context, this study presents a novel methodology to estimate the electrical parameters of a typical submarine power cable with seven phases by using several machine learning methods. Our proposed methodology encompasses an end-to-end framework, starting from the modeling of the cable using the FEM method, numerical simulation in the time domain, phasor acquisition in real-world scenarios, i.e., considering the presence of random and systematic errors, and finally the estimation using a regression approach.

Concerning the machine learning solution, diverse hypothesis classes were tested during the process, including tree-based methods, generalized additive models, and linear regression. Also, best practices of the pipeline for a data-centric solution were adopted, including cross-validation, the use of a separate test dataset, and feature engineering.

The parameters of the cable were obtained considering the current models for synchronized phasor measurements obtained from PMUs [2,7], i.e., considering Gaussian and systematic errors. Also, the model-free approach based on a regression problem is able to estimate the parameters using only one-side measurement of the cable, which is an improvement over the classical solutions present in the literature [7,9,24].

The main contributions of this paper can be summarized as follows:

- An end-to-end methodology to build a dataset for estimating the parameters of a submarine power cable, from physical and geometric properties to phasor measurements;
- The estimation of electrical parameters for a complex cable configuration with seven propagation modes, which comprise three phase conductors and four armor wires;
- The use of multiple machine learning methods that do not rely on explicit physical models of the cable or predefined noise distributions;
- A method capable of handling both random and systematic measurement errors, as well as one-sided measurements of the cable.

2. Simulation of the synchrophasor dataset

This section outlines the procedure employed to simulate synchronized voltage and current phasors at both terminals of the submarine power cable. The process starts with the calculation of the impedance and admittance parameters, which are essential for accurately modeling

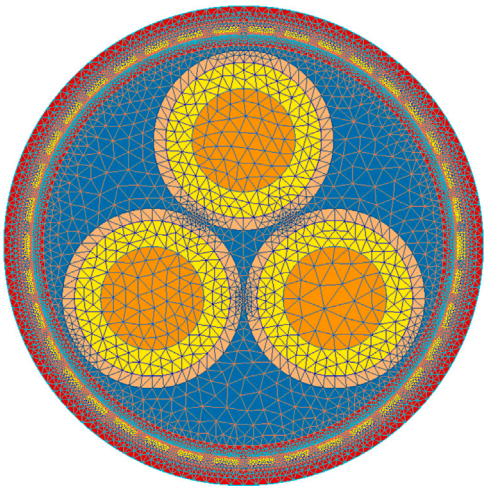


Fig. 1. Finite element mesh of a three-core high-voltage power cable.

the cable's electrical behavior. Subsequently, these parameters are used to simulate a synchrophasor measurement dataset. As previously noted, the generation of this synthetic database is necessitated by the unavailability of real phasor measurement data.

The theory of multi-conductor transmission lines is formulated in terms of electrical circuit quantities, allowing for an abstraction from the physical aspects of the electromagnetic fields and enabling a more tractable analytical approach. Nevertheless, the governing equations are fundamentally derived from Maxwell's equations, thereby retaining consistency with the underlying principles of electromagnetism [25]. The equations that model the multi-conductor transmission line are characterized by a set of electric parameters that represents the cable and can be accurately calculated by using the Finite Element Method FEM, since the characteristics of its physical and structural cross section are previously known. The technical literature shows several procedures for calculating the electric parameters of cables with multiple propagation modes and simulation [26–28]. Conventionally, a FEM-base algorithm is applied to compute the electrical parameters of the cable, and thereafter a well-established line or cable modeling is used for transient or steady-state simulations.

The FEM has been widely used as an accurate and reliable approach for the calculation of series and transversal parameters of cables with arbitrary and non-conventional cross-sectional geometries [27,28]. This method offers substantial advantages in situations involving complex or irregular geometries, as it enables precise modeling of the system's physical behavior. In contrast, some well-established analytical methods are available in the technical literature for similar calculations [29,30]. However, these methods often rely on a series of simplifying assumptions and approximations, which can lead to inaccuracies, particularly when applied to cables with arbitrary cross sections.

The schematic representation of the cross-section of the power cable, considering the finite element mesh, is shown in Fig. 1. In this figure, each conductor (orange) is enclosed by an insulation layer (yellow) and additional shielding layers (light brown) to ensure proper electrical performance and mechanical integrity. The outer sheath (blue) encapsulates the entire cable, providing environmental protection and electrical insulation. The finely discretized triangular mesh enables precise numerical analysis of the electric field distribution, thermal behavior, and mechanical stress concentration. The results of the FEM-based method is presented in Appendix A.1.

Since the reference electrical parameters are known, the phasorial measurements can be simulated through the two-port network representation of the cable. This model provides an accurate characterization of frequency-domain propagation and the characteristic impedance, thereby enabling the precise simulation of voltage and current phasors at both terminals of the cable [31,32].

The multi-phase representation can be analyzed using modal decomposition, wherein the phase conductors, shields, and armor are decoupled into seven independent propagation modes. These modes are subsequently modeled through a two-port network representation, enabling a precise characterization of the system's electromagnetic behavior [33].

Subsequently, random and systematic errors are applied to the dataset, in order to reproduce the real behavior of the PMU measurements [2,7]. The currents and voltages values, obtained from the introduced cable model, are the synchrophasors that compose the dataset to be applied in the learning process of the AI-based methodology for parameter estimation.

3. The parameter estimation problem

The previous section provided a detailed description of the procedure employed for generating the phasor measurements, as well as the calculation of the reference parameters for this purpose. This section introduces the building process of the synchrophasor dataset as well as a general supervised learning approach for the parameter estimation by various machine learning hypothesis classes.

3.1. Building the database

In a supervised learning framework, the process of constructing the database represents a crucial to achieve an effective solution. The technical literature addressing parameter estimation in transmission systems, particularly through the application of machine learning techniques, remains relatively limited and underexplored [34–36]. Therefore, this study proposes an innovative approach for generating training datasets, which capitalizes on characteristic patterns and statistical properties extracted from historical data reported in the technical literature.

According to the literature, parameter errors typically range from approximately 10% to 30% of the true parameter value [37,38]. Consequently, the database should include parameter values within this interval. Thus, if p^* denotes the exact value of a given parameter, the inaccurate parameter p should be constrained within the following bounds:

$$p^*(1-0.30) \leq p \leq p^*(1+0.30) \Rightarrow 0.7p^* \leq p \leq 1.3p^*$$

Since extreme parameter values occur less frequently than those with smaller deviations, the Gaussian distribution is an appropriate choice for data generation. This distribution naturally aligns with the expected error pattern, where values farther from the mean occur with lower frequency, reflecting the typical distribution of parameter deviations. The probability density function (PDF) of a Gaussian distribution is given by:

$$f_p(p) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(p-\mu)^2}{2\sigma^2}\right), \tag{1}$$

where μ represents the mean parameter value, and σ denotes the standard deviation, which determines the spread of the distribution.

To establish a relationship between the standard deviation and the range of parameter errors, the empirical 3σ rule can be applied [39]. This rule states that approximately 99.7% of the values in a normal distribution lie within three standard deviations from the mean:

$$\mathcal{P}(\mu - 3\sigma \leq \mathrm{X} \leq \mu + 3\sigma) \approx 0.997,$$

where $\mathcal{P}(\cdot)$ represents the probability function, and X is a Gaussian random variable with mean μ and standard deviation σ.

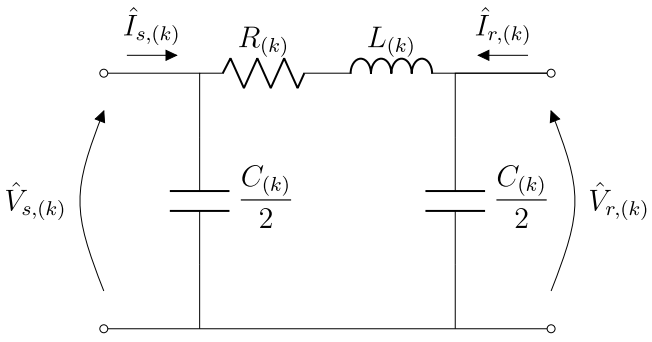


Fig. 2. Analysis of the residuals for sequence parameters.

Given that the maximum observed error is defined as $E_{\max}$, the standard deviation can be expressed as:

$$\sigma \approx \frac{E_{\max}}{3}. \tag{2}$$

Applying this formulation to the problem, where parameter values range from 70% to 130% of the true value p^*, we obtain:

$$p^* + 3\sigma \leq 1.3p^* \Rightarrow \sigma = 0.1p^*.$$

Therefore, to generate data for parameters with a mean value of p^* and observed maximum and minimum values of $1.3p^*$ and $0.7p^*$, respectively (with a probability of 0.997), a Gaussian distribution with mean $\mu = p^*$ and standard deviation $\sigma = 0.1p^*$ should be used.

