

Article

Phasor Measurement Unit-Driven Estimation of Transmission Line Parameters Using Variable Noise Model

Felipe Proença de Albuquerque ^{*}, Rafael Nascimento, Carlos A. Prete, Jr.  and Eduardo Coelho Marques da Costa

Escola Politécnica, Universidade de São Paulo, São Paulo 05508-220, Brazil; rafael2.nascimento@usp.br (R.N.); carlos.prete@usp.br (C.A.P.J.); educosta@usp.br (E.C.M.d.C.)

* Correspondence: felipe.proenca.albuquerque@usp.br

Abstract: Accurate parameters are crucial in modern energy systems to ensure the reliable operation of all components. Given the substantial volume of data in monitored systems, high-performance methods are necessary. This paper proposes a new Bayesian multi-output regressor for estimating the parameters of a three-phase transmission line. The presented approach achieves acceptable accuracy in parameter estimation using only one end of the line. The Bayesian regressor is developed using information derived from the data themselves, eliminating the need to explicitly model the system. This capability allows the method to estimate parameters while accommodating different noise models, even in the presence of systematic errors and non-Gaussian random noise. The methodology was validated on various systems, including a two-bus system, IEEE 14-bus, IEEE 39-bus, and IEEE 118-bus, under diverse conditions such as varying sample sizes, loads, and noise levels. These tests demonstrate the robustness of the proposed approach.

Keywords: three-phase transmission lines; Bayesian regressor; phasor measurement units; non-Gaussian noise; parameter estimation



Citation: de Albuquerque, F.P.; Nascimento, R.; Prete, C.A., Jr.; Coelho Marques da Costa, E. Phasor Measurement Unit-Driven Estimation of Transmission Line Parameters Using Variable Noise Model. *Energies* **2024**, *17*, 3587. <https://doi.org/10.3390/en17143587>

Academic Editor: Zheng Xu

Received: 1 June 2024

Revised: 16 July 2024

Accepted: 19 July 2024

Published: 21 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate estimation of power system parameters is a significant concern in modern systems, especially within the context of smart grids. Transmission line parameters, considering both positive and zero sequences, find application in various scenarios, including modeling of protection relays for distance protection [1], power system state estimation [2], and investigation of voltage stability [3]. Thus, evaluating these parameters with a high accuracy is imperative for ensuring the security and reliability of power system operations.

Studies in the literature have shown that the difference between the stored values of the transmission line parameters and the actual values can reach up to 30% [4,5]. These deviations directly affect the performance of the state estimators and the power flow solutions [6]. Therefore, a robust methodology to estimate the parameters of a given power system using the available noisy measurements is necessary.

Over the years, several papers have attempted to address transmission line parameter estimation [2,6–13]. Each of them focuses on a specific model of the transmission line or a different methodology to describe the observed data. In [6], a method to estimate the parameters of a three-phase transmission line employing robust linear regression in the presence of outliers was proposed. However, the methodology did not address the challenge of modeling random noise in measurements obtained from PMUs.

The authors of [7] presented a methodology to calculate the parameters of a short transmission line by developing a rigorous model for random noise based on an accurate noise covariance matrix, considering the realistic specifications of noise introduced by instrument transformers and PMUs. In this work, the solution is obtained using weighted least squares using the covariance matrix of the measurements. Nevertheless, systematic errors are neglected, and the projections are assumed to be Gaussian for all noise levels, which is not valid in general [11].

The work developed by [14] proposed a solution based on the Kalman filter, considering the presence of noise in voltage and current measurements. Additionally, the authors presented an analytical derivation for the covariance matrix for the lumped model of a transmission line. However, the authors ignored the presence of systematic errors and the possibility of non-Gaussian noise distributions. Moreover, the application of the method requires the presence of PMUs at both ends of the transmission lines.

In [12,13], the authors presented a multi-branch method to estimate the parameters of the transmission lines and the systematic errors present in the chain of measurements used to obtain a phasor measurement. The random errors are modeled as Gaussian distributions for both magnitude and phase, while the systematic error is assumed to follow a uniform distribution, i.e., it has the same contribution for each sample. The solution is calculated by applying a Tikhonov regularization framework.

The need for a method that requires measurements only from one side of a transmission line was introduced in [15]. This paper focused on identifying erroneous parameters and subsequently estimating them by applying the extended error function for state estimation in power systems. However, the authors did not address the problem of modeling noise in phasor measurements, although they heavily relied on it in the estimation process.

Additionally, it is important to highlight the work proposed by [8], where the authors developed a method to handle non-Gaussian random noise in voltage and current phasor measurements. This methodology was supported by recent studies that examined the Gaussian assumption of random errors [16,17], indicating that the process of obtaining phasors is multimodal and that the Gaussian assumption may not be the most suitable choice.

In this context, this work presents a novel, model-free methodology to estimate transmission line parameters. This approach does not rely on transmission line equations to evaluate parameters; instead, it is a data-centric solution. Moreover, the proposed method is capable of estimating parameters using only one-sided measurements with an acceptable performance. Additionally, the developed multi-output regressor can accommodate different noise models, including Gaussian random error, systematic errors, and non-Gaussian random noise, outperforming existing methods under the same studied conditions. Therefore, the main innovations introduced by this paper are as follows:

- A model-free methodology is developed to assess the parameters of a three-phase transmission line that does not necessitate measurements from both ends of the transmission line;
- The Bayesian regressor enables the construction of credible intervals for the estimation, thereby enhancing confidence in the results;
- The method is capable of handling various noise models, resulting in estimations with high performances across all cases;
- The method was tested on systems with varying dimensions, numbers of samples in the database, load conditions, and noise levels.

2. Description of the Employed Transmission Line Model

In this work, the model for long-transmission lines presented in [18] was studied. The multi-phase model, which includes consideration of ground effects, is depicted in Figure 1. In this figure, the parameters z_{aa} , z_{bb} , and z_{cc} represent the self-impedance per unit length (Ω/km), while z_{ab} , z_{ac} , and z_{bc} represent the mutual inductances. As for capacitances, c_{ab} , c_{ac} , and c_{bc} denote the capacitances between phases per unit length (F/km), and c_g denotes the ground capacitance.

The three-phase transmission lines are represented by the impedance matrix $\mathbf{Z}'$ and the capacitance matrix $\mathbf{C}'$. The parameters are shown in (1).

$$\mathbf{Z}' = \begin{bmatrix} r_{aa} + j\omega l_{aa} & r_{ab} + j\omega l_{ab} & r_{ac} + j\omega l_{ac} \\ r_{ba} + j\omega l_{ba} & r_{bb} + j\omega l_{bb} & r_{bc} + j\omega l_{bc} \\ r_{ca} + j\omega l_{ca} & r_{cb} + j\omega l_{cb} & r_{cc} + j\omega l_{cc} \end{bmatrix},$$

$$\mathbf{C}' = \begin{bmatrix} c_g + c_{ab} + c_{ac} & c_{ab} & c_{ac} \\ c_{ba} & c_g + c_{ba} + c_{bc} & c_{bc} \\ c_{ca} & c_{cb} & c_g + c_{ca} + c_{cb} \end{bmatrix}, \quad (1)$$

where $\omega = 2\pi f$, and f is the frequency of the power system.

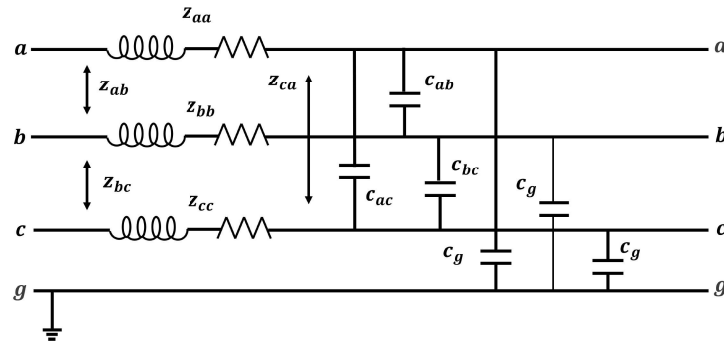


Figure 1. Three-phase transmission line model with mutual coupling between phases used to obtain the sequence parameters.

