

Equivalent wide bandwidth transformer modeling using physics-consistent error compensation model

Qianchao Wang^a, Chakhung Yeung^a, Quan Zhou^a, Yuxuan Ding^{a,*}, Yaping Du^a, Lei Jia^b, Song Zhang^c

^a Department of Building Environment and Energy Engineering, Hong Kong Polytechnic University, 999077, Hong Kong, China

^b The Power Research Institute, China Southern Power Grid, Guangzhou, 510000, China

^c The Guizhou Electric Power Research Institute, China Southern Power Grid, Guiyang, 550000, China

ARTICLE INFO

Keywords:

Wide bandwidth transformer
Deep learning
Physics-consistent model
Power system

ABSTRACT

Wide bandwidth transformer modeling has been a critical part in the power system. However, achieving a balance between modeling accuracy and computational efficiency remains a significant challenge. To address this, the present study introduces a novel physics-consistent error compensation model (PCEC), which integrates a data reconstruction module, an adaptive parameter generation module, and an error compensation module. First, the data reconstruction module effectively denoises the input signals. The processed data are then utilized to generate RLC model parameters, incorporating time-varying physical properties into the PCEC framework. Meanwhile, the same data are re-parameterized and used as input to the error compensation module, which corrects deviations in the RLC model. To ensure the reliability of PCEC, a field experiment is conducted to gather essential voltage data. The effectiveness of the proposed model is validated through three analytical experiments, while the importance of individual modules is further demonstrated through two ablation studies.

1. Introduction

The transient performance of power distribution networks plays a critical role in ensuring the reliability and stability of modern power systems, especially with the increasing integration of renewable energy sources, distributed generation, and advanced power electronics. Transient events, such as lightning strikes and switching operations, can induce high-frequency disturbances that pose significant risks to equipment and system operation [1]. Effective protection against such events requires accurate simulation and optimization of the network's transient response [2]. Transformers, as critical core components of the power grid, play a pivotal role in shaping the system's behavior during such disturbances. Accurate modeling of transformers across a wide frequency range is therefore essential for capturing the complex dynamics of transient phenomena and enabling the design of robust protection and mitigation strategies.

In this vein, some researchers proposed a lot modeling methods for transformers. The most popular and simplest methods are the multi-conductor transmission line (MTL) model [3] and the RLC model [4], each of which is valid within a certain frequency range. Nevertheless,

some inherent problems, such as the difficulty in modeling the differences in winding turns length, being computationally intensive and time-consuming limit the application of MTL [5]. On the other hand, RLC models have been used in the analysis and simulation of transformers with large-scale winding and a large number of turns/disks, thanks to their lower complexity and lower computational requirements. In [6], a detailed transformer high frequency electric circuit model is proposed considering the winding structure, inter-turn capacitance and all mutual inductance. By calculating the model's parameters and comparing with other methods, the accuracy of the proposed model is assessed. [7] proposed an improved RLC model by analyzing and extracting the circular multi-conductor transmission line (CMTL) [8] equations. The accuracy of the new RLC model is proved theoretically and practically. A winding-core high-frequency equivalent circuit model is proposed in [9] to improve the accuracy of simulated signals. These researches try to add some electronic components based on physical mechanisms of transformers to achieve a trade-off between computational complexity and model accuracy. However, these methods are computationally intensive and cost a lot, especially when the model is infinitely subdivided into equivalent capacitors, resistors, and

* Corresponding author.

E-mail addresses: qianchao.wang@polyu.edu.hk (Q. Wang), chakhung.yeung@connect.polyu.hk (C. Yeung), q3zhou@polyu.edu.hk (Q. Zhou), yx.ding@connect.polyu.hk (Y. Ding), Ya-ping.du@polyu.edu.hk (Y. Du), jialei@csg.cn (L. Jia), 1151077158@qq.com (S. Zhang).

<https://doi.org/10.1016/j.ijepes.2025.110817>

Received 14 February 2025; Received in revised form 18 April 2025; Accepted 4 June 2025

Available online 10 July 2025

0142-0615/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

inductors. Therefore, it is necessary to find a way to effectively balance computational complexity and physical consistency for equivalent wide bandwidth transformer modeling. A potential way is to use the RLC model as the core component and to use the efficient computing power of artificial intelligence (AI) to compensate for the deviation of the RLC model. The details of transformers can be fully compensated for by AI models as long as the basic mechanism remains unchanged.