However, this methodology may yield physically invalid parameter values, particularly with respect to the inherent limitation imposed by the speed of light, which is intrinsically related to the electrical parameters of the transmission line. Therefore, it becomes essential to evaluate the resulting expressions to ensure their physical plausibility. Specifically, with regard to the speed of light in vacuum c:

$$v_{(1)} = \frac{\omega}{\mathrm{Im}(\gamma_{(1)})} < c,$$
$$v_{(0)} = \frac{\omega}{\mathrm{Im}(\gamma_{(0)})} < c,$$

where the index (1) represents the positive sequence, while (0) the zero sequence.

Even after these verifications, the proposed methodology may still result in inconsistent values across individual rows, i.e. bad data values that might affect the performance of the machine learning method. In this sense, the proposed method introduces a new approach to verify whether each row of the dataset represents a consistent record of the measurements. This new methodology is based on the equating of sequence models for the submarine power cable.

Consider the equivalent circuit per sequence ($k = 0$ or 1) shown in Fig. 2, where $\hat{V}_{s,(k)}$ and $\hat{I}_{s,(k)}$ represent the voltage and current of sequence k at the sending end, $\hat{V}_{r,(k)}$ and $\hat{I}_{r,(k)}$ are the voltage and current of sequence k at the receiving end of the system. Terms $L_{(k)}$, $R_{(k)}$, and $C_{(k)}$ are the lumped parameters of sequence k for the equivalent circuit. If the measurements generated by the FEM method associated with the propagation model are coherent, then f_1, f_2, and f_3 defined by (3), (4), and (5) should be null or then very close to zero.

$$f_1 \overset{def}{=} \hat{I}_{s,(k)} - \hat{I}_{r,(k)} - (\hat{V}_{r,(k)} + \hat{V}_{s,(k)}) \cdot j\omega C_{(k)}/2 \tag{3}$$

$$f_2 \overset{def}{=} \hat{I}_{r,(k)} + \hat{V}_{r,(k)} \cdot j\omega C_{(k)}/2 - \frac{\hat{V}_{r,(k)} - \hat{V}_{r,(k)}}{R_{(k)} + j\omega L_{(k)}} \tag{4}$$

$$f_3 \overset{def}{=} \hat{I}_{s,(k)} - \hat{V}_{s,(k)} \cdot j\omega C_{(k)}/2 - \frac{\hat{V}_{r,(k)} - \hat{V}_{r,(k)}}{R_{(k)} + j\omega L_{(k)}} \tag{5}$$

3.2. Machine learning approach

To estimate the sequence parameters of a submarine power cable, we propose a supervised regression framework using terminal phasor measurements of voltage and current. Specifically, the input features are constructed from three-phase voltage and current signals measured at both the sending (s) and receiving (r) ends of the cable, covering phases A, B, and C.

The objective is to estimate the six real-valued sequence parameters:

$$\{R_{(0)},\ L_{(0)},\ C_{(0)},\ R_{(1)},\ L_{(1)},\ C_{(1)}\},$$

based on the observed terminal phasors:

$$\{\hat{V}_s,\ \hat{I}_s,\ \hat{V}_r,\ \hat{I}_r\}_{\mathrm{A,B,C}}.$$

Here, the phasor quantities are expressed as complex numbers with magnitude and phase angle:

$$\hat{V}_s = |\hat{V}_s|e^{j\theta_s}, \quad \hat{I}_s = |\hat{I}_s|e^{j\varphi_s}, \quad \hat{V}_r = |\hat{V}_r|e^{j\theta_r}, \quad \hat{I}_r = |\hat{I}_r|e^{j\varphi_r},$$

where $j = \sqrt{-1}$ and all magnitudes and angles are included as features.

The physical relation between voltages and currents along the cable follows the frequency-domain telegrapher's equations:

$$\frac{d\hat{\boldsymbol{V}}(x)}{dx} = -\boldsymbol{Z}(\omega)\,\hat{\boldsymbol{I}}(x), \quad \frac{d\hat{\boldsymbol{I}}(x)}{dx} = -\boldsymbol{Y}(\omega)\,\hat{\boldsymbol{V}}(x),$$

where $\boldsymbol{Z}(\omega)$ and $\boldsymbol{Y}(\omega)$ are the per-unit-length series impedance and shunt admittance matrices obtained via FEM simulations. From these matrices, it is performed the Kron reduction method and after is applied the symmetrical components, resulting in the sequence parameters:

$$\left\{R_{(0)},\ L_{(0)},\ C_{(0)},\ R_{(1)},\ L_{(1)},\ C_{(1)}\right\}.$$

The phasors obtained from the signal propagation model were intentionally corrupted by introducing both random noise and systematic errors into the original measurements. The resulting corrupted phasors were used to construct the dataset with N samples:

$$\mathcal{D} = \left\{(\boldsymbol{x}^{(i)},\ \boldsymbol{y}^{(i)})\right\}_{i=1}^{N},$$

where $\boldsymbol{x}^{(i)}$ denotes the input features derived from the corrupted phasors, and $\boldsymbol{y}^{(i)}$ corresponds to the true electrical parameters.

Let $\boldsymbol{x} \in \mathbb{R}^k$ denote the real-valued feature vector derived from the corrupted phasors, and let $\boldsymbol{y} \in \mathbb{R}^6$ represent the sequence parameters. The estimation task is to learn a function f that maps input features to the target parameters:

$$f : \mathbb{R}^k \to \mathbb{R}^6, \quad \boldsymbol{y} = f(\boldsymbol{x}).$$

The model is trained to minimize the Mean Squared Error (MSE) between predicted and true parameters:

$$\mathcal{L}(f) = \frac{1}{N}\sum_{i=1}^{N} \|f(\boldsymbol{x}^{(i)}) - \boldsymbol{y}^{(i)}\|^2,$$

or equivalently:

$$\mathcal{L}(f) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{6}\left(f_j(\boldsymbol{x}^{(i)}) - y_j^{(i)}\right)^2.$$

Several model types are explored to approximate f, including linear regression, tree-based models, and Generalized Additive Models (GAMs).

These models define the hypothesis space $\mathcal{H}$ over which the optimal function is sought:

$$f^* = \arg\min_{f \in \mathcal{H}} \mathcal{L}(f).$$

This approach enables direct estimation of sequence parameters from terminal measurements, while being resilient to noise and modeling uncertainties. The complete methodology of the proposed solution is presented in Fig. 3.

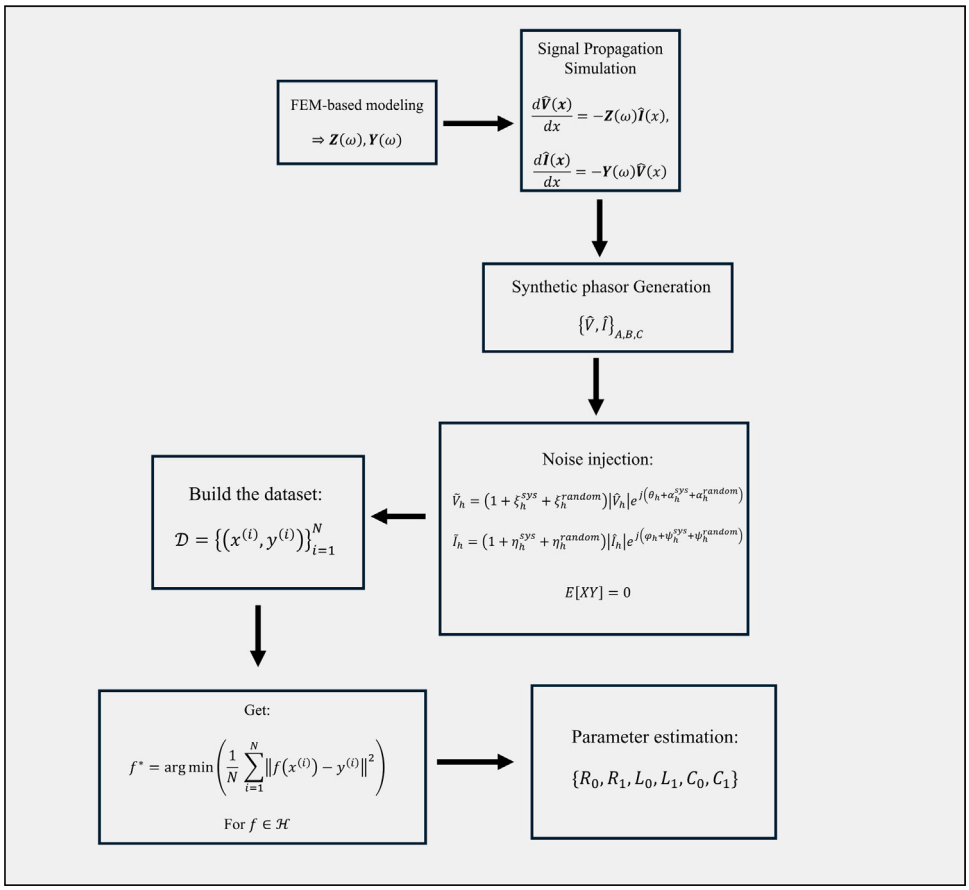


Fig. 3. Flowchart of the parameter estimation methodology.