Since we are assuming that the method deals with long transmission lines, it is reasonable to presume that the line is transposed. Thus, the model for the per-length parameters is represented by the matrices in (2):

$$\mathbf{Z}' = \begin{bmatrix} z_p & z_m & z_m \\ z_m & z_p & z_m \\ z_m & z_m & z_p \end{bmatrix} \quad \mathbf{C}' = \begin{bmatrix} c_p & c_m & c_m \\ c_m & c_p & c_m \\ c_m & c_m & c_p \end{bmatrix}. \quad (2)$$

Therefore, by applying symmetrical components, the sequence parameters are obtained while still considering the per-length units:

$$z_{(1)} = z_p - z_m, \quad c_{(1)} = c_p - c_m, \quad z_{(0)} = z_p + 2z_m, \quad c_{(0)} = c_p + 2c_m.$$

Using the parameters per length, it is possible to obtain the parameters using hyperbolic corrections [19] using the following equations:

$$y_{(s)} = j\omega c_{(s)}, \quad Z_{C,(s)} = \sqrt{z_{(s)}/y_{(s)}}, \quad \gamma_{(s)} = \sqrt{z_{(s)}y_{(s)}},$$

$$Z_{(s)} = R_{(s)} + j\omega L_{(s)} = Z_{C,(s)} \sinh(\gamma_{(s)} \cdot l_{sec})$$

$$Y_{(s)} = \frac{2}{Z_{C,(s)}} \tanh\left(\gamma_{(s)} \frac{l_{sec}}{2}\right)$$

$$R_{(s)} = \text{Re}\{Z_{(s)}\}, \quad L_{(s)} = \frac{\text{Im}\{Z_{(s)}\}}{\omega}, \quad C_{(s)} = \frac{\text{Im}\{Y_{(s)}\}}{\omega},$$

where the index (s) might be 1 for a positive sequence or 0 for a zero sequence.

The provided expressions delineate the circuit per sequence, as illustrated in Figure 2. Within this representation, $\dot{V}_{(s),k}$ and $\dot{V}_{(s),k}$ denote the phasors of voltage and current at the bus k for sequence s , while $\dot{V}_{(s),l}$ and $\dot{V}_{(s),l}$ represent the respective phasors at bus l . It is worth noting that each sequence needs its own circuit depiction. Notably, the

negative sequence mirrors the positive sequence in terms of parameters, thereby rendering it unnecessary for the analysis.

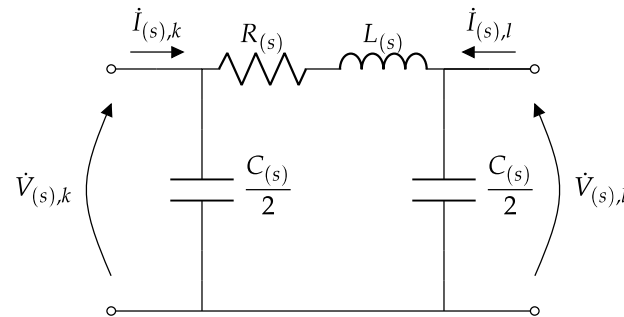


Figure 2. Π -model of each sequence circuit of the transmission line used to generate the measurements at the sender and receiver ends.

To simplify the notation, the sequence notation will be omitted from this point forward, as the equations can be handled identically for positive and zero sequences. Any distinctions between sequences will be highlighted as needed.

3. Problem Formulation

In this section, we describe the methodology used to generate the database utilized for estimating the parameters, as well as the formulation of the problem concerning the supervised learning approach and the algorithm built to evaluate the parameters.

3.1. Building the Database

In a supervised learning framework, the methodology employed to construct the database holds significant importance in achieving a viable solution. Given the scarcity of literature addressing transmission line parameter estimation through a machine learning paradigm [20–22], this paper introduces a novel approach to generating the training data. This methodology takes into account the characteristics obtained from historical data documented in the technical literature.

According to studies in the literature, parameter errors typically fall within the range of approximately 10% to 30% of the true parameter value [4,15]. Consequently, it is anticipated that the database should encompass parameter values within this range. Thus, if p^* represents the exact value of a specific parameter, then p , the parameter with inaccuracy, should fall within the following bounds:

$$p \in [0.7p^*, 1.3p^*].$$

Given that extreme parameter values are less frequently observed compared to those with smaller errors, the Gaussian distribution is suitable for data generation. By employing the empirical rule of 3σ [23], it is possible to establish a relationship between the maximum observed error and the standard deviation of the Gaussian distribution. This relationship can be succinctly expressed through the following rule:

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

where $\mathcal{P}(\cdot)$ is the probability density function, X is a Gaussian random variable with mean μ and standard deviation σ . Using this rule with the formulation of the problem, we have the following:

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

Hence, to generate data for parameters with a mean value of p^* and maximum observed values of $1.3p^*$ and minimum values of $0.7p^*$, we use a truncated Gaussian

distribution with a mean of $\mu^* = p^*$ and a standard deviation of $\sigma^* = 0.1p^*$, truncated in the interval $[\mu^* - 3\sigma^*, \mu^* + 3\sigma^*]$.

For distribution lines, the approach of varying the parameters within a specific range of values was tested in [24]. However, in that work, the authors used a maximum variation of 5%, which simplifies the problem in terms of supervised learning because the target space is smaller. Additionally, for transmission lines, it is well-established that parameter values can vary by up to 30%.

Although this method for building the database is based on practical scenarios of transmission lines, it presents some concerns. If the parameter to be estimated falls outside the range of possible parameters, the algorithm needs to be retrained, considering the new range of possible values. Furthermore, the random generation of cases can lead to unfeasible values for the speed of signal propagation. To address this problem, an additional step was necessary to verify the generated speed of propagation. Regarding the speed of light in a vacuum (c), the values for the positive ($v_{(1)}$) and zero ($v_{(0)}$) sequences should be as follows:

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

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

The entire dataset consists of N samples or instances for valid cases, after the verification of both conditions. Before applying any machine learning algorithm, this dataset is split into 80% for training the algorithm and 20% for testing. It is important to note that this process ensures that the training process is completely blind to the test dataset.

3.2. Supervised Learning Setup

In modern power systems, advances in monitoring technology are increasing the use of Phasor Measurement Units in several buses of the system [25–27]. Additionally, the greater frequency reporting rate compared with common supervisory systems (SCADA) and the possibility to employ synchronized data in energy system problems make this device more suitable to deal with linear and nonlinear problems. Generally, the transmission line parameter estimation problem is nonlinear [11], requiring more sophisticated noise models [8].

Considering the presented context, this work presents a methodology to estimate the parameters based on a supervised learning approach. For such a task, initially, we have considered as features the voltage and current phasors for the sender (bus k) and receiver (bus l) ends in a three-phase transmission line. The goal is to estimate the parameters for positive and zero sequences. Mathematically, this is expressed as follows:

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

$$\text{Using } \{\dot{V}_k, \dot{I}_k, \dot{V}_l, \dot{I}_l\}_{A,B,C}$$

where the indices A , B , and C represent the system phases, and k and l are the indices of buses connected by a transmission line. Thus, the phasor measurements can be written as follows:

$$\dot{V}_k = |\dot{V}_k|e^{j\theta_k}, \quad \dot{I}_k = |\dot{I}_k|e^{j\varphi_k},$$

$$\dot{V}_l = |\dot{V}_l|e^{j\theta_l}, \quad \dot{I}_l = |\dot{I}_l|e^{j\varphi_l},$$

where both magnitude and angles are treated as features in the estimation process.

This way, each vector of features, denoted as $\mathbf{x}_n$, contains the information of the three-phase phasors in the sender and receiver ends. Each row in the matrix of features $\mathbf{X}$ represents an instance or sample for a specific observation of the set of features $\mathbf{x}_n$. Additionally, each parameter is a different target t . A method to deal with multi-output regression is to perform one model per target [28]. Thus, the theory can be developed for a

single target and applied for each desired parameter, ignoring the remaining parameters in the estimation.

3.3. Statistical Model

In this section, we describe the Bayesian ridge regression model[29] used to estimate the transmission line parameters. Even though this is a hierarchical Bayesian model, it assumes convenient prior distributions and approximates the posterior distributions of some parameters based on their maximum a posteriori estimates to produce low-complexity approximations of the transmission line parameters. As a Bayesian model, the output is a posterior distribution, with the mean serving as a point estimate and the standard deviation indicating the level of uncertainty.

To train the model, we use N pairs $\{\mathbf{x}_n, t_n\}, n = 1, \dots, N$, where $\mathbf{x}_n = [x_{j1}, x_{j2}, \dots, x_{jd}]^T$ is the vector of features and t_n is the label; that is, the value of the target for the n -th sample. We assume the measurement model is given by the following:

$$t_n = \mathbf{x}_n^T \mathbf{w} + \varepsilon_n,$$

where $\mathbf{w} = [w_1, w_2, \dots, w_d]^T$ and ε_n is an independent and identically distributed zero-mean Gaussian process with variance σ^2 . Thus, $p(t_n|\mathbf{x}_n, \sigma^2) = \mathcal{N}(t_n|\mathbf{x}_n^T \mathbf{w}, \sigma^2)$, being $\mathcal{N}(\mu, \sigma^2)$ a Gaussian distribution with mean μ and variance σ^2 .