In recent research, AI has been used mainly for transformer fault diagnosis and frequency response analysis (FRA) instead of regression models [10]. Several machine learning methods are tested in partial discharge (PD) localization of transformers, including neural networks, support vector machine (SVM), and k-nearest neighbor [11]. They are analyzed for near real-time identification of high-risk PDs. The convolutional neural network (CNN) is widely utilized in transformer fault diagnosis [12,13] using different input signals. [14] uses CNN to interpret the frequency responses of power transformer windings faults, improving detection accuracy and speed. The balanced isolation forest is combined with CNN in [15] to suppress the level of noise of the vibration signals for higher fault diagnostic accuracy. Some other methods such as residual networks [16], state coding [17] are also combined with deep learning methods for fault diagnosis. Although AI is widely used for transformer fault diagnosis at this stage, establishing a physically consistent deep learning regression model for transformer is still a gap.

Another problem is the parameter calculation of RLC models. For most research, the parameters are determined by the inherent properties of transformers and some assumptions. Finite element simulations can be employed to calculate frequency-dependent circuit parameters via the magneto-static and electrostatic equations [7,9]. Optimization algorithms can be used for parameters searching of RLC as well such as particle swarm optimization [18], shark smell optimization [19], and the enhanced logistic chaotic marine predator algorithm [20]. Neural networks are also utilized as the parameter searchers based on FRA for different transformers with different ratings, sizes, and winding structures [21,22]. Nevertheless, these calculated parameters are deterministic and can only reflect the average of the properties instead of the real-time features of the transformers.

In this context, the primary aim of this research is to develop a deep learning regression model that effectively incorporates the underlying physical principles of transformers, focusing on error compensation and parameter optimization to improve the transformer model accuracy in high frequencies. Building on the foundation of the RLC model, we present a novel physics-consistent error compensation model, designated as PCEC. This model integrates three essential components: a data reconstruction module, an error compensation module, and an adaptive parameter generation module. By employing variational loss and reparameterization techniques [23] in the data reconstruction module, we successfully decompose the raw sampled data into distinct input signals and Gaussian noise. The input signals are subsequently utilized as inputs for the RLC models and the adaptive parameter generation module to ascertain the outputs and the corresponding parameters for the RLC system. Following this, the reparameterized signals from the reconstruction module are leveraged to minimize the discrepancy between the observed outputs and the simulated outputs of the RLC models. The loss functions embedded within PCEC are designed to drive the compensated error toward zero, ensuring that the outputs of the RLC models align closely with the actual data. Crucially, this innovative architecture possesses the potential for adaptation across various domains within regression modeling. To empirically demonstrate the efficacy of PCEC, we conducted experiments utilizing real transformers and gathered the necessary data for modeling. Comparative analyses were performed involving three RLC-based models and four machine learning models through rigorous analytical experiments. Furthermore, ablation studies were executed to investigate the contributions of the data reconstruction module and the adaptive parameter generation module. In summary, the key contributions of this paper encompass:

1. For RLC-based transformer modeling, a novel physics-consistent error compensation model named PCEC is proposed, using deep learning methods, to improve the modeling accuracy.
2. PCEC integrates three primary components: a data reconstruction module, an error compensation module, and an adaptive parameter generation module. These modules effectively denoise the raw data, produce real-time parameters for the RLC model that accurately reflect the inherent characteristics of transformers, and compensate for errors in the RLC model.
3. A field-based transformer experiment is conducted to collect empirical data, which is used to validate the effectiveness of the PCEC framework.
4. The analytical and ablation experiments demonstrate the efficiency of PCEC and the importance of the proposed modules, especially compared with RLC-based models and machine learning models.

The paper is organized as follows. Section 2 provides background about equivalent wide bandwidth RLC model of transformer and variational auto encoder. Section 3 describes the detailed information of the proposed PCEC, including the data reconstruction module, the error compensation module, and the adaptive parameter generation module. Section 4 shows the transformer experiments and other experimental results, and Section 5 concludes the paper.