4. Results

The submarine power cable was simulated under the conditions specified in Table 11. Initially, a total of 4285 valid samples were obtained for generating the Gaussian distribution of the parameters, following the verification procedures detailed in Section 3.1.

In terms of implementation, a dataset comprising 4285 samples can be obtained in under two minutes of system observation, assuming a typical PMU reporting rate of 60 Hz (i.e., one sample per cycle) [40]. This highlights the feasibility of rapidly collecting sufficient data for parameter estimation using phasor measurements. The machine learning models were trained on a standard workstation equipped with an Intel Core i7 processor (3.40 GHz) and 16 GB of RAM, without requiring GPU acceleration. Given the lightweight nature of the linear regression model, both training and inference procedures exhibit low computational overhead, making the proposed approach suitable for real-time or near-real-time deployment in practical monitoring environments.

To demonstrate the effectiveness of the proposed methodology in handling the residuals introduced in Section 3.1, the results for the residuals associated with the positive and zero sequence components were analyzed. The residuals corresponding to the positive sequence are depicted in Fig. 4, whereas those related to the zero sequence are presented in Fig. 5.

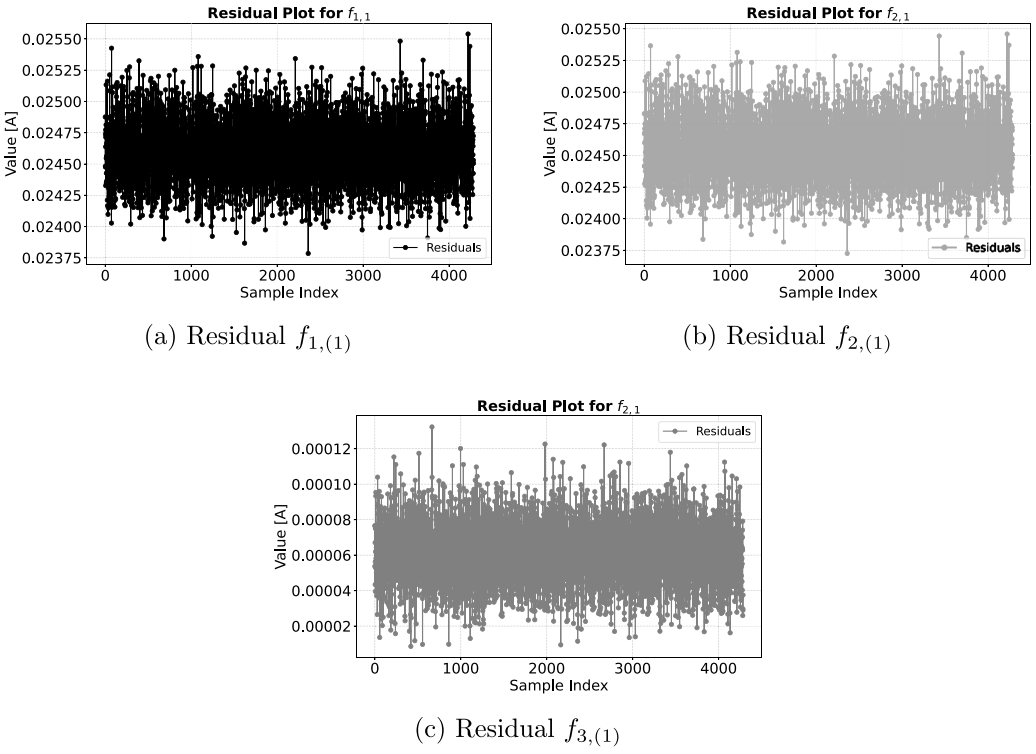
As expected from the theoretical analysis, the residuals (f_1, f_2, f_3) are smaller for each sequence, demonstrating the coherence of the measurements produced by the developed methodology, considering the power cable models and the dataset construction. Additionally, it is important to highlight that the presented approach can be used as a bad data detection method, allowing the selection of only suitable measurements for use in the estimation method. By setting a specific threshold, rows with residuals exceeding the established limit can be excluded.

The proposed method utilizes an 80%–20% data split, with 80% of the dataset designated for training and 20% for testing. Model training was conducted using k-fold cross-validation, where the dataset was partitioned into k subsets. In each iteration, the model was trained on $k - 1$ folds and validated on the remaining fold. This iterative process improved the model's generalization, enhancing its ability to perform accurately on unseen data.

Moreover, this methodology reduces the risk of overfitting, i.e., the tendency of the machine learning algorithm to memorize patterns from the training dataset rather than learning generalizable features. By repeatedly training on different subsets of the data, the model is exposed to a broader variety of patterns, improving its robustness and predictive performance.

The problem is a multi-output regression task, since it requires the estimation of six parameters. The approach employed for this task involves treating each parameter separately. Specifically, regression models are fitted for each parameter individually, while the remaining parameters are disregarded during the estimation process for the parameter in question. This method is referred to as the single-target method [41].

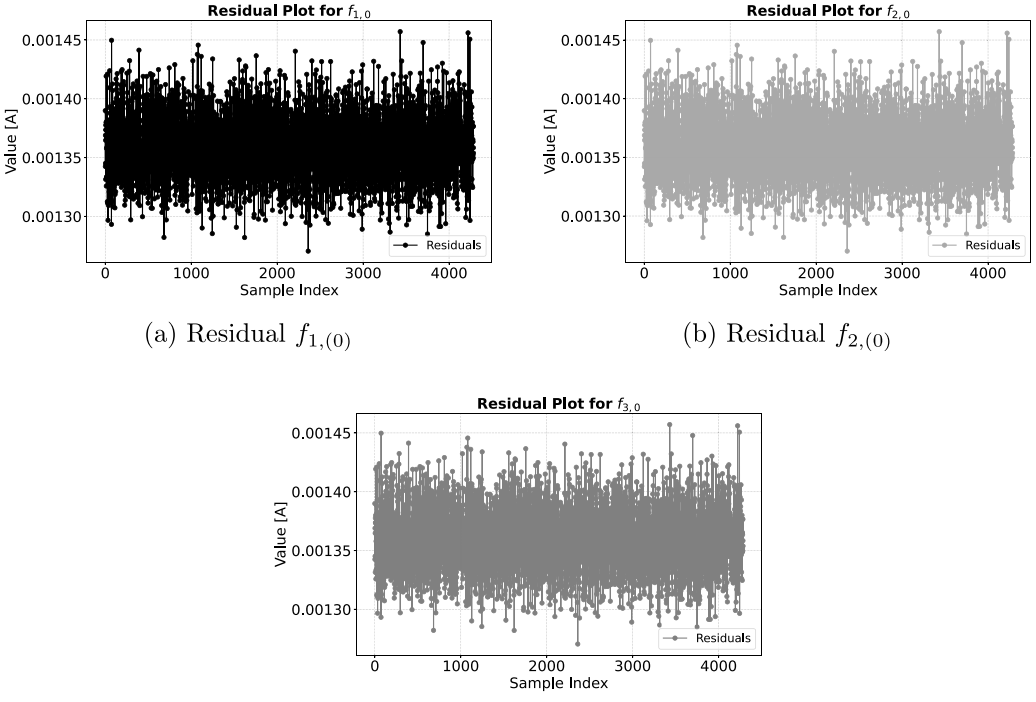
We have performed a comparison of different hypothesis classes $(\mathcal{H})$, this comparison should be done in the validation base, which is part of the training process [42]. To evaluate and compare the performance of the different methods, three metrics were applied: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), defined in (6). The results were presented considering each different parameter $(R_1, R_0, L_1, L_0, C_1, C_0)$ and can be seen in Tables 2, 3, 4,5, 6, and 7.



(a) Residual $f_{1,(1)}$ (b) Residual $f_{2,(1)}$

(c) Residual $f_{3,(1)}$

Fig. 4. Residuals plots for $f_{1,(1)}$, $f_{2,(1)}$, and $f_{3,(1)}$.