Defining $\beta = \frac{1}{\sigma^2}$ and assuming that each sample is independent and identically distributed [30], the likelihood of the observed data can be written as follows:

$$p(\mathbf{t}|\mathbf{w}, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \exp\left\{-\frac{\beta}{2}\|\mathbf{t} - \mathbf{X}\mathbf{w}\|^2\right\},$$

where $\mathbf{t} = [t_1, t_2, \dots, t_N]^T, \mathbf{X} \in \mathbb{R}^{N \times d}$, such that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$. The prior distribution for β is chosen as a Gamma distribution; a common choice for scale parameters:

$$p(\beta) = \text{Gamma}(\beta|c, d),$$

with

$$\begin{aligned} \text{Gamma}(\alpha|a, b) &= \Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha}, \\ \Gamma(a) &= \int_0^\infty t^{a-1} e^{-t} dt. \end{aligned}$$

Following [29], we assume a hierarchical prior distribution for the parameter $\mathbf{w}$ by defining a vector of hyperparameters $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_d]$ representing the standard deviation of $\mathbf{w}$:

$$\begin{aligned} p(\mathbf{w}|\boldsymbol{\alpha}) &= \prod_{i=1}^d \mathcal{N}(w_i|0, \alpha_i^{-1}), \\ p(\boldsymbol{\alpha}) &= \prod_{i=1}^d \text{Gamma}(\alpha_i|a, b). \end{aligned}$$

Instead of jointly estimating $(\boldsymbol{\alpha}, \beta)$ and $\mathbf{w}$, these hyperparameters are inferred using a maximum a posteriori (MAP) Bayesian inference approach; that is, we select

$$(\boldsymbol{\alpha}_{\text{MP}}, \beta_{\text{MP}}) = \arg \max_{\boldsymbol{\alpha}, \beta} p(\boldsymbol{\alpha}, \beta|\mathbf{t})$$

and calculate $\sigma_{\text{MP}}^2 = \frac{1}{\beta_{\text{MP}}}$. The approach to numerically optimize the parameters $\boldsymbol{\alpha}$ and β is presented in [29].

Because $p(\mathbf{w}|\boldsymbol{\alpha})$ and $p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha})$ are multivariate Gaussian distributions, the posterior distribution of $\mathbf{w}$ is also Gaussian:

$$p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_{\text{MP}}, \sigma_{\text{MP}}^2) = \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with the mean and covariance matrix given by:

$$\boldsymbol{\mu} = \sigma_{\text{MP}}^{-2} \boldsymbol{\Sigma} \mathbf{X}^T \mathbf{t}, \quad \boldsymbol{\Sigma} = \sigma_{\text{MP}}^2 \left(\mathbf{X}^T \mathbf{X} + \mathbf{A} \right)^{-1},$$

where $\mathbf{A} = \text{diag}(\alpha_{\text{MP},1}, \alpha_{\text{MP},2}, \dots, \alpha_{\text{MP},d})$.

3.4. Inference

After the matrices $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ were computed, the model can predict a target t_* given a new test point $\mathbf{x}_*$. The posterior distribution of the predicted target is as follows:

$$p(t_*|\mathbf{t}, \boldsymbol{\alpha}_{\text{MP}}, \sigma_m^2) = \int p(t_*|\mathbf{w}, \boldsymbol{\alpha}_{\text{MP}}, \sigma_{\text{MP}}^2) p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_{\text{MP}}, \sigma_{\text{MP}}^2) d\mathbf{w}.$$

Since both terms in the integral are Gaussian functions, $p(t_*|\mathbf{t}, \boldsymbol{\alpha}_{\text{MP}}, \sigma_m^2)$ is a Gaussian distribution:

$$p(t_*|\mathbf{t}, \boldsymbol{\alpha}_{\text{MP}}, \sigma_{\text{MP}}^2) = \mathcal{N}(t_*|y_*, \sigma_*^2) \quad (3)$$

with the following parameters:

$$y_* = (\sigma_{\text{MP}}^{-2} \boldsymbol{\Sigma} \mathbf{X}^T \mathbf{t})^T \mathbf{x}_*, \quad \sigma_*^2 = \sigma_{\text{MP}}^2 + \mathbf{x}_*^T \boldsymbol{\Sigma} \mathbf{x}_*. \quad (4)$$

Therefore, the hierarchical Bayesian model learns $\boldsymbol{\alpha}_{\text{MP}}, \sigma_{\text{MP}}$ during the training step and use these estimates to calculate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Unlike traditional machine learning models, this Bayesian framework produces as outputs both a central estimate y_* (that is, the predicted mean value) and the variance of the prediction σ_*^2 , given by (4).

The methodology used to generate the dataset and learn the parameters $\mathbf{w}, \boldsymbol{\alpha}, \sigma^2$ from the model, as well as the procedure used for inference, are summarized in Algorithm 1.

Algorithm 1: Methodology used to generate the dataset, train the model, and infer the parameters.

Dataset generation

- 1 Read the values of $r_0, \ell_0, x_0, r_1, \ell_1, x_1$
 - 2 Generate 10,000 samples $p^{(1)}, \dots, p^{(10,000)}$ from truncated normal distribution with mean $\mu^* = p_{0,1}^*$, standard deviation $\sigma^* = 0.1 p_{0,1}^*$ truncated in the interval $[\mu^* - 3\sigma^*, \mu^* + 3\sigma^*]$.
 - 3 **foreach** $s \in \{1, \dots, 10,000\}$ **do**
 - 4 Calculate $Z_c = \sqrt{\frac{L}{C}}$, $P_i = \frac{3V_i^2}{Z_c}$ and $Q_i = P_i \cdot \frac{\sqrt{1-f_p^2}}{f_p}$ from $p^{(s)}$.
 - 5 Solve the following system of non-linear equations using the Gauss–Newton method, obtaining $V_1, \dots, V_n, \theta_1, \dots, \theta_N$:

$$V_i \sum_{k=1}^n V_k (G_{i,k} \cos(\theta_i - \theta_k) + B_{i,k} \sin(\theta_i - \theta_k)) = P_i$$

$$V_i \sum_{k=1}^n V_k (G_{i,k} \sin(\theta_i - \theta_k) - B_{i,k} \cos(\theta_i - \theta_k)) = Q_i$$
 - 6 Calculate $I_k = \frac{P_k - jQ_k^*}{V_k^*}$, $1 \leq k \leq n$
 - 7 Discard $p^{(s)}$ if $v_1 \geq c$ or $v_2 \geq c$ or $V_k < \frac{\sqrt{2}}{2} p_u$
 - 8 Add noise to V_k and I_k according to (5) or (6), $1 \leq k \leq n$
-

Algorithm 1: *Cont.***Training**

9 Find

$$(\alpha_{MP}, \beta_{MP}) = \arg \max_{\alpha, \beta} p(\mathbf{t}|\alpha, \beta)p(\alpha)p(\beta)$$

and define $\sigma_{MP}^2 = \frac{1}{\beta_{MP}}$;

10

11 Calculate the posterior mean and covariance matrix of $\mathbf{w}$:

$$\Sigma = \sigma_{MP}^2 (\mathbf{X}^T \mathbf{X} + \mathbf{A})^{-1}, \quad \mathbf{A} = \text{diag}(\alpha_{MP,1}, \dots, \alpha_{MP,d})$$

$$\mu = \sigma_{MP}^{-2} \Sigma \mathbf{X}^T \mathbf{t}$$

Inference12 Given a vector of features $\mathbf{x}_*$, calculate the posterior mean y_* and variance σ_*^2 of the predicted parameter t_* as

$$y_* = \mu^T \mathbf{x}_*$$

$$\sigma_*^2 = \sigma_{MP}^2 + \mathbf{x}_*^T \Sigma \mathbf{x}_*$$

4. Results

In the initial analysis, we studied a system with a non-ideal three-phase source connected to the transmission line, and at the end, an unbalanced three-phase load. The corresponding system is shown in Figure 3. The load was given as a percentage of the surge impedance loading (SIL) of the transmission line [18]. Using the lossless case, the SIL is given by:

$$SIL_{1\phi} = \frac{|\dot{V}_{A,n}|^2}{\sqrt{\frac{l_{(1)}}{c_{(1)}}}},$$

where $SIL_{1\phi}$ is the equivalent one-phase SIL, or the three-phase SIL divided by three, and $\dot{V}_{A,n}$ is the phase-ground voltage of phase A.

Additionally, to estimate the zero-sequence parameters, an unbalance factor in the load was introduced, designated as f_{ac} . This way, we propose the following manner to generated different active powers for each phase:

$$P_A = (1 + f_{ac})SIL_{1\phi}, \quad P_B = (1 - f_{ac})SIL_{1\phi}, \quad P_C = (1 - 2f_{ac})SIL_{1\phi}.$$

Furthermore, to determine the influence of the reactive power, we fixed a power factor (f_p). The numerical values of the parameters for the studied system are presented in Appendix A.

The presented system was simulated using MATLAB/Simulink using a notebook equipped with an Intel Core i7-1165G7 2.8 GHz CPU and 8GB of memory. In the initial analysis, the dataset size was considered to be 10,000 samples, which were split into training and test sets in an 80/20 proportion.

In terms of practical applications, this approach aims to develop software for post-processing data obtained from transmission lines monitored by PMUs, even when monitoring is limited to one side. The assumption of monitoring a transmission line with at least one PMU is common in the literature [7–9]. However, many papers assume that PMUs are installed at both ends [15], which can increase system costs. In this work, our approach achieved satisfactory performance using measurements from only one side. The proposed methodology for estimating transmission line parameters leverages readily available data

in transmission systems to enhance parameter monitoring, requiring only a computational environment for storing and processing the data used in the model.