2. Preliminaries

2.1. Equivalent wide bandwidth RLC model of transformer

The frequency range of lightning-induced over-voltage is relatively broad, typically spanning from 1 Hz to 1 MHz. Hence, an accurate equivalent RLC model needs to depict the characteristics of transformers in this situation. However, achieving higher model accuracy inherently increases the computational complexity, posing significant challenges for practical application. In this section, we introduce an RLC model which has the same structure as our real transformer to simulate the transformer's characteristics when struck by lightning [24].

Fig. 1 shows the structure of equivalent wide bandwidth RLC model. The left side of the diagram depicts the high-voltage winding, while the right side illustrates the low-voltage winding. Points A, B, C denote the three-phase taps on the high-voltage side, whereas points a, b, c, and o represent the corresponding three-phase taps and the neutral point on the low-voltage side. Both ends of the windings are connected in a star configuration. In this model, capacitance is employed to represent capacitive coupling in the high-frequency range, inductance is utilized to signify electromagnetic coupling in the low-frequency range, and resistance accounts for energy losses.

Specifically, Chg and Clg in the figure indicate the capacitance between the windings and the transformer shell. Ch and Cl represent the total equivalent capacitance between layers as well as inter-turn capacitance. $C1 = C2 = C3 = C4$ denotes the inter-phase capacitance of the high-voltage winding. Chl illustrates inter-winding capacitance. $R1$ and $R2$ indicate leakage resistance. $L1$ and $L2$ represent leakage inductance. Rm and Lm correspond to excitation resistance and excitation inductance, respectively. The ideal ratio is represented by k . The calculation of the parameters are based on the transformer nameplate and the inter structure parameters [25].

2.2. Variational auto-encoder

Variational auto-encoder (VAE) is a probabilistic model created by variational inference.

Assuming that the input dataset is denoted as $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^N$ and the output is denoted as $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}_i\}_{i=1}^N$. The implicit variable is denoted as $\mathbf{z} = \{\mathbf{z}_i\}_{i=1}^N$. The encoder is used to approach the real posterior

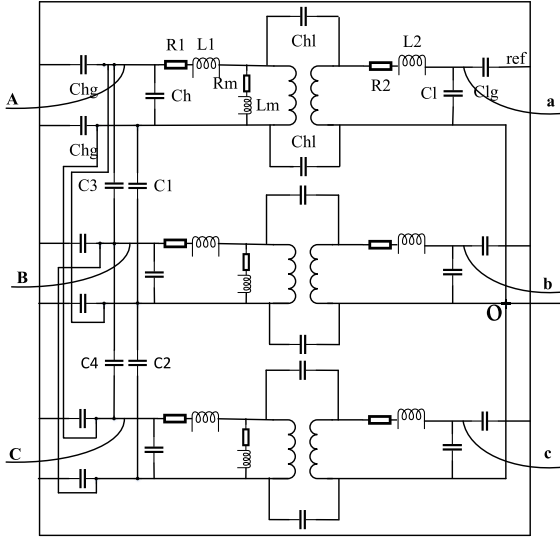


Fig. 1. Equivalent wide bandwidth RLC model [24].

distribution $p_\theta(z|x)$ by estimated $q_\phi(z|x)$. Assuming that $p_\theta(z|x)$ follows the standard normal distribution $N(0, 1)$ and can be written as

$$p_\theta(z|x) = \frac{p_\theta(x|z)p_\theta(z)}{p_\theta(x)} \quad (1)$$

where $p_\theta(z)$ is the prior distribution following the standard normal distribution. $p_\theta(x|z)$ is the probabilistic decoder. To evaluate the difference between real and estimated posterior distributions, the Kullback-Leibler (KL) divergence is utilized in loss function.

$$\begin{aligned} KL(N(\mu, \sigma^2) \parallel N(0, 1)) \\ &= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \left(\log \frac{e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}}{e^{-x^2/2}/\sqrt{2\pi}} \right) dx \\ &= \frac{1}{2} \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} [-\log\sigma^2 + x^2 - (x-\mu)^2/\sigma^2] dx \\ &= \frac{1}{2} (-\log\sigma^2 + \mu^2 + \sigma^2 - 1) \end{aligned} \quad (2)$$

where μ and σ are separately the estimated the mean and variance. The decoder is then used to reconstruct the inputs after sampling by introducing the reconstruction loss in loss function.

$$L_r = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (3)$$

Therefore, the training goal for VAE is minimizing the KL divergence and the reconstruction loss.