(a) Residual $f_{1,(0)}$ (b) Residual $f_{2,(0)}$

(c) Residual $f_{3,(0)}$

Fig. 5. Residuals plots for $f_{1,(0)}$, $f_{2,(0)}$, and $f_{3,(0)}$.

Some considerations must be addressed regarding the obtained results. The performance metrics were evaluated based on the complete set of 19 regression models available in the PyCaret library. However, for clarity and conciseness, only the nine best-performing models were selected for presentation in this manuscript. The algorithms were assessed using K-fold cross-validation with five folds, enabling a consistent comparison of the models based on the validation dataset. Concerning the set of considered measurements, it can be fully described as:

$$\mathbb{F} = \{|\hat{V}_{As}|, |\hat{V}_{Bs}|, |\hat{V}_{Cs}|, \angle\hat{V}_{As}, \angle\hat{V}_{Bs}, \angle\hat{V}_{Cs}, |\hat{I}_{As}|, |\hat{I}_{Bs}|, |\hat{I}_{Cs}|, \angle\hat{I}_{As}, \angle\hat{I}_{Bs}, \angle\hat{I}_{Cs}$$
$$|\hat{V}_{Ar}|, |\hat{V}_{Br}|, |\hat{V}_{Cr}|, \angle\hat{V}_{Ar}, \angle\hat{V}_{Br}, \angle\hat{V}_{Cr}, |\hat{I}_{Ar}|, |\hat{I}_{Br}|, |\hat{I}_{Cr}|, \angle\hat{I}_{Ar}, \angle\hat{I}_{Br}, \angle\hat{I}_{Cr}\},$$

Table 2
Performance comparison of regression models for predicting **R1**.

Algorithm	MAPE (%)	RMSE (Ω)	MAE (Ω)
Linear Regression	**0.0383**	**0.000641**	**0.000401**
Extra Trees Regressor	3.95	0.0330	0.0178
Orthogonal Matching Pursuit	1.18	0.0191	0.0139
Gradient Boosting Regressor	4.15	0.0390	0.0267
Huber Regressor	1.09	0.0209	0.0135
Random Forest Regressor	5.21	0.0506	0.0310
Light Gradient Boosting Machine	6.26	0.0671	0.0390
Extreme Gradient Boosting	4.46	0.0479	0.0330
Decision Tree Regressor	6.12	0.0765	0.0489

Table 3
Performance comparison of regression models for predicting **R0**.

Algorithm	MAPE (%)	RMSE (Ω)	MAE (Ω)
Linear Regression	**0.1746**	**0.00394**	**0.00294**
Extra Trees Regressor	4.08	0.0817	0.0623
Orthogonal Matching Pursuit	1.74	0.0407	0.0281
Gradient Boosting Regressor	3.67	0.0752	0.0584
Huber Regressor	2.82	0.0590	0.0459
Random Forest Regressor	4.38	0.0891	0.0667
Light Gradient Boosting Machine	4.72	0.0962	0.0699
Extreme Gradient Boosting	4.37	0.0895	0.0684
Decision Tree Regressor	5.88	0.1204	0.0908

Table 4
Performance comparison of regression models for predicting **L1**.

Algorithm	MAPE (%)	RMSE (mH)	MAE (mH)
Linear Regression	**0.0142**	**0.00165**	**0.00114**
Extra Trees Regressor	0.562	0.1045	0.0404
Orthogonal Matching Pursuit	0.408	0.0436	0.0312
Gradient Boosting Regressor	0.847	0.1229	0.0651
Huber Regressor	0.462	0.0506	0.0326
Random Forest Regressor	1.33	0.1733	0.0939
Light Gradient Boosting Machine	2.30	0.3023	0.1419
Extreme Gradient Boosting	1.32	0.1758	0.0976
Decision Tree Regressor	1.41	0.2338	0.1294

Table 5
Performance comparison of regression models for predicting **L0**.

Algorithm	MAPE (%)	RMSE (mH)	MAE (mH)
Linear Regression	**0.1588**	**0.0154**	**0.0118**
Extra Trees Regressor	2.18	0.221	0.164
Orthogonal Matching Pursuit	1.04	0.0932	0.0642
Gradient Boosting Regressor	2.49	0.222	0.170
Huber Regressor	1.92	0.185	0.146
Random Forest Regressor	3.06	0.264	0.193
Light Gradient Boosting Machine	4.36	0.348	0.221
Extreme Gradient Boosting	2.87	0.258	0.188
Decision Tree Regressor	3.18	0.328	0.262

where the operator $\angle$ denotes the phase angle of the complex number associated with each phasor measurement. This way, initially, it was considered 24 features for the multioutput regression problem. This set can be divided into the features from the sender-end (s) and from the receiver-end (r). It will be proved that our proposed methodology requires only data from one side to achieve acceptable performance. The magnitudes are expressed in per unit (p.u.), and therefore do not require any additional scaling. Regarding the target variables, since they exhibit different scales, a standardization procedure was adopted to improve the numerical stability of the solution.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y^{(i)} - y^{(i)}_{pred}|}{|y^{(i)}|} * 100, \qquad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y^{(i)} - y^{(i)}_{pred})^2},$$
$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y^{(i)} - y^{(i)}_{pred}| \tag{6}$$

where $y^{(i)}_{pred}$ is the predicted target of the sample i.

An analysis of the results presented in Tables 2, 3, 4, 5, 6, and 7 indicates that the best-performing algorithm for the studied case was

Table 6
Performance comparison of regression models for predicting **C1**.

Algorithm	MAPE (%)	RMSE (μF)	MAE (μF)
Linear Regression	**0.00164**	**0.0000791**	**0.0000532**
Extra Trees Regressor	0.787	0.0447	0.0248
Orthogonal Matching Pursuit	0.0411	0.00186	0.00132
Gradient Boosting Regressor	1.09	0.0499	0.0347
Huber Regressor	2.69	0.111	0.0873
Random Forest Regressor	1.21	0.0631	0.0382
Light Gradient Boosting Machine	1.28	0.0651	0.0418
Extreme Gradient Boosting	1.32	0.0631	0.0426
Decision Tree Regressor	1.88	0.0865	0.0606

Table 7
Performance comparison of regression models for predicting **C0**.

Algorithm	MAPE (%)	RMSE (μF)	MAE (μF)
Linear Regression	**1.04**	**0.0387**	**0.0263**
Extra Trees Regressor	5.41	0.172	0.138
Orthogonal Matching Pursuit	5.42	0.175	0.138
Gradient Boosting Regressor	5.24	0.173	0.134
Huber Regressor	6.01	0.205	0.157
Random Forest Regressor	5.27	0.171	0.134
Light Gradient Boosting Machine	5.39	0.177	0.138
Extreme Gradient Boosting	5.51	0.180	0.141
Decision Tree Regressor	7.22	0.235	0.179

Table 8
Performance of **Linear Regression** on the Test Dataset.

Parameter	MAPE (%)	RMSE	MAE
R_0	0.144434	0.003230 [Ω]	0.002437 [Ω]
R_1	0.032516	0.000554 [Ω]	0.000344 [Ω]
L_0	0.117050	0.0121 [mH]	0.00921 [mH]
L_1	0.011067	0.00151 [mH]	0.000995 [mH]
C_0	1.232133	0.0451 [μF]	0.0370 [μH]
C_1	0.001480	0.000079 [μF]	0.000049 [μF]

linear regression, followed by the Extremely Randomized Trees (Extra Trees), an ensemble learning method based on the bagging technique. Furthermore, it was observed that the parameter C_0 exhibited the poorest estimation performance. The best performance was observed for the linear regression, on the test dataset is detailed in Table 8.

To enhance the interpretability of the proposed approach with respect to the machine learning methodology, explainable AI techniques can be employed, for instance, SHAP (SHapley Additive exPlanations) values. SHAP is a game-theoretic method that quantifies the contribution of each feature by computing its average marginal effect on the model's output across all possible combinations of input features. This allows for a consistent and theoretically grounded interpretation of how each input influences the model's predictions. The SHAP value ϕ_i is defined as the average marginal contribution of feature i across all possible subsets of the remaining features by using (7). The results of such analysis are presented in Fig. 6.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (M - |S| - 1)!}{M!} \left[f_{S\cup\{i\}}(\boldsymbol{x}_{S\cup\{i\}}) - f_S(\boldsymbol{x}_S) \right] \tag{7}$$

where:

- M is the total number of features.
- S is any subset of $N \setminus \{i\}$.
- $f_S(\boldsymbol{x}_S)$ denotes the expected value of the model output when only the features in S are known, with the remaining features marginalized.
- The weight $\frac{|S|!(M-|S|-1)!}{M!}$ corresponds to the fraction of all possible orderings in which the subset S precedes feature i.