Some studies in the literature have utilized practical systems to validate developed methods. Access to data from actual systems is often limited or restricted due to privacy concerns. For example, in [6], the method was tested using an experimental setup based on a simplified system. In contrast, [8] validated their solution for consistency using raw PMU data obtained from a U.S. power utility in the Eastern interconnection.

Using the approach of [6], it is possible to access the method performance directly, since the parameters of a simplified laboratory system can be calculated via offline techniques. Meanwhile, in [8], the consistency of the method can be tested by applying it to similar ambient temperature and loading conditions using different samples, but with the same length. This way, if the method produces the same results, it is possible to conclude that it is consistent.

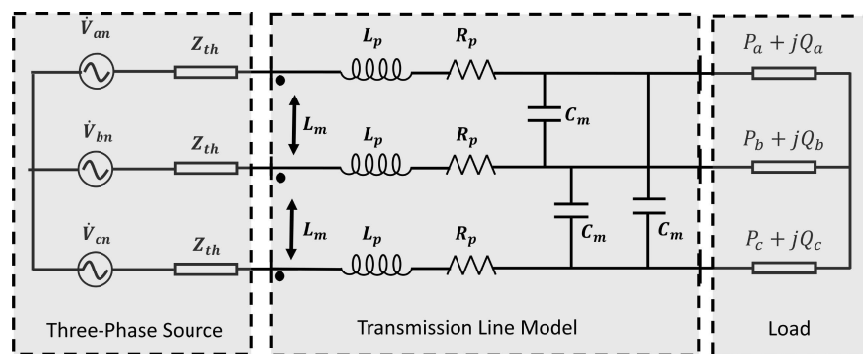


Figure 3. Simplified 2-bus system used to obtain the measurements in the computational environment for different conditions.

Numerical Analysis

The model is built using a training database and evaluated on a test database that is completely excluded from the training process. To present the performance, it is necessary to use some metrics. For this purpose, consider the actual values of the target ($\bar{t} = \{\bar{t}_1, \bar{t}_2, \dots, \bar{t}_k\}$) and the estimated values ($\hat{t} = \{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_k\}$), where k is the size of the test database. In this work, we have adopted the root mean squared error (RMSE) and mean absolute percentage error (MAPE), defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{k} \sum_{j=1}^k (\bar{t}_j - \hat{t}_j)^2}, \quad \text{MAPE} = \frac{100}{k} \sum_{j=1}^k \frac{|\bar{t}_j - \hat{t}_j|}{|\bar{t}_j|}.$$

It is important to note that the RMSE is given in the unit of the parameter, e.g., for resistance, the RMSE is in Ω , whereas the MAPE is given in % using the presented configurations. Additionally, it is relevant to note that the metric is taken considering different k values for the target in the test database. The result for the noiseless case is shown in Table 1. Regarding the computational time required to obtain the solution, we conducted an average time analysis based on running 100 realizations to train and test the algorithm, resulting in an average time of 0.0397 s. As a benchmark, we implemented a simple linear regression solution to that commonly employed in the literature [7,10,31], and we found an average computational time of 0.0213 s. Thus, our method exhibits a computational burden that is very similar to the standard methods used in the literature.

The use of a Bayesian method to estimate the parameters presents another advantage compared to other machine learning-based solutions or even common maximum likelihood estimators. In (3), the developed method for prediction provides the estimator with the distribution of the target t^* in the test database. Consequently, the mean value and the standard deviation of each sample in the test set are calculated. The mean value corresponds to the prediction used for a given set of features. Additionally, the standard deviations can

be used to illustrate the possible variations in the estimation process. To demonstrate this behavior, we have performed the estimation for the parameters R_0 and R_1 , along with their error bars, considering a variation of $\pm 3\sigma$. The obtained results are shown in Figure 4.

Table 1. Performance metrics for the proposed approach considering a 2-bus system and the noiseless case.

Parameter	MAPE (%)	RMSE
R_1	0.0361	0.00586 Ω
L_1	0.0302	0.000127 H
B_1	0.00148	4.09×10^{-11} F
R_0	0.0636	0.271 Ω
L_0	0.0575	0.000947 L
B_0	0.398	6.56×10^{-9} F

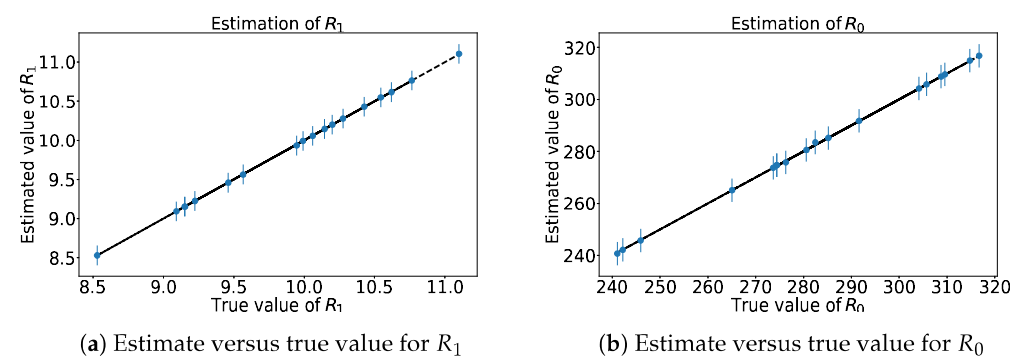


Figure 4. Comparison between the estimated and true parameters for the noiseless case for R_1 (a) and R_0 (b).

The results depicted in Figure 4 reveal some important aspects of the solution. Firstly, the estimation performance is remarkably high, as the true values closely align with the estimated ones. Another noteworthy observation is the reliability of the estimations, evidenced by the small standard deviations. Consequently, the proposed method generates stable predictions with a high performance for estimating three-phase transmission line parameters.

An important influencing factor in a machine learning solution is the size of the dataset, i.e., the number of samples used to build the statistical model and obtain results. Some papers in the literature have employed various dataset sizes. In the context of data-based solutions in power systems, [32] used 5000 samples to train the model, whereas [33] employed 32,000 samples to develop the model used to address the problem. Generally, non-machine learning solutions utilize a day of PMU measurements considering a certain reporting rate for the device. For instance, in [7], a reporting rate of 50 samples per second was utilized, resulting in approximately 4.3×10^6 samples.

The presented work initially employed 10,000 samples to obtain the results shown in Table 1. However, as shown in Figure 5, the required number of samples to achieve the best performance is considerably lower. By analyzing this figure, it is possible to conclude that, for the noiseless case, 200 samples are sufficient to produce results with a high performance.

Another relevant aspect of the solution is the necessity of measurements at both ends. In real-world applications, obtaining PMUs to monitor each transmission line of the system is challenging due to the high cost of implementation [16]. Thus, it is expected that some non-critical lines may only be monitored using a single PMU. Consequently, we tested the proposed algorithm considering measurements from only one side. This process involves removing either the receiver or sender end measurement from the feature matrix $\mathbf{X}$ and applying the method as described in Section 3.2. The results of this analysis are shown in Table 2.

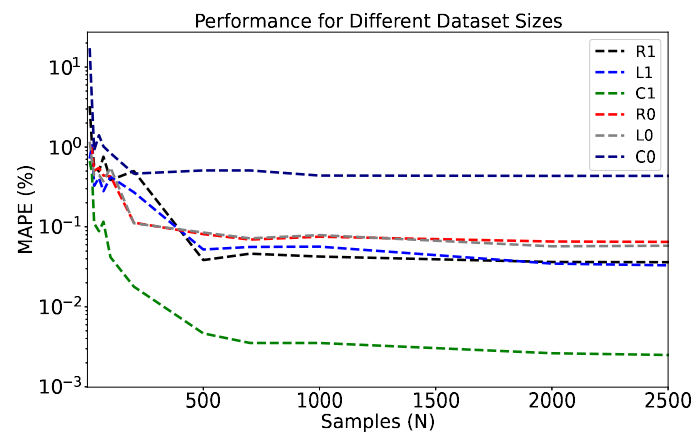


Figure 5. Influence of the number of samples in the estimation method considering the noiseless case.

Table 2. Performance metrics of the proposed approach considering only one-sided measurement in the noiseless case.

Parameter	MAPE (%)	RMSE
R_1	0.324	0.04707Ω
L_1	0.0416	0.000169 H
B_1	0.00415	$1.08 \times 10^{-10} \text{ F}$
R_0	1.55	0.271Ω
L_0	0.200	0.00320 L
B_0	2.67	$3.81 \times 10^{-8} \text{ F}$

The results presented in Table 2 confirm that, as expected, the method's performance decreases when using measurements from only one side. However, the obtained values still demonstrate an acceptable performance [6], as it is desirable to achieve parameter values that are within 10% of the maximum values. It is worth highlighting that the worst-performing parameter is the zero-sequence susceptance. Generally, estimating zero-sequence parameters is more challenging due to their smaller magnitude compared to the positive sequence, which poses an additional barrier to achieving numerical stability in the estimation process.