3. Physics-consistent error compensation model

The physics-consistent error compensation model is divided into three modules: data reconstruction module, adaptive parameter generation module, and error compensation module to deal with the inherent noise of sampled data, the lack of real-time features in RLC model, and the potential error compensation.

3.1. Data reconstruction module

For industrial modeling, the noise is inevitable when data is sampled from field devices, resulting in model drift. Thus, it is necessary to reconstruct the input dataset for denoising. We assume that the input sampled data $\mathbf{x} = \{x_i\}_{i=1}^N \in \mathbb{R}^N$ is consist of the real signals and the

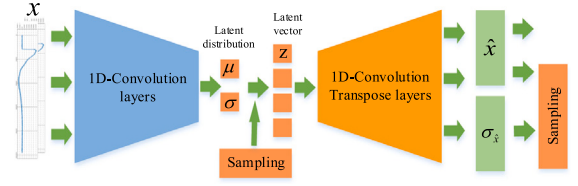


Fig. 2. Data reconstruction module.

white noise:

$$x_i = \bar{x}_i + \epsilon_i \quad (4)$$

where $\bar{x}_i$ is the actual input signals and ϵ_i is the white noise following the Gaussian distribution $N(0, \sigma_i^2)$. Therefore, the $x_i \sim N(\bar{x}_i, \sigma_i^2)$. To improve the modeling accuracy, the noise ϵ_i needs to be extracted before modeling and parameter tuning.

The data reconstruction module is composed of 1D-convolution layers and 1D convolution transpose layers, which is depicted in Fig. 2. Before being imported into the data reconstruction module, the raw sampled data is cut into multiple segments with k sampling points as data pre-processing. Based on the principle of VAE, the input data can be projected to a standard normal distribution $q_\phi(z|x)$ and then decoded into the original data. However, due to the existence of noise, the outputs are assumed to be drawn from the distribution $x_i \sim N(\bar{x}_i, \sigma_i^2)$. Thus, the reconstruction loss becomes the maximum likelihood estimation (MLE) instead of mean squared error (MSE) in Eq. (3). This is achieved by minimizing the negative log likelihood loss function:

$$L_r = \frac{1}{k} \sum_{i=1}^k \frac{1}{2} \log(2\pi\sigma_{\hat{x}_i}^2) + \frac{(x_i - \hat{x}_i)^2}{2\sigma_{\hat{x}_i}^2} \quad (5)$$

where $\hat{x}_i$ is the estimated mean of $\bar{x}_i$ and the $\sigma_{\hat{x}_i}$ is the estimated variance of x_i . By minimizing the $loss_D = L_{KL} + L_r$, the data reconstruction module can separate noise ϵ from raw data.

3.2. Adaptive parameter generation module

For RLC models, the essential parameters such as capacitance, resistance, and inductance determines the performance of models. However, the optimized deterministic parameters are hard to present the time-varying characteristics of transformers especially in a fixed model structure. Therefore, we create a super-network called the adaptive parameter generation module to generate these time-varying key parameters.

The adaptive parameter generation module is composed of several fully connected layers after flattening the estimated mean $\hat{x}_i$, which is shown in Fig. 3. The model can be denoted as $C_i, L_i, R_i = \Phi_\theta(\hat{x}_i)$. Here, we need to notice that the input of the module $\Phi_\theta(\cdot)$ is a vector, including the input from time $t - k$ to t instead of a scalar. This is due to one of our assumptions: the characteristics of the transformer do not change in a very short time. Hence, within k sampling points, the parameters of the RLC model remain unchanged. Consequently, we will simulate the RLC model k times to calculate the output y_{RLC} , since that we have k inputs. The corresponding loss is

$$L_{RLC} = \frac{1}{k} \sum_{i=1}^k (y_i - y_{RLC,i})^2 \quad (6)$$

where, the y is the sampled output, y_{RLC} is calculated output.