The results presented in Fig. 6 indicate that the most relevant features for parameter estimation are the magnitude and phase of the current phasors in phases A and B. This highlights the crucial role of PMUs in the estimation process, as such high-resolution phasor

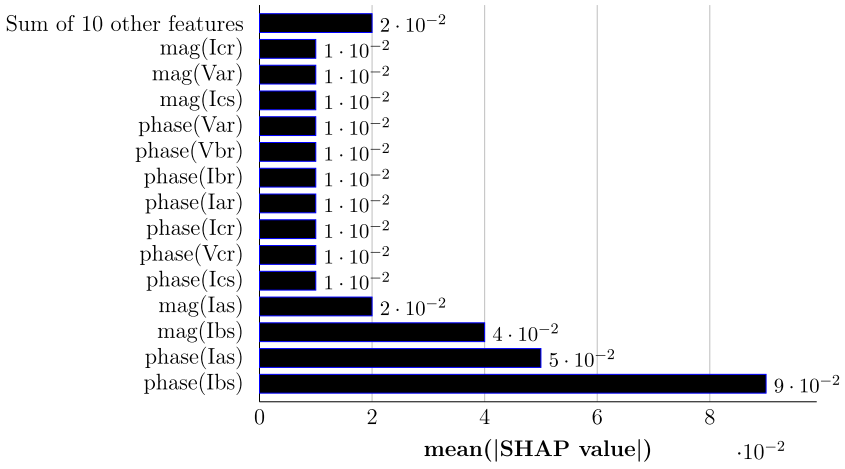


Fig. 6. Mean absolute SHAP value for each feature.

Table 9
Performance of **Linear Regression** on the Test Dataset using measurements from only one terminal.

Parameter	MAPE (%)	RMSE	MAE
R_0	0.61	$1.3513 \cdot 10^{-2}$ [Ω]	$1.0590 \cdot 10^{-2}$ [Ω]
R_1	0.26	$5.0426 \cdot 10^{-3}$ [Ω]	$3.8012 \cdot 10^{-3}$ [Ω]
L_0	0.09	$1.0106 \cdot 10^{-2}$ [mH]	$7.1696 \cdot 10^{-3}$ [mH]
L_1	0.06	$7.0576 \cdot 10^{-3}$ [mH]	$4.9921 \cdot 10^{-3}$ [mH]
C_0	4.88	$1.5890 \cdot 10^{-1}$ [μF]	$1.2519 \cdot 10^{-1}$ [μF]
C_1	0.01	$5.8761 \cdot 10^{-4}$ [μF]	$4.4256 \cdot 10^{-4}$ [μF]

measurements are exclusively available through phasor measurement units.

The presented solution has employed data from both ends of the cable. However, in certain situations, measurements may be available from only one side. Therefore, the parameter estimation methodology must be capable of handling this constraint, that is, estimating the parameters using one-sided measurements. In classical approaches [7,9,43,44], which rely on the explicit modeling and equation, based representation of the cable, such estimation is not feasible when only one side is observed. Nevertheless, by adopting a data-centric approach, it becomes possible to estimate the electrical parameters even with one-sided measurements. Therefore, we have considered available the measurements only from the sender-end of the power cable. The results of such analysis are presented in Table 9.

The results presented in Table 9 show that the performance of the method decreases when only the measurements from one side of the power cable are considered. This outcome was expected, since the performance of a machine learning-based method depends on the variability of the features present in the dataset. Furthermore, the exclusive use of sender-end data naturally reduces the method's accuracy, as observed. However, even under this condition, the method still achieved suitable performance for steady-state applications (errors below 5%) [45], demonstrating its ability to estimate the parameters using measurements from a single terminal.

Although the proposed methodology has shown promising results in estimating electrical parameters of submarine cables under controlled conditions, some practical scenarios may present additional challenges. For instance, in configurations involving multi-segment cables with different physical or electromagnetic characteristics along their length, the estimation model may require re-training or feature re-engineering to account for the varying propagation dynamics. Furthermore, the presence of underwater disturbances such as sea currents, temperature gradients, or biofouling can introduce additional noise and variability into the phasor measurements, potentially reducing the estimation accuracy. For example, this kind of disturbance might lead the random errors to follow non-stationary and non-Gaussian distributions [12]. Even though these conditions require adjustments in the estimation process, the data-centric approach can still be adapted through modifications in the dataset, allowing the parameters to be estimated under these new configurations.

5. Estimation considering errors in the measurements

Error modeling represents a crucial aspect during parameter estimation process using PMU measurements. Several studies have examined the measurement models generated by the data acquisition chain in which PMUs are integrated [2,46–48]. Essentially, errors in PMU measurements can be divided into two categories: random errors and systematic errors. Random errors are typically modeled as a Gaussian noise term, whereas systematic errors are modeled by a uniform distribution, meaning that their contribution is equally distributed across all frequencies [7].

In the technical literature, measurement uncertainties are typically categorized into two distinct types: random errors, which are inherently associated with PMUs, and systematic errors, which are primarily attributed to the limited precision of Instrument Transformers (ITs), such as current and voltage transformers [2,7]. Using the values presented in [2]:

- ITs are assumed to be of class 0.5, thus using 0.5% for maximum voltage and current ratio errors, 0.9 crad for maximum CT phase-angle displacement, and 0.6 crad for maximum VT phase-angle displacement;
- For the errors associated with PMUs, a maximum amplitude error of 0.1% (α_{PMU}) and a maximum phase-angle error of 0.1 crad (β_{PMU}) are usually considered.

The systematic error is modeled using a uniform distribution. Thus, since the values are given by their maximum deviations, the distribution limits can simply be adjusted to account for the maximum possible errors. For random errors, the maximum values must be set to the standard deviation of the Gaussian distribution. Following a methodology analogous to that described in Section 3.1, the standard deviation of the distribution can be estimated from the maximum allowable deviations in the PMU magnitude and phase measurements, as given by the following expressions:

$$\sigma_{mag} = \frac{\alpha_{PMU}}{3}, \quad \sigma_{phase} = \frac{\beta_{PMU}}{3},$$

where σ_{mag} is the standard-deviation of the distribution that models the magnitude and σ_{phase} is related to the Gaussian distribution for the phase.

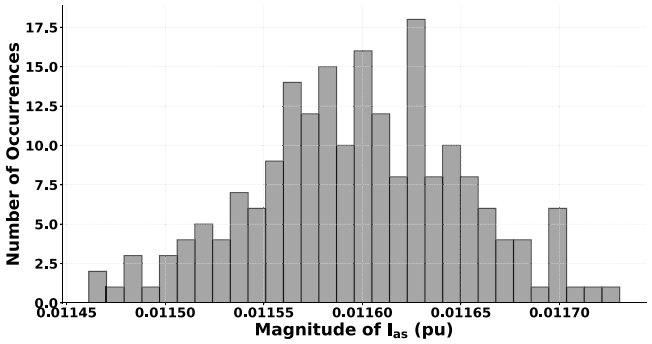


Fig. 7. Histogram for $|\hat{I}_{As}|$.

The field measurements can be simulated by using the following model:

$$\begin{cases} \tilde{\mathrm{V}}_h &= (1+\xi_h^{sys}+\xi_h^{random})|\hat{V}_h|\ e^{j(\theta_h+\alpha_h^{sys}+\alpha_h^{random})} \\ \tilde{\mathrm{I}}_h &= (1+\eta_h^{sys}+\eta_h^{random})|\hat{I}_h|\ e^{j(\varphi_h+\psi_h^{sys}+\psi_h^{random})} \end{cases} \tag{8}$$

where h represents the phase $\{A, B, C\}$, $\tilde{V}_h$ and $\tilde{I}_h$ represent the phasors corrupted by noise, the term "sys" denotes systematic error, while "random" pertains to the random errors inherent in the measurements. The terms ξ, α, η, and Ψ are uncorrelated and separately modeled for each magnitude and phase in complex data. The key distinction between systematic and random errors lies in their behavior: systematic errors contribute consistently across all samples, whereas random errors fluctuate throughout the time series. As the errors are uncorrelated, it is possible to write the following:

$$\mathbb{E}[\mathrm{XY}] = 0,$$

for $\mathrm{X}, \mathrm{Y} \in \{\xi_h^{sys}, \xi_h^{random}, \eta_h^{sys}, \eta_h^{random}, \alpha_h^{sys}, \alpha_h^{random}, \psi_h^{sys}, \psi_h^{random}\}$. As an example, the random error and systematic error for phase A in voltage is calculated as:

$$\xi_A^{random} \sim \mathcal{N}(0, \alpha_{PMU}|\hat{V}_A|/3), \quad \xi_A^{sys} \sim \mathcal{U}(-0.005|\hat{V}_A|, 0.005|\hat{V}_A|).$$

To generate the synthetic dataset, 1000 Monte Carlo (MC) trials were performed for each original measurement. In each trial, a corrupted measurement window consisting of 200 samples was created. The original measurement remained constant within this window; however, after the addition of noise, the resulting signal varied according to the error model described earlier. Specifically, each trial included a fixed systematic error applied uniformly across the 200 samples, and a random Gaussian component independently applied to each sample. Fig. 7 illustrates the result of a single MC trial, showing all 200 corrupted samples generated within that window. Notably, the resulting distribution deviates significantly from Gaussianity due to the combined influence of the Gaussian and Uniform error components, as further analyzed in Appendix A.3. This non-Gaussian behavior is consistent with observations reported in recent literature [12,24].