Another relevant parameter in the estimation process is the active power carried through the line. In the initial analysis, we assumed that the active power was 1 SIL, equivalent to 100% of its value. To consider the influence of this parameter, the active power was varied in a range of $\pm 30\%$ of the 1 SIL. This range is a commonly observed value for the variation in the active power in a period of 24 h [34]. The results of this analysis are shown in Figure 6.

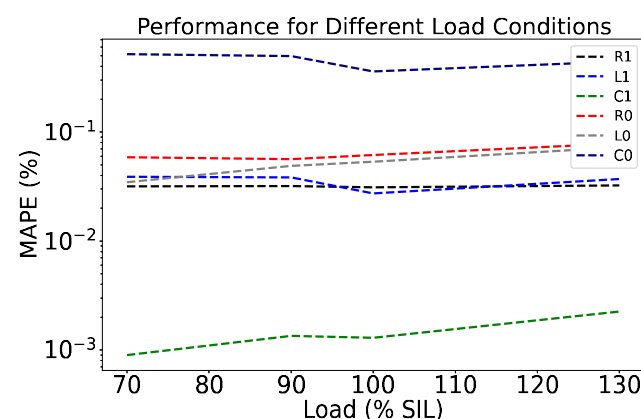


Figure 6. Analysis of the MAPE of the solution under different load conditions and in the noiseless case.

The results presented in Figure 6 show that the proposed methodology is robust concerning the active power carried through the line. This way, the variation in the load did not have a significant influence on the performance of the method, since there was only a slight difference among the results for the studied loads. This is an interesting behavior of the proposed approach due to the fact that common methods in the literature [11], based on nonlinear optimization, present extreme variation in performance when the load is modified, especially for extreme values of load.

Another relevant scenario to test the performance of the algorithm is to estimate the parameters in a complete power system with the presence of other components, such as transformers, generators, loads, and transmission lines. To study the performance of the proposed method under this situation, we implemented the methodology to estimate the transmission line parameters in the IEEE 39-bus system. The IEEE 39-bus standard system is a widely used power network in the New England area of the United States [35], providing an extensive scope to evaluate power system solutions. The system consists of 10 generators, 39 bus bars, and 12 transformers, as illustrated in Figure 7. The methodology employed to generate the data is consistent with that presented in Section 3.1. It is important to note that the simulation was carried out for positive sequence analysis, which is the standard case for these test systems [36]. Thus, the results for several transmission lines of the system are presented in Table 3.

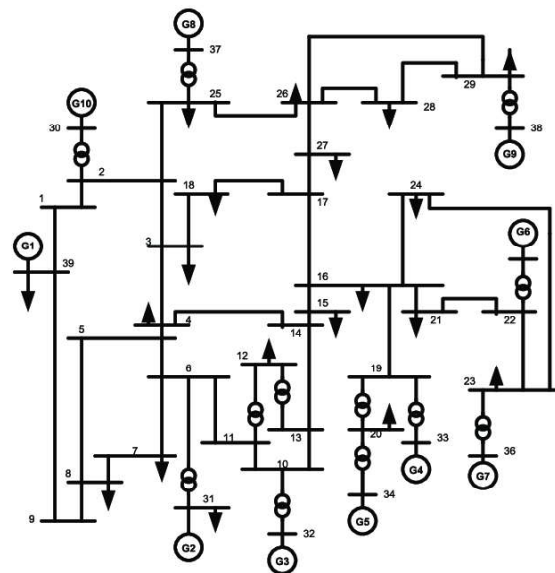


Figure 7. The IEEE 39-bus system [37].

As observed for the simplified 2-bus system, Table 3 demonstrates that the method exhibits a high performance in estimating the positive sequence parameters of a three-phase transmission line. Particularly noteworthy is the fact that the worst parameter yielded a mean absolute percentage error (MAPE) smaller than 1%. Furthermore, it is relevant to emphasize that the parameters with the worst performances shared the characteristic of having one of their buses (sender or receiver) controlled, meaning that the voltage was constant in practical applications. This characteristic poses a challenge for conventional methods, which rely on algebraic solutions and often neglect such restrictions in their solutions. In contrast, the proposed methodology, based on statistical learning, is capable of handling this characteristic simply by ignoring this feature in the regression.

Table 3. Performance metrics of the proposed solution considering the IEEE 39-bus system.

$r_{i,j}$	MAPE (%)	$x_{i,j}$	MAPE (%)	$b_{i,j}$	MAPE (%)
$r_{1,2}$	0.250	$x_{1,2}$	0.0471	$b_{1,2}$	0.00120
$r_{1,39}$	0.998	$x_{1,39}$	0.0161	$b_{1,39}$	0.000344
$r_{2,3}$	0.0719	$x_{2,3}$	0.0235	$b_{2,3}$	0.000827
$r_{2,25}$	0.0308	$x_{2,25}$	0.0201	$b_{2,25}$	0.00457
$r_{3,4}$	0.392	$x_{3,4}$	0.0497	$b_{3,4}$	0.00185
$r_{3,18}$	0.221	$x_{3,18}$	0.0548	$b_{3,18}$	0.000529
$r_{4,5}$	0.0977	$x_{4,5}$	0.0502	$b_{4,5}$	0.00200
$r_{4,14}$	0.118	$x_{4,14}$	0.0583	$b_{4,14}$	0.00243
$r_{5,8}$	0.167	$x_{5,8}$	0.0810	$b_{5,8}$	0.00401
$r_{6,7}$	0.0942	$x_{6,7}$	0.0746	$b_{6,7}$	0.00503
$r_{6,11}$	0.0362	$x_{6,11}$	0.0337	$b_{6,11}$	0.00268
$r_{7,8}$	0.0698	$x_{7,8}$	0.0451	$b_{7,8}$	0.00236
$r_{8,9}$	0.417	$x_{8,9}$	0.0525	$b_{8,9}$	0.00170
$r_{10,11}$	0.0253	$x_{10,11}$	0.0167	$b_{10,11}$	0.00244
$r_{10,13}$	0.0186	$x_{10,13}$	0.0203	$b_{10,13}$	0.00214
$r_{13,14}$	0.0367	$x_{13,14}$	0.0451	$b_{13,14}$	0.00334
$r_{14,15}$	0.519	$x_{14,15}$	0.0974	$b_{14,15}$	0.000309

To directly compare different methods, it is necessary to perform the comparison using scenarios similar to those used to build and test the proposed methodology. In general, papers in the literature develop methods to address specific problems, such as considering the simplified model for the line in the dq domain [38] or in three-phase untransposed short transmission lines [7], which makes it difficult to propose a fair comparison. Additionally, the data models employed are generally different due to the use of other measurement resources, such as power measurements, or even due to the methods adopted to model the inaccuracies present in the data. However, it is possible to use references that present results using noiseless cases and some IEEE test systems. In [31,36], a method was implemented to estimate the parameters based on the extended vector of states in a power system. The comparison of the proposed solution with such references is shown in Table 4. Note that the works have calculated only the admittance branch parameters, which are simply obtained by inverting the elements $(r + jx)^{-1} = g_{sr} + jb_{sr}$.

By observing the results presented in Table 4, it is possible to conclude that the developed methodology achieved better or at least comparable performance with the analyzed previous works in the literature. Furthermore, it is important to highlight that the methods developed in [31,36] may not work in some situations depending on the detection method used to evaluate the parameters. For example, for the branch 9–10, it was not possible to detect such a parameter as erroneous using the approach used in [39].

Table 4. Single line errors in the IEEE 14-bus system.

Line (bus#–bus#)	Proposed Solution (%)		Ref [36] (%)		Ref [31] (%)	
	g_{sr}	b_{sr}	g_{xr}	b_{sr}	g_{sr}	b_{sr}
(1–2)	0.18	0.11	0.35	0.34	0.22	0.17
(2–3)	0.19	0.07	0.25	0.49	0.52	0.66
(2–4)	0.09	0.10	0.07	0.56	0.17	0.09
(2–5)	0.03	0.05	2.00	0.03	0.00	0.01
(6–11)	0.05	0.06	0.44	0.30	0.76	0.01
(6–12)	0.11	0.07	0.37	0.23	0.32	0.06
(6–13)	0.04	0.04	0.34	0.17	0.25	0.03
(9–10)	0.01	0.01	-	-	0.00	0.01
(9–14)	0.05	0.05	0.14	0.18	0.01	0.01
(7–9)	0.00	0.01	0.00	1.66	0.00	0.42

5. Estimation Considering Noise

The noise model for phasor measurements is a fundamental aspect of solutions utilizing data obtained from PMUs. In general, the chain of instruments and components introduces inaccuracies in the data. The instrument transformers (IT), namely voltage and current transformers associated with the PMU device, introduce both random and systematic errors in the temporal series generated by the sequence of obtained phasors.