3.3. Error compensation module

Since the RLC model cannot fully reflect the detailed information of transformer under lightning strike, the error compensation module is used to compensate for this potential bias.

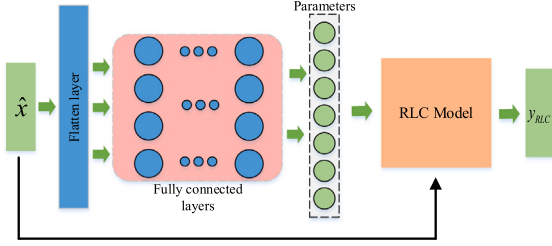


Fig. 3. Adaptive parameter generation module.

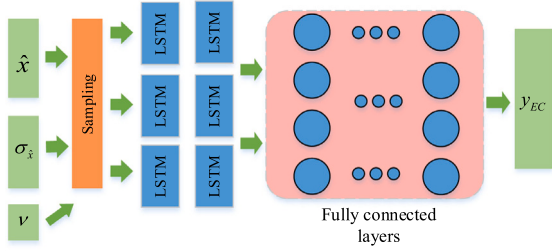


Fig. 4. Error compensation module.

Fig. 4 depicts the data flow in error compensation module. The model can be denoted as $y_{EC,i} = E(\hat{x}_i, \sigma_{\hat{x}_i}, v_i)$, where $E(\cdot)$ is the nonlinear function and v is a new variable for reparameterization. Different from the adaptive parameter generation module, the input of this module is the reparameterized $x_{r,i}$ instead of $\hat{x}_i$ to improve the robustness of error compensation. Because the noise ϵ_i follows the Gaussian distribution $N(0, \sigma_i^2)$, the reparameterization $x_{r,i}$ can be achieved by introducing a new variable v , which follows the standard normal distribution:

$$x_{r,i} = \hat{x}_i + v_i * \sigma_i \quad (7)$$

where v_i is sampled from $N(0, 1)$. Then, the long short-term memory (LSTM) is utilized to extract the underlying temporal relationships in the data for further error compensation. To ensure that the module is the error compensation of RLC model, the output y_{EC} is constrained in loss function:

$$L_{EC} = \frac{1}{k} \sum_{i=1}^k (y_i - y_{RLC,i} - y_{EC,i})^2 \quad (8)$$

3.4. Model structure and loss function

The physics-consistent error compensation model is realized by cleverly combining above mentioned modules, which is shown in Fig. 5.

The sampled input data is reconstructed to reduce the effect of noise on the RLC parameter generation and the corresponding RLC model output. In another aspect, the reconstructed data is reparameterized for error compensation module to improve the robustness of the whole model. The whole loss function of physics-consistent error compensation model is

$$Loss = L_{KL} + L_r + L_{RLC} + L_{EC} \quad (9)$$

where L_{RLC} and L_{EC} make sure that RLC model takes the dominant position in the whole model. Based on VAE theory, the new introduced parameter v does not affect the training process.

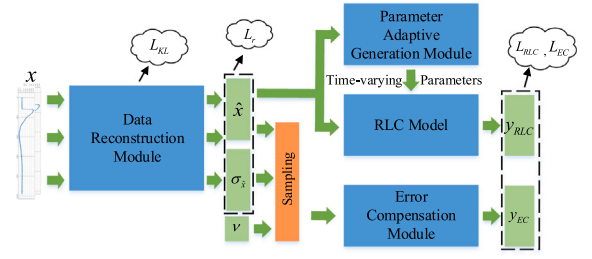


Fig. 5. Physics-consistent error compensation model.

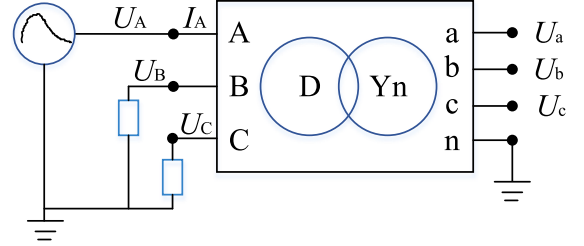


Fig. 6. The schematic diagram of the experiment.