Let $\hat{\theta}_i$ denote the estimated parameter vector obtained in the ith Monte Carlo trial, and let N_T be the total number of trials. The final estimated parameter vector $\hat{\theta}$ is then computed as the average over all trials:

$$\hat{\theta} = \frac{1}{N_T}\sum_{i=1}^{N_T}\hat{\theta}_i$$

For simplicity, we will omit the notation indicating that the reported results correspond to the average over the Monte Carlo realizations.

Considering this noise modeling, the results employing the metric MAPE considering 5-fold cross-validation for all the parameters are presented in Table 10, where the abbreviations are Linear Regression (LR), ETR (Extra Tree Regressor), Orthogonal Matching Pursuit (OMP), Gradient Boosting Regressor (GBR), Huber Regressor (HR), and LGBM (Light Gradient Boosting Machine).

Table 10
Comparison of the algorithms considering noise.

Algorithm	R_0	R_1	L_0	L_1	C_0	C_1	Mean MAPE
LR	3.06	**2.92**	1.62	**0.88**	5.64	**0.56**	**2.78**
ETR	**2.47**	7.50	**1.54**	2.07	5.62	1.39	3.43
OMP	7.94	10.19	3.52	2.82	5.64	1.04	5.86
GBR	2.77	7.12	**1.54**	1.99	5.74	1.17	3.72
HR	5.42	5.86	2.32	1.61	**5.61**	0.58	3.23
LGBM	2.78	6.76	1.67	1.84	5.63	1.12	3.30

Analyzing Table 10, the best algorithm varies depending on the parameter considered. However, the Linear Regressor consistently demonstrates the best performance in most cases and achieves the lowest average MAPE across all parameters. It is important to note that linear regression, despite being simpler than the other algorithms, also imposes a lower computational burden. Another key aspect is that linear regression requires fewer samples to achieve satisfactory performance. To verify this behavior, we conducted an analysis by varying the number of samples and evaluating the average MAPE across all parameters. The results are presented in Fig. 8.

Fig. 8 shows that the method achieved acceptable performance with a reduced number of samples, approximately 200. This demonstrates that a high-accuracy solution could be obtained while accounting for both random and systematic errors, with lower computational cost and a smaller dataset. Therefore, the proposed methodology makes a significant contribution to power cable parameter estimation by considering both the complete modeling of the equipment and the estimation of electrical parameters using a non-invasive, measurement-based approach.

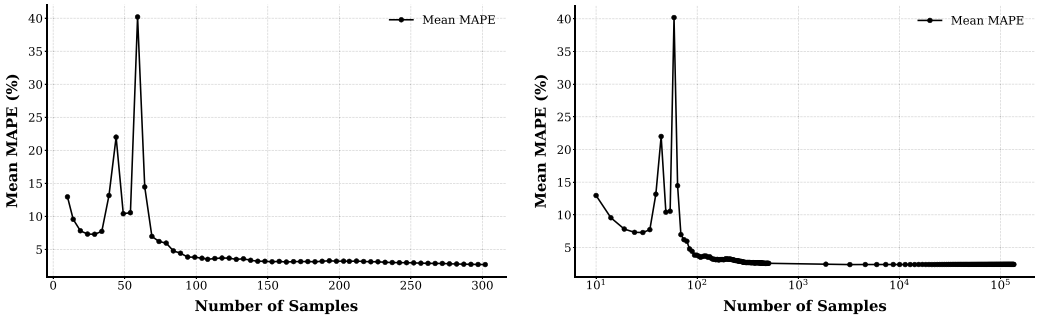
As a direction for future work, the proposed methodology will be validated using real-world data obtained from operational power systems. This includes leveraging PMU measurements collected from submarine cables through collaborations with utility companies or, alternatively, by employing a scaled prototype in a controlled laboratory environment. Such partnerships will enable the evaluation of the model under realistic conditions, including environmental variability, measurement noise, and hardware constraints. This step is essential to assess the robustness and generalizability of the approach beyond synthetic data scenarios.

6. Conclusion

This research presented two main contributions: the proposal of a novel parameter estimation methodology for submarine power cables based on machine learning techniques, and the performance ranking of different hypothesis classes applied to the estimation process.

The phasorial measurements were simulated at the terminals of the submarine cable through modeling based on FEM, which was also applied to calculate the reference values of the cable parameters. Phasor measurement series were considered both with and without noise, which was modeled with a Gaussian distribution. In this context, the framework combines FEM-based modeling, frequency-domain simulations with multi-mode propagation, and realistic noise insertion aligned with PMU measurement characteristics.

Most linear and tree-based regression models demonstrated accurate parameter estimations both in the presence and absence of noise in the phasor measurements. Furthermore, linear regression proved to be notably accurate and robust for parameter estimation even when noisy measurements were available at only one terminal of the power cable. This characteristic represents a significant advantage of the proposed methodology, particularly in scenarios where measurement series are accessible exclusively at a single end of the submarine cable. For example, even with a limited dataset composed of only 200 training



(a) Mean MAPE Evolution for first samples. (b) Mean MAPE Evolution part 2.

Fig. 8. Analysis of the method considering different samples.

samples under severe noise conditions, the estimation errors remained below 1%, thereby demonstrating the robustness and accuracy of the proposed methodology for real-world offshore transmission systems.

CRediT authorship contribution statement

Felipe P. de Albuquerque: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rafael Nascimento:** Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Gabriel de Castro Biage:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Rooney R.A. Coelho:** Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Ronaldo F. Ribeiro Pereira:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Eduardo C. Marques da Costa:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Mario L. Pereira Filho:** Visualization, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation. **Cassio G. Lopes:** Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **José R. Cardoso:** Visualization, Supervision, Resources, Project administration, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Eduardo Coelho Marques da Costa reports financial support, administrative support, and article publishing charges were provided by University of São Paulo. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the support of the RCGI – Research Centre for Greenhouse Gas Innovation (23.1.8493.1.9), hosted by the University of São Paulo (USP), sponsored by FAPESP – São Paulo Research Foundation, Brazil (2020/15230-5), and sponsored by TotalEnergies, Brazil, and the strategic importance of the support given by ANP, Brazil (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&DI levy regulation

CAPES, Brazil - *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*.

CNPq - National Council for Scientific and Technological Development, Brazil (Grants 402830/2021-0, 307485/2021-8, and 176725/2023-7).

FAPESP - São Paulo Research Foundation, Brazil (Grants 2021/01325-7 and 2024/00261-3).

Appendix

A.1. Parameters of the cable using FEM-based method

The cable is composed of three similar phase conductors, which are covered by three similar shields, and an external armor. Thus, the cable is characterized by three square matrices with dimension seven: resistance $\boldsymbol{R}'$, inductance $\boldsymbol{L}'$, and capacitance $\boldsymbol{C}'$.

The resistance matrix (9) is determined from the current density and magnetic flux, and given in Ω/km. The first three elements in the main diagonal are the longitudinal resistance of the phase conductors, whereas the following three are the same values for the respective shields. The last term in the main diagonal is the external armor of the cable. The terms out of the main diagonal are mutual resistance between conductors, shield and external armor.