Some papers in the literature have introduced the necessity of modeling the noise as a separated term for the magnitude and phase; i.e., the noise should be included in the polar coordinates and the distribution should be verified for the projections into the real and imaginary axes [7,9,11]. This model makes the solution more sophisticated, since it is not possible to obtain an analytical solution with a closed form. Instead, approximated numerical solutions are used to evaluate the parameters in different conditions.

In [16], the authors challenge the Gaussian assumption of random errors by arguing that the process of obtaining phasors is multimodal, time-variant, and characterized by non-stationary error statistics. Consequently, the central limit theorem is deemed inapplicable to this scenario, as the chosen sampling distribution may not accurately represent the actual random process underlying the measurement data generation, potentially leading to erroneous performance assessments. To address this challenge, the paper proposed modeling the errors using a Gaussian mixed model (GMM). Using this approach, the authors of [8] developed a solution by equating the statistical model and subsequently applying Newton's method for solving nonlinear problems.

In summary, each different noise model leads to a specific solution, requiring algebraic work to formalize the cost function to optimize and the testing of different algorithms to obtain the parameters. In this work, we propose a solution that is able to deal with different noise models that do not require any algebraic work to equate the transmission line model. By employing a data-centric approach, it is possible to evaluate the parameters with acceptable performance considering several scenarios of the system.

5.1. Gaussian Random Noise with Uniform Systematic Error

Following recent papers in the literature that consider Gaussian models for the random errors [12,13], the field measurements can be simulated by using the following model:

$$\begin{cases} \tilde{V}_h &= (1 + \xi_h^{sys} + \xi_h^{random}) |\dot{V}_h| e^{j(\theta_h + \alpha_h^{sys} + \alpha_h^{random})} \\ \tilde{I}_h &= (1 + \eta_h^{sys} + \eta_h^{random}) |\dot{I}_h| e^{j(\varphi_h + \psi_h^{sys} + \psi_h^{random})} \end{cases} \quad (5)$$

where h represents the index of a specific bus, $\tilde{V}_h$ 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:

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

for $X, Y \in \{\xi_h^{sys}, \xi_h^{random}, \eta_h^{sys}, \eta_h^{random}, \alpha_h^{sys}, \alpha_h^{random}, \psi_h^{sys}, \psi_h^{random}\}$, where $\mathbb{E}[X]$ is the expected value of the random variable X .

This model relies on papers that deal with actual data from PMUs, and its design is based on the analyzed statistics of time series obtained from field applications. According to previous studies [40–42], a Gaussian distribution is often a meaningful choice, depending on the specific device, measured quantity, and considered channel. It is also important to highlight that the studies were conducted by fitting one model for the magnitude data and another for the phase data, as adopted in (5).

In terms of noise level, although the values vary significantly across the diverse data, a value of 0.5% is a good approximation for magnitude [40], and 0.52 degrees is a good approximation for the phase [13]. In this paper, we have used more conservative values for the noise: 1% for the magnitude noise level and 1.03 degrees for the phase, demonstrating that the developed methodology does not assume the availability of very high-quality data.

Concerning the values, the proposed approach considers that [13]:

- Instrument transformers, including both voltage and current transformers, are categorized as class 1.0 [11]. Systematic errors are modeled to follow a uniform distribution with a standard deviation equal to $\Delta/\sqrt{3}$, where Δ represents the maximum deviation. For example, in class 1.0, a systematic error of $\Delta = 1.0\%$ corresponds to the systematic error model.
- The random errors are assumed to be Gaussian based on the maximum total vector error (TVE) of 1% for the PMUs [13], which is decomposed into a maximum amplitude error of 1% and a maximum phase angle error of 0.57° .

It is worth highlighting that to evaluate the performance of the model under random noise, 1000 Monte Carlo simulations were carried out. This approach helps mitigate the bias introduced by a single numerically generated distribution and yields a more stable response that can be compared in a fairer manner with the proposed solutions in the literature.

When applying the presented model using the literature values, the results for the Gaussian random noise with uniform systematic errors are displayed in Table 5.

Table 5. Performance metrics of the estimation for the 2-bus system considering Gaussian noise and systematic error in the measurements.

Parameter	MAPE (%)	RMSE
R_1	1.50	0.182 Ω
L_1	0.228	0.000879 H
B_1	0.0349	8.53×10^{-10} F
R_0	1.26	4.39 Ω
L_0	0.838	0.0117 L
B_0	3.15	4.71×10^{-8} F

Similar to the noiseless case, it is possible to generate results with their corresponding error bars for $\pm 3\sigma$. The obtained result is shown in Figure 8. Upon analyzing this figure, it is evident that the performance deteriorates when considering noise compared to the noiseless case, as expected since noise introduces complications in the parameter estimation problem. It is worth highlighting that the presented approach not only provides the estimated parameters but also presents the bounds for the expected error in a generic realization of the method. This represents an innovation in solutions dealing with transmission line parameter estimation.

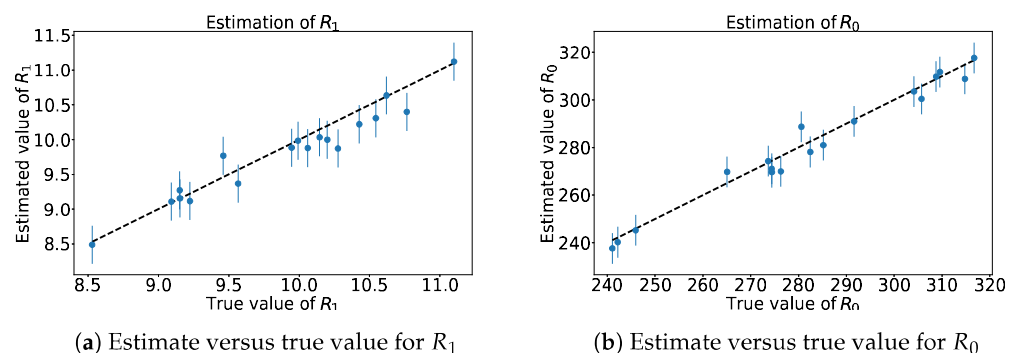


Figure 8. Comparison between the estimated and true parameters considering Gaussian noise and systematic errors for R_1 (a) and R_0 (b).

In order to demonstrate the robustness of the proposed solution under Gaussian noise compared with other works in the literature, the model presented in [9] was implemented. In this model, the authors developed a recursive least squares-based solution to estimate the transmission line parameters. In their paper, they used the IEEE 118-bus test system and applied Gaussian noise to phasor measurements using the errors in the instrument transformers shown in Table 6. To establish a fair comparison, we implemented the same system using the same noise model. It is important to emphasize that the authors of [9] used zero-mean Gaussian noise, neglecting the presence of systematic errors. Therefore, we adopted the same strategy to make a fair comparison of the methods. The results are shown in Table 7.

Table 6. Standard deviations of measurement errors [9].

Error	Voltage		Current	
	Magnitude	Phase	Magnitude	Phase
Assumed	0.3%	0.4°	0.4%	0.7°

Table 7. Errors in the IEEE 118-bus system.

Line (bus#–bus#)	Proposed Solution (%)			Ref [9] (%)		
	<i>r</i>	<i>x</i>	<i>b</i>	<i>r</i>	<i>x</i>	<i>b</i>
(8–9)	0.0245	0.717	0.0334	1.64	1.38	0.0403
(9–10)	0.0159	0.216	0.252	2.29	0.140	0.291
(23–25)	0.396	0.311	0.497	0.774	0.681	1.77
(25–27)	0.441	0.373	0.342	0.142	0.0837	0.0871
(26–30)	0.242	0.116	0.0776	0.124	0.146	0.198
(42–49)	0.486	0.435	0.232	1.50	0.563	0.268
(49–54)	0.814	0.574	0.0755	1.05	0.867	0.0738
(54–59)	0.586	0.566	0.0562	1.25	1.61	0.0750
(56–59)	0.241	0.600	0.0521	2.15	1.06	0.103
(59–61)	0.701	0.404	0.103	1.29	1.31	0.605
(38–65)	0.0448	0.246	0.0569	0.0445	0.284	0.174
(49–66)	0.272	0.473	0.0928	1.52	0.964	0.182
(62–66)	1.31	0.428	0.0639	1.37	1.64	0.315
(49–69)	0.769	0.284	0.0464	0.462	0.499	0.564
(69–70)	0.344	0.586	0.0932	0.280	0.665	0.155
(76–77)	0.338	0.319	0.0852	0.221	1.56	0.194
(89–90)	0.370	0.328	0.192	2.05	0.688	0.814
(89–92)	0.700	0.381	0.119	1.72	0.616	0.046
(92–100)	0.998	0.136	0.0279	1.54	1.31	0.444
(100–106)	0.222	0.462	0.249	0.164	0.729	1.20

The results presented in Table 7 demonstrate that the developed methodology is superior to the one in the literature for most of the analyzed lines of the IEEE 118-bus system. Some aspects are relevant in the comparison of the solutions. The largest observed deviations are in the resistance parameter, and the smallest deviations are in the susceptance parameter, which occurred in both solutions. Additionally, the highest encountered deviation in [9] was greater than 2%, while in the presented solution, it was less than 1.5%.