4. Experimental results

4.1. Experiment setup

Surge response experiments are conducted on a 'Dyn11' transformer with the load applied to the high-voltage side. The schematic diagram of the experiment is shown in Fig. 6.

This distribution transformer is an oil-immersed type with a rated capacity of 630 kVA, a rated voltage of 10/0.4 kV, and a rated frequency of 50 Hz. The core is a three-legged laminated type, made of silicon steel sheets. The windings adopt a three-phase, dual-winding structure, and are made of enameled rectangular copper wire. Two extra 400 Ω resistances are added on the high-voltage side in experiments to equalize the cable resistance. The transformer tank is a rectangular corrugated oil tank. The low-voltage bushing uses a standard porcelain bushing for lead-out. A current transformer is installed at the low-voltage terminal to monitor the current on the low-voltage side, and the lead is connected to the end of the bushing and routed outside the tank for low-voltage side voltage measurement. The high-voltage bushing integrates an AC sensor that measures voltage on the high-voltage side. The combination wave generator is connected to either the high-voltage or low-voltage side of phase A.

Fig. 7 gives the detailed information of the experiment, including the transformer, the combined wave generator, the oscilloscope, and the isolation transformer. The parameters of the 'Dyn11' transformer are summarized in Table 1. We use the CWS 1000N combined wave generator to generate the 1.2/50~8/20 μ s hybrid-surge pulse, which conforms to IEC/EN 61000-4-5. The isolation transformer GBK₂ - 500 VA is used to protect the TDS3054C oscilloscopes. We collect the output voltages of all phases, and each voltage contains 10,000 sampled data with 0.1 μ s sample time.

When the generated lightning waves are injected into the high-voltage side of a three-phase distribution transformer, the high-frequency characteristics of the waves can significantly impact the internal characteristics of the transformer. The distributed capacitance

Table 1

The parameters of the ‘Dyn11’ transformer.

Basic Parameters		Structural Parameters	
Rated capacity (kVA)	630	Phase number	3
Rated voltage (KV)	25	Number of windings	2
Tap range	$\pm 2 \times 2.5$	Winding material	Copper
Rated frequency (Hz)	50	Core structure	Three-column laminated core
Maximum ambient temperature ($^{\circ}$ C)	40	Insulation type	Mineral oil
No-load loss (W)	510	Insulation class	A
Load loss (W)	4960	Cooling method	ONAN
Impedance voltage (%)	4.5	Model	Dyn11
Winding average temperature rise (K)	65	Voltage regulation method	non-excitation voltage regulation

of the transformer windings provides a coupling path for the lightning waves, resulting in strong electromagnetic coupling between the three-phase windings, which induces interference and ultimately manifests as high-frequency voltage spikes on the low-voltage side. Furthermore, high-frequency lightning waves trigger complex multiple reflections and superpositions within the windings, leading to an uneven distribution of voltage between winding layers and phases. In addition to capacitive coupling and high-frequency reflections, the injection of lightning waves can exacerbate interference through magnetic flux coupling in the windings. The combined effects of these factors result in more uneven distributions of voltage, current, and magnetic fields within the transformer, increasing insulation stress and operational risks for the equipment.

Our experiments are used to explore the characteristics of the transformer when it receives a sudden high-frequency signal. Therefore, the collected data is limited to explore all characteristics of transformers. Meanwhile, due to limited resources, the experiments are conducted on a ‘Dyn11’ transformer, which cannot reflect the characteristics of all transformers with different configurations.