Analogously, the inductance matrix $\boldsymbol{L}'$ is expressed in mH/km.

$$\boldsymbol{R}' = \begin{bmatrix} 0.1066 & 0.0580 & 0.0580 & 0.0639 & 0.0580 & 0.0580 & 0.0598 \\ 0.0580 & 0.1066 & 0.0580 & 0.0580 & 0.0639 & 0.0580 & 0.0598 \\ 0.0580 & 0.0580 & 0.1066 & 0.0580 & 0.0580 & 0.0639 & 0.0598 \\ 0.0639 & 0.0580 & 0.0580 & 0.0926 & 0.0580 & 0.0580 & 0.0598 \\ 0.0580 & 0.0639 & 0.0580 & 0.0580 & 0.0926 & 0.0580 & 0.0598 \\ 0.0580 & 0.0580 & 0.0639 & 0.0580 & 0.0580 & 0.0926 & 0.0598 \\ 0.0598 & 0.0598 & 0.0598 & 0.0598 & 0.0598 & 0.0598 & 0.0828 \end{bmatrix} \tag{9}$$

$$\boldsymbol{L}' = \begin{bmatrix} 0.0021 & 0.0017 & 0.0017 & 0.0018 & 0.0017 & 0.0017 & 0.0017 \\ 0.0017 & 0.0021 & 0.0017 & 0.0017 & 0.0018 & 0.0017 & 0.0017 \\ 0.0017 & 0.0017 & 0.0021 & 0.0017 & 0.0017 & 0.0018 & 0.0017 \\ 0.0018 & 0.0017 & 0.0017 & 0.0018 & 0.0017 & 0.0017 & 0.0017 \\ 0.0017 & 0.0018 & 0.0017 & 0.0017 & 0.0018 & 0.0017 & 0.0017 \\ 0.0017 & 0.0017 & 0.0018 & 0.0017 & 0.0017 & 0.0018 & 0.0017 \\ 0.0017 & 0.0017 & 0.0017 & 0.0017 & 0.0017 & 0.0017 & 0.0017 \end{bmatrix} \tag{10}$$

Table 11
Transmission system parameters used to simulate the measurements in a computational environment.

Parameter	Value
$\dot{V}_{A,n}; \dot{V}_{B,n}; \dot{V}_{C,n}$	69 kV
$r_{(1)}$	0.0589 Ω/km
$l_{(1)}$	0.347 mH/km
$c_{(1)}$	77.4 nF/km
$r_{(0)}$	0.0665 Ω/km
$l_{(0)}$	0.325 mH/km
$c_{(0)}$	52.1 nF/km
Rated frequency (f)	60 Hz
Length (ℓ_{sec})	30 km
Power factor	$f_p = 0.83$
Unbalance factor	$f_{ac} = 0.1$
Z_C	70.19 Ω
P_A	$(1) \cdot \mathrm{P}_C$
P_B	$(1 - f_{ac}) \cdot \mathrm{P}_C$
P_C	$(1 + f_{ac}) \cdot \mathrm{P}_C$
Q_A	$P_A \cdot \sqrt{1 - fp^2}/fp$
Q_B	$P_B \cdot \sqrt{1 - fp^2}/fp$
Q_C	$P_C \cdot \sqrt{1 - fp^2}/fp$

The self and mutual capacitances in $\boldsymbol{C}'$ are determined from the electric field and difference of potential among phase conductors, shields and armor. Eq. (11) expresses the longitudinal capacitance in nF/km.

$$\boldsymbol{C}' = \begin{bmatrix} 94.3 & 0 & 0 & -94.3 & 0 & 0 & 0 \\ 0 & 94.3 & 0 & 0 & -94.3 & 0 & 0 \\ 0 & 0 & 94.3 & 0 & 0 & -94.3 & 0 \\ -94.3 & 0 & 0 & 425 & -101 & -101 & -129 \\ 0 & -94.3 & 0 & -101 & 425 & -101 & -129 \\ 0 & 0 & -94.3 & -101 & -101 & 425 & -129 \\ 0 & 0 & 0 & -129 & -129 & -129 & 3980 \end{bmatrix} \tag{11}$$

A.2. Parameters of the system

A.3. Statistical behavior of the measurements

Understanding the statistical behavior of the corrupted phasor measurements is essential to properly assess the performance and robustness of estimation and monitoring techniques. In particular, many classical estimation methods and data-driven approaches rely on the assumption that the measurement noise follows a Gaussian distribution. However, the presence of both random and systematic errors in the proposed noise model leads to a more complex statistical profile.

To demonstrate that the magnitude measurements do not follow a Gaussian distribution, various methodologies can be applied. These can be broadly categorized into graphical and statistical hypothesis-based approaches.

The first graphical method is the histogram, as illustrated in Fig. 7, which already suggests the presence of asymmetry and deviations from normality. Another powerful graphical tool is the Quantile-Quantile (QQ) plot [47], shown in Fig. 9 for the magnitude of $\hat{I}_{As}$. From this figure, it becomes evident that the data deviates from a normal distribution in multiple regions, particularly in the tails. These deviations are indicative of heavy-tailed behavior or skewness, which are not consistent with Gaussian statistics.

In addition to graphical techniques, formal statistical tests can be employed to assess normality. According to [49], the Shapiro–Wilk test is among the most powerful tests for detecting departures from normality, particularly for sample sizes smaller than 5000. This test was applied to the simulated magnitude measurements, and the results are

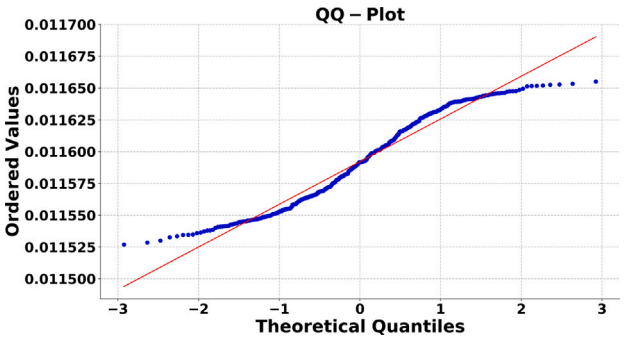


Fig. 9. QQ-plot analysis of $|\hat{I}_{As}|$ for a single MC realization.

Table 12
Results of the Shapiro–Wilk normality test for different measurements.

Measurement	Statistic	p-value
$\lvert\hat{I}_{As}\rvert$	0.9547	$9.76 \cdot 10^{-10}$
$\lvert\hat{V}_{As}\rvert$	0.9562	$1.56 \cdot 10^{-9}$
$\lvert\hat{I}_{Ar}\rvert$	0.9643	$2.66 \cdot 10^{-8}$
$\lvert\hat{V}_{Ar}\rvert$	0.9598	$5.26 \cdot 10^{-9}$
$\lvert\hat{I}_{Bs}\rvert$	0.9665	$6.18 \cdot 10^{-8}$
$\lvert\hat{V}_{Bs}\rvert$	0.9496	$1.96 \cdot 10^{-10}$
$\lvert\hat{I}_{Br}\rvert$	0.9483	$1.30 \cdot 10^{-10}$
$\lvert\hat{V}_{Br}\rvert$	0.9609	$8.11 \cdot 10^{-9}$
$\lvert\hat{I}_{Cs}\rvert$	0.9668	$7.02 \cdot 10^{-8}$
$\lvert\hat{V}_{Cs}\rvert$	0.9653	$6.88 \cdot 10^{-9}$
$\lvert\hat{I}_{Cr}\rvert$	0.9669	$7.29 \cdot 10^{-8}$
$\lvert\hat{V}_{Cr}\rvert$	0.9541	$7.80 \cdot 10^{-10}$

summarized in Table 12. The rejection of the normality hypothesis at a 5% significance level reinforces the graphical evidence and supports the conclusion that the corrupted measurements exhibit non-Gaussian behavior.

This observation is particularly relevant for the monitoring and estimation problem, as it justifies the need for robust or non-Gaussian-aware methods. By characterizing the statistical behavior of the inputs, the study provides useful insights into the limitations of traditional assumptions and motivates the development or application of techniques that are resilient to such deviations.

In Shapiro–Wilk test, the statistic is calculated using the samples and very close to 1 (greater than 0.98) suggest that the data has a normal distribution. For the p-value, if the calculated value is smaller than the significance level, typically 5%, the null hypothesis is rejected and the data are assumed non-Gaussian. This way, considering the results shown in Table 12, it is possible to conclude that there is no statistical evidence that the data presents a Gaussian distribution.

Data availability

Data will be made available on request.

References

[1] de Albuquerque FP, et al. Data-centric approach for online P-margin estimation from noisy phasor measurements. Energy Rep 2023;10:2194–205.
[2] Sitzia C, Muscas C, Pegoraro PA, Solinas AV, Sulis S. Enhanced PMU-based line parameters estimation and compensation of systematic measurement errors in power grids considering multiple operating conditions. IEEE Trans Instrum Meas 2022;71:1–12.
[3] Patel B, Bera P. A new transmission line parameter estimation technique and its impact on fault localization. IEEE Trans Instrum Meas 2023;72:1–8.