5.2. Non-Gaussian Random Noise

Recent papers in the literature have assessed the assumption that random errors in phasor measurements obtained from PMUs follow a Gaussian distribution. The authors of [16,17] used field data to demonstrate that phasor measurements, particularly for current measurements, do not conform to a Gaussian distribution [16]. In place of the Gaussian model, the authors in [16] propose a model based on GMMs for these measurements.

Using this model, let $\mathbf{x}_n \in \mathbb{R}^d$ be a sample of the phasor measurements. The probability density function $p(\cdot)$ of this sample is given by:

$$p(\mathbf{x}_n) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

where K is the number of clusters or Gaussian distributions and π_k is the mixing coefficients of the normal distributions, such that:

$$\sum_{k=1}^K \pi_k = 1.$$

Since $\mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the multivariate normal distribution, it is possible to write the following:

$$\mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}_k|} \cdot (2\pi)^d} \exp\left(-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k)\right).$$

An important parameter for modeling the Gaussian mixed model is the size of the window used to fit the centroids (average values) and the format of the distribution (covariance matrix). This parameter will be referred to as N . Thus, for N independent and identically distributed samples, the random vector $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ is obtained using the following:

$$p(\mathbf{X}) = \prod_{n=1}^N \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \quad (6)$$

To accurately describe the noise model for the data, it is necessary to provide the number of clusters, the mean values for each measurement, their corresponding variances, and the vector of weights. In the model, it is assumed that each random variable is uncorrelated with the others; hence, the covariance matrix is diagonal.

In possession of real field data, these parameters could be estimated using a certain method. Initially, determining the optimal number of clusters is accomplished by employing the Bayesian information criterion (BIC). This approach aids in estimating the model that best fits the data while penalizing models with a high complexity. In the literature [8,16], the reported optimal number of clusters typically falls between 2 and 4. The means and variances can then be evaluated based on the maximization of the likelihood using the model outlined in Equation (6).

This way, to prove the robustness of the proposed methodology, we have investigated the performance of the method under two different conditions of Non-Gaussian noise for 2 and 4 clusters.

Utilizing the methodology elucidated in [8], the efficacy of the approach was explored for two clusters (Scenario 1) and four clusters (Scenario 2).

The subsequent parameters were employed for Scenarios 1 and 2:

$$\boldsymbol{\mu}_k = \begin{bmatrix} 0 \\ 0.005 \end{bmatrix}, \quad \boldsymbol{\Sigma}_k = 0.0015 \cdot \mathbf{I}_2, \quad \pi_k = \{0.3; 0.7\} \quad (\text{Scenario 1})$$

$$\boldsymbol{\mu}_k = \begin{bmatrix} -0.002 \\ 0 \\ 0.005 \\ 0.008 \end{bmatrix}, \quad \boldsymbol{\Sigma}_k = 0.001 \cdot \mathbf{I}_4, \quad \pi_k = \{0.1; 0.2; 0.5; 0.2\}, \quad (\text{Scenario 2})$$

where I_n is the identity matrix of dimension n .

The corresponding results for both situations are presented in Table 8. By observing the results for the association between Gaussian and systematic noise and non-Gaussian noise, it is possible to conclude that, under the standard conditions of noise levels in both cases, the Gaussian noise associated with a uniform distribution is more difficult for the model to deal with. Generally speaking, this occurred because the former model introduced a greater bias value into the data, leading to a loss of accuracy in the method. However, in both cases, the proposed methodology was able to estimate the six parameters with acceptable performance, proving its robustness under different noise models.

According to [16], the noise level in GMM might vary in terms of the mean values and the variance of each Gaussian distribution. This way, following the approach presented in [8], it is possible to consider different values for the mean and variance by scaling using a factor expressed as a percentage. The magnitude of 1% noise closely resembles the GMM noise attributes employed in the previous study. For the remaining values, the GMM noise parameters were derived by appropriately adjusting the mean and standard deviations of the noise. For instance, to derive the GMM noise attributes for a 2% noise level, the GMM noise characteristics associated with a 1% noise level were doubled. The result is shown in Figure 9. This study proves that the method is robust under different noise levels. Particularly, for all the studied noise levels, the observed MAPE was less than 1%.

Table 8. Performance metrics the proposed method considering non-Gaussian noise in a 2-bus system.

Parameter	Scenario 1 MAPE (%)	Scenario 2 MAPE (%)
R_1	0.223	0.221
L_1	0.115	0.107
C_1	0.00215	0.00228
R_0	0.153	0.138
L_0	0.162	0.149
C_0	0.402	0.431

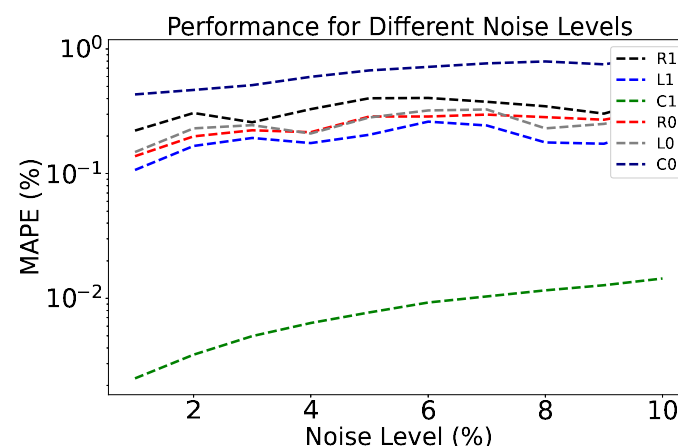


Figure 9. Analysis of the MAPE of the solution under different noise levels using Gaussian mixed models.

6. Conclusions

This paper introduces a new Bayesian regression approach to estimate transmission line parameters. The proposed methodology is model-free, meaning that the estimation method does not rely on the transmission line equations to estimate the parameters. Instead, the developed method is data-centric, allowing for the evaluation of parameters even with measurements from only one side. Consequently, the method does not necessitate PMUs at both ends of the transmission lines, making it more realistic for real-world applications.

Furthermore, due to the nature of the proposed method, it was possible to obtain credible intervals for the solutions, providing bounds for the estimation of each parameter.

This aspect is often overlooked in solutions dealing with transmission line parameter estimation. The performance of the solution was investigated under different noise models, considering the state-of-the-art noise modeling in phasor measurement units. In particular, it was concluded that the standard case of Gaussian random noise associated with systematic errors is more challenging to estimate compared to the model that employs GMMs. However, for all the studied models, our method achieved acceptable performance, outperforming solutions from the literature under the same studied conditions.

Finally, our methodology was tested across systems of varying scales, including several IEEE test systems of different sizes. This testing encompassed various load conditions, numbers of samples (size of the database), and noise levels. Therefore, the proposed solution remained stable under different conditions of the power system and data models.

Author Contributions: Conceptualization, F.P.d.A., R.N. and C.A.P.J.; methodology, F.P.d.A., R.N. and E.C.M.d.C.; software, F.P.d.A., R.N., E.C.M.d.C. and C.A.P.J.; validation, F.P.d.A., R.N. and C.A.P.J.; formal analysis, F.P.d.A., R.N. and E.C.M.d.C.; investigation, F.P.d.A., R.N. and C.A.P.J.; resources, F.P.d.A., R.N. and E.C.M.d.C.; data curation, F.P.d.A., R.N., C.A.P.J. and E.C.M.d.C.; writing—original draft preparation, F.P.d.A., R.N., C.A.P.J. and E.C.M.d.C.; writing—review and editing, F.P.d.A., R.N., C.A.P.J. and E.C.M.d.C.; visualization, F.P.d.A. and R.N.; supervision, R.N. and E.C.M.d.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed by São Paulo Research Foundation (FAPESP)—Grant 2021/01325-7, 2024/00261-3 and 2019/21858-0.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Appendix A

Parameters of the System

Table A1. Transmission system parameters used to simulate the measurements in a computational environment.

Parameter	Value
$\dot{V}_{A,n}; \dot{V}_{B,n}; \dot{V}_{C,n}$	132.8 kV
$r_{(1)}$	0.052986 Ω/km
$l_{(1)}$	1.2819 mH/km
$c_{(1)}$	9.2876 nF/km
$r_{(0)}$	1.4559 Ω/km
$l_{(0)}$	5.546 mH/km
$c_{(0)}$	5.4995 nF/km
Rated frequency (f)	60 Hz
Length (ℓ_{sec})	200 km
Power factor	f_p
Unbalance factor	f_{ac}
$SIL_{1\phi}$	46.67 MW
P_A	$(1 + f_{ac}) \cdot SIL_{1\phi}$
P_B	$(1 - f_{ac}) \cdot SIL_{1\phi}$
P_C	$(1 - 2f_{ac}) \cdot SIL_{1\phi}$
Q_A	$P_A \cdot \sqrt{1 - f_p^2} / f_p$
Q_B	$P_B \cdot \sqrt{1 - f_p^2} / f_p$
Q_C	$P_C \cdot \sqrt{1 - f_p^2} / f_p$

References

- Patel, B.; Bera, P. A New Transmission Line Parameter Estimation Technique and Its Impact on Fault Localization. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 9003908.
- Costa, E.C.M.; Kurokawa, S. Estimation of transmission line parameters using multiple methods. *IET Gener. Transm. Distrib.* **2015**, *9*, 2617–2624. [\[CrossRef\]](#)
- de Albuquerque, F.P.; Nascimento, R.; Liboni, L.H.; Pereira, R.F.R.; da Costa, E.C.M. Data-centric approach for online P-margin estimation from noisy phasor measurements. *Energy Rep.* **2023**, *10*, 2194–2205. [\[CrossRef\]](#)
- Kusic, G.; Garrison, D. Measurement of transmission line parameters from SCADA data. In Proceedings of the IEEE PES Power Systems Conference and Exposition, New York, NY, USA, 10–13 October 2004; pp. 440–445.
- Mansani, P.K.; Pal, A.; Rhodes, M.; Keel, B. Estimation of transmission line sequence impedances using real PMU data. In Proceedings of the 2018 North American Power Symposium (NAPS), Fargo, ND, USA, 9–11 September 2018; pp. 1–6.
- Milojević, V.; Čalić, S.; Rietveld, G.; Ačanski, M.V.; Colangelo, D. Utilization of PMU measurements for three-phase line parameter estimation in power systems. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2453–2462. [\[CrossRef\]](#)
- 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*, 6143–6154. [\[CrossRef\]](#)
- Varghese, A.C.; Pal, A.; Dasarathy, G. Transmission line parameter estimation under non-Gaussian measurement noise. *IEEE Trans. Power Syst.* **2022**, *38*, 3147–3162. [\[CrossRef\]](#)
- Lin, J.; Song, J.; Lu, C. Synchrophasor data analytics: Transmission line parameters online estimation for energy management. *IEEE Trans. Eng. Manag.* **2019**, *69*, 671–681. [\[CrossRef\]](#)
- Dobakhshari, A.S.; Abdolmaleki, M.; Terzija, V.; Azizi, S. Online non-iterative estimation of transmission line and transformer parameters by SCADA data. *IEEE Trans. Power Syst.* **2020**, *36*, 2632–2641. [\[CrossRef\]](#)
- De Albuquerque, F.P.; Da Costa, E.C.M.; Pereira, R.F.R.; Liboni, L.H.B.; De Oliveira, M.C. Nonlinear analysis on transmission line parameters estimation from noisy phasorial measurements. *IEEE Access* **2021**, *10*, 1720–1730. [\[CrossRef\]](#)
- Pegoraro, P.A.; Sitzia, C.; Solinas, A.V.; Sulis, S. Transmission line parameters estimation in the presence of realistic PMU measurement error models. *Measurement* **2023**, *218*, 113175. [\[CrossRef\]](#)
- Sitzia, C.; Muscas, C.; Pegoraro, P.A.; Solinas, A.V.; 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. [\[CrossRef\]](#)
- Moghanian, M.; Dobakhshari, A.S. Accurate Kalman filter based estimation of transmission line parameters utilizing synchronized phasor measurements. *Electr. Power Syst. Res.* **2024**, *230*, 110218. [\[CrossRef\]](#)
- Khalili, R.; Abur, A. Transmission line parameter error identification and estimation in three-phase networks. *IEEE Trans. Power Syst.* **2021**, *37*, 2271–2282. [\[CrossRef\]](#)
- Ahmed, M.M.; Amjad, M.; Qureshi, M.A.; Imran, K.; Haider, Z.M.; Khan, M.O. A critical review of state-of-the-art optimal pmu placement techniques. *Energies* **2022**, *15*, 2125. [\[CrossRef\]](#)
- Wang, S.; Zhao, J.; Huang, Z.; Diao, R. Assessing Gaussian assumption of PMU measurement error using field data. *IEEE Trans. Power Deliv.* **2017**, *33*, 3233–3236. [\[CrossRef\]](#)
- Kundur, P. Power system stability. *Power Syst. Stab. Control* **2007**, *10*, 7-1.
- Paul, C.R. *Analysis of Multiconductor Transmission Lines*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
- Givaki, K.; Seyedzadeh, S. Machine learning based impedance estimation in power system. In Proceedings of the 8th Renewable Power Generation Conference (RPG 2019), Shanghai, China, 24–25 October 2019.
- Qin, C.; Vyakaranam, B.; Etingov, P.; Venetos, M.; Backhaus, S. Machine learning based network parameter estimation using AMI data. In Proceedings of the 2022 IEEE Power & Energy Society General Meeting (PESGM), Denver, CO, USA, 17–21 July 2022; pp. 1–5.
- 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*, 1299–1305. [\[CrossRef\]](#)
- Grafarend, E.W. *Linear and Nonlinear Models: Fixed Effects, Random Effects, and Mixed Models*; de Gruyter: Berlin, Germany, 2006.
- 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. [\[CrossRef\]](#)
- Sun, L.; Chen, T.; Chen, X.; Ho, W.K.; Ling, K.V.; Tseng, K.J.; Amaratunga, G.A. Optimum placement of phasor measurement units in power systems. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 421–429. [\[CrossRef\]](#)
- Monti, A.; Muscas, C.; Ponci, F. *Phasor Measurement Units and Wide Area Monitoring Systems*; Academic Press: Cambridge, MA, USA, 2016.
- Johnson, T.; Moger, T. A critical review of methods for optimal placement of phasor measurement units. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e12698. [\[CrossRef\]](#)
- Borchani, H.; Varando, G.; Bielza, C.; Larranaga, P. A survey on multi-output regression. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2015**, *5*, 216–233. [\[CrossRef\]](#)
- Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
- Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
- Costa, L.F.; Giraldo, J.S.; Castro, C.A. Identification and correction of transmission line parameter errors using SCADA and synchrophasor measurements. *Int. J. Electr. Power Energy Syst.* **2022**, *135*, 107509. [\[CrossRef\]](#)

32. Kumar, S.; Tyagi, B.; Kumar, V.; Chohan, S. PMU-based voltage stability measurement under contingency using ANN. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 1–11. [[CrossRef](#)]
33. Gao, H.; Wang, L.; Yang, D.; Cai, G.; Liu, C.; Yang, H.; Sun, Z.; Wang, B. Mechanism Enhanced Data-Driven Method for Reliability Improvement of Load Margin Estimation. *IEEE Trans. Power Syst.* **2023**, *39*, 3715–3724. [[CrossRef](#)]
34. Zhou, D.Q.; Annakkage, U.D.; Rajapakse, A.D. Online monitoring of voltage stability margin using an artificial neural network. *IEEE Trans. Power Syst.* **2010**, *25*, 1566–1574. [[CrossRef](#)]
35. Ghaedi, A.; Golshan, M.E.H.; Sanaye-Pasand, M. Transmission line fault location based on three-phase state estimation framework considering measurement chain error model. *Electr. Power Syst. Res.* **2020**, *178*, 106048. [[CrossRef](#)]
36. Asprou, M.; Kyriakides, E. Identification and estimation of erroneous transmission line parameters using PMU measurements. *IEEE Trans. Power Deliv.* **2017**, *32*, 2510–2519. [[CrossRef](#)]
37. Sahoo, B.; Samantaray, S.R.; Kamwa, I. Supervising Vulnerable Third Zone Distance Relay to Enhance Wide-Area Back-Up Protection Systems. *IEEE Access* **2022**, *10*, 49862–49872. [[CrossRef](#)]
38. Pereira, R.F.R.; de Albuquerque, F.P.; Liboni, L.H.B.; Costa, E.C.M.; de Oliveira, M.C. Impedance Parameters Estimation of Transmission Lines by an Extended Kalman Filter-Based Algorithm. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–10. [[CrossRef](#)]
39. Asprou, M.; Kyriakides, E. Estimation of transmission line parameters using PMU measurements. In Proceedings of the 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 26–30 July 2015; pp. 1–5.
40. Brown, M.; Biswal, M.; Brahma, S.; Ranade, S.J.; Cao, H. Characterizing and quantifying noise in PMU data. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 17–21 July 2016; pp. 1–5.
41. Castello, P.; Muscas, C.; Pegoraro, P.A. Statistical behavior of PMU measurement errors: An experimental characterization. *IEEE Open J. Instrum. Meas.* **2022**, *1*, 1–9. [[CrossRef](#)]
42. Frigo, G.; Derviškić, A.; Bach, A.; Paolone, M. Statistical model of measurement noise in real-world PMU-based acquisitions. In Proceedings of the 2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA), College Station, TX, USA, 21–23 May 2019; pp. 1–8.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.