[4] Hofmman L. Series expansions for line series impedances considering different specific resistances, magnetic permeabilities and dielectric permittivities of conductors, air, and ground. IEEE Trans Power Deliv 2003;18(2):564–70.
[5] Shaikh MS, Raj S, Latif SA, Mbasso WF, Kamel S. Optimizing transmission line parameter estimation with hybrid evolutionary techniques. IET Gener Transm Distrib 2024;18(9):1795–814.
[6] Schultze R, Schegner P, Zivanovic R. Parameter identification of unsymmetrical transmission lines using fault records obtained from protective relays. IEEE Trans Power Deliv 2011;26(2):1265–72.
[7] Pegoraro PA, Sitzia C, Solinas AV, Sulis S. Transmission line parameters estimation in the presence of realistic PMU measurement error models. Measurement 2023;218:113175.
[8] Moghanian M, Dobakhshari AS. Accurate Kalman filter based estimation of transmission line parameters utilizing synchronized phasor measurements. Electr Power Syst Res 2024;230:110218.
[9] De Albuquerque FP, et al. Nonlinear analysis on transmission line parameters estimation from noisy phasorial measurements. IEEE Access 2021;10:1720–30.
[10] Gerez C, Ângelos EWS, Albuquerque FP, Costa ECM, Sguarezi Filho AJ, Liboni LHB. Estimation of the frequency-dependent parameters of transmission lines by using synchronized measurements. IEEE Access 2022;10:17526–41.
[11] Wehenkel A, Mukhopadhyay A, Le Boudec J-Y, Paolone M. Parameter estimation of three-phase untransposed short transmission lines from synchrophasor measurements. IEEE Trans Instrum Meas 2020;69(9):6143–54.
[12] Ahmad T, Senroy N. Statistical characterization of PMU error for robust WAMS based analytics. IEEE Trans Power Syst 2019;35(2):920–8.
[13] Adalja D, Kalita K, Čepová L, Patel P, Mashru N, Jangir P, et al. Advancing truss structure optimization—A multi-objective weighted average algorithm with enhanced convergence and diversity. Results Eng 2025;25:104241.
[14] Aljaidi M, Mashru N, Patel P, Adalja D, Jangir P, Arpita, et al. MORIME: A multi-objective RIME optimization framework for efficient truss design. Results Eng 2025;25:103933.
[15] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw 2014;69:46–61.
[16] Bandopadhyay J, Roy PK. Application of hybrid multi-objective moth flame optimization technique for optimal performance of hybrid micro-grid system. Appl Soft Comput 2020;95:106487.
[17] Zaki AM, Mahmood S. K-nearest neighbors approach to analyze and predict air quality in Delhi. J Artif Intell Metaheuristics 2025;9(1):34–43.
[18] El-kenawy EM. A review of machine learning models for predicting air quality in urban areas. Metaheuristic Optim Rev 2025;3(2):33–46.
[19] Wang S, Ma C, Gao H, Deng D, Fernandez C, Blaabjerg F. Improved hyperparameter Bayesian optimization-bidirectional long short-term memory optimization for high-precision battery state of charge estimation. Energy 2025;136598.
[20] Wang S, Fan Y, Jin S, Takyi-Aninakwa P, Fernandez C. Improved anti-noise adaptive long short-term memory neural network modeling for the robust remaining useful life prediction of lithium-ion batteries. Reliab Eng Syst Saf 2023;230:108920.
[21] Wang S, Gao H, Takyi-Aninakwa P, Guerrero JM, Fernandez C, Huang Q. Improved multiple feature-electrochemical thermal coupling modeling of lithium-ion batteries at low-temperature with real-time coefficient correction. Prot Control Mod Power Syst 2024;9(3):157–73.
[22] Van Breugel F, Kutz JN, Brunton BW. Numerical differentiation of noisy data: A unifying multi-objective optimization framework. IEEE Access 2020;8:196865–77.
[23] Yang N-C, Sen A. Parameter estimation in unbalanced three-phase distribution lines using general regression neural networks with inconsistent data handling capacity. Appl Soft Comput 2023;133:109936.
[24] Varghese AC, Pal A, Dasarathy G. Transmission line parameter estimation under non-Gaussian measurement noise. IEEE Trans Power Syst 2023;38(4):3147–62.
[25] Paul CR. Analysis of multiconductor transmission lines. John Wiley & Sons; 2007.
[26] Benato R, Sessa SDambone, Poli M, Sanniti F. Sequence impedances of land single-core insulated cables: direct formulae and multiconductor cell analyses compared with measurements. Energies 2020;13(5):1084.
[27] Hafner AA, Caballero PT, Monteiro JHA, Costa ECM, Kurokawa S, Luz MVF, et al. Modeling of power cables with arbitrary cross section: from parameter calculation to electromagnetic transients simulation. J Control Autom Electr Syst 2017;28:405–17.
[28] Salles MBC, Costa MC, Filho MLP, Cardoso JR, Marzo GR. Electromagnetic analysis of submarine umbilical cables with complex configurations. IEEE Trans Magn 2010;46(8):3317–20.
[29] A. Ametani. Approximate method for calculating the impedances of multiconductors with cross sections of arbitrary shapes. Electr Eng Japan 1992;112(2):117–23.
[30] Ametani A. A general formulation of impedance and admittance of cables. IEEE Trans Power Appar Syst PAS- 1980;99(3):902–10.
[31] A. Budner. Introduction of frequency-dependent line parameters into an electromagnetic transients program. IEEE Trans Power Appar Syst PAS-1970;89(1):88–97.
[32] Gómez P, Uribe FA. The numerical Laplace transform: an accurate technique for analyzing electromagnetic transients on power system devices. Int J Electr Power Energy Syst 2009;31(2):116–23.
[33] Costa ECM, Kurokawa S, Pissolato J, Prado AJ. Efficient procedure to evaluate electromagnetic transients on three-phase transmission lines. IET Gener Transm Distrib 2010;4(9):1069–81.
[34] Givaki K, Seyedzadeh S. Machine learning based impedance estimation in power system. In: 8th renewable power generation conference - RPG 2019. IET; 2019, p. 1–6.
[35] Qin C, Vyakaranam B, Etingov P, Venetos M, Backhaus S. Machine learning based network parameter estimation using AMI data. In: 2022 IEEE power & energy society general meeting. IEEE; 2022, p. 1–5.
[36] Ramasamy S, Ganesan K, Arunachalam V. Power flow parameter estimation in power system using machine learning techniques under varying load conditions. Int J Electr Electron Res 2022;10(4):1299–305.
[37] Khalili R, Abur A. Transmission line parameter error identification and estimation in three-phase networks. IEEE Trans Power Syst 2021;37(3):2271–82.
[38] Kusic GL, Garrison DL. Measurement of transmission line parameters from SCADA data. In: IEEE PES power systems conference and exposition. IEEE; 2004, p. 440–5.
[39] Grafarend EW. Linear and nonlinear models: fixed effects, random effects, and mixed models. de Gruyter; 2006.
[40] Frigo G, Pegoraro PA, Toscani S. Tracking power systems events: PMU, reporting rate, interpolation. In: Proc. 2022 int. conf. on smart grid synchronized measurements and analytics. IEEE; 2022, p. 1–6.
[41] Borchani H, Varando G, Bielza C, Larranaga P. A survey on multi-output regression. Wiley Interdiscip Rev: Data Min Knowl Discov 2015;5(5):216–33.
[42] Shalev-Shwartz S, Ben-David S. Understanding machine learning: from theory to algorithms. New York, NY, USA: Cambridge University Press; 2014.
[43] Dasgupta K, Soman SA. Line parameter estimation using phasor measurements by the total least squares approach. In: Proc. 2013 IEEE power & energy society general meeting. p. 1–5.
[44] Asprou M, Kyriakides E. Estimation of transmission line parameters using PMU measurements. In: Proc. 2015 IEEE power & energy society general meeting. p. 1–5.
[45] Milojević V, Čalija S, Rietveld G, Ačanski MV, Colangelo D. Utilization of PMU measurements for three-phase line parameter estimation in power systems. IEEE Trans Instrum Meas 2018;67(10):2453–62.
[46] Frigo G, Derviškadić A, Bach A, Paolone M. Statistical model of measurement noise in real-world PMU-based acquisitions. In: Proc. 2019 international conference on smart grid synchronized measurements and analytics. IEEE; 2019, p. 1–8.
[47] Castello P, Muscas C, Pegoraro PA. Statistical behavior of PMU measurement errors: An experimental characterization. IEEE Open J Instrum Meas 2022;1:1–9.
[48] Brown M, Biswal M, Brahma S, Ranade SJ, Cao H. Characterizing and quantifying noise in PMU data. In: Proc. 2016 IEEE power and energy society general meeting. IEEE; 2016, p. 1–5.
[49] Yap BW, Sim CH. Comparisons of various types of normality tests. J Stat Comput Simul 2011;81(12):2141–55.