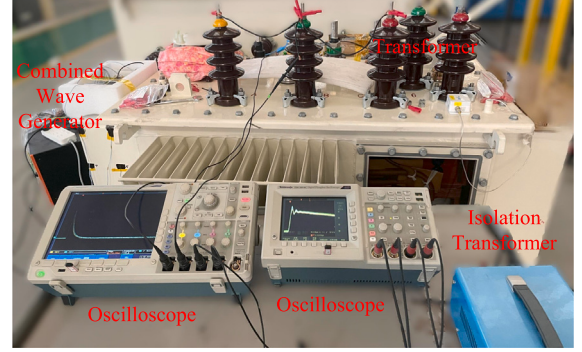
4.2. Model setup

For each block of PCEC, we have different parameters. In the data reconstruction module, the features are extracted by two 1D-convolution layers with 64 filters, and each kernel is 3×3 . The transpose layers have the corresponding parameters to reconstruct the data. The adaptive parameter generation module is composed of four fully connected layers after flattening the estimated mean $\hat{x}$. Each layer has 512 neurons and ‘PReLU’ [26] as the activation function. The error compensation module has two LSTM layers with 256 neurons to find the potential relationships in the reparameterized data. The subsequent fully connected layers have the same parameters as the adaptive parameter generation module.

We train the model with a batch size of 64 using the “NAdam” optimizer [27], with an initial learning rate of 0.01. The learning rate decays with a factor of 0.1 if the accuracy change over 5 consecutive epochs is no more than 1×10^{-4} . We set the number of epochs to 100 to ensure proper convergence. The dataset is split into training and test sets with a ratio of 0.8: 0.2. The experiments are implemented in Tensorflow using a CPU Intel i7-11800H processor at 2.3 Hz and a GPU NVIDIA T600.

4.3. Evaluation indicators

In order to compare the performance of different models, we use four different evaluation indicators to evaluate the regression accuracy, including mean absolute error (MAE), MSE, R-Square (R^2), and mean absolute percentage error (MPAE). The indicators have been widely accepted in regression model evaluation. MAE measures the average magnitude of errors between predicted and actual values. MSE measures the average of the squared errors, emphasizing larger errors due to squaring. R^2 quantifies the proportion of variance in the dependent

**Fig. 7.** Transformer experiment.

variable explained by the model and MAPE expresses errors as a percentage of actual values, providing a relative measure.

$$\begin{aligned}
 MAE &= \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \\
 MSE &= \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \\
 R^2 &= 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\bar{y}_i - y_i)^2} \\
 MAPE &= \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|
 \end{aligned} \tag{10}$$

where y_i is the experimental output, $\hat{y}_i$ is the model output and $\bar{y}_i$ is the mean of experimental output.

4.4. Analytical experiments

In this section, the effectiveness of the proposed method is evaluated by comparing PCEC with several RLC-based models, including those with physics-based parameters [25] (RLC) and optimization-based parameters [28] (OptRLC). Additionally, the efficiency of PCEC is assessed through comparisons with several machine learning techniques, including traditional neural networks (NN), K-nearest neighbor (KNN) regression, extra trees regression (ET), random forest regression (RF), convolution neural networks (CNN) and eXtreme Gradient Boosting (XGB). In order to make the comparison more targeted, we retain the data reconstruction module and adaptive parameter generation module when machine learning models are used.

4.4.1. The comparisons among RLC models and PCEC

Using the parameters of the ‘Dyn11’ transformer, we calculate the capacitance, leakage, and resistance variables for the RLC model. These calculated variables also serve as the initial input for the Particle Swarm Optimization (PSO) algorithm to obtain potentially improved variables.

Fig. 8 illustrates the output voltages of two selected phases when the combined wave generator is connected to Phase A and Phase A,

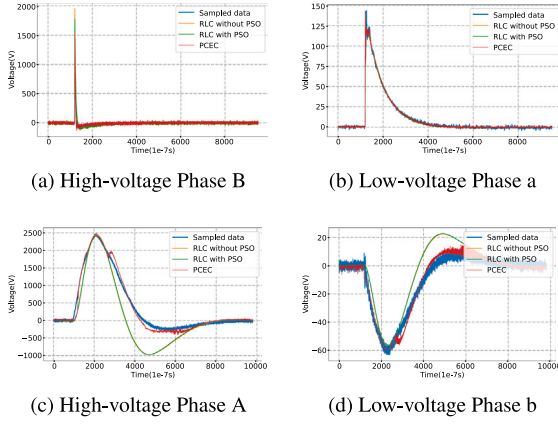


Fig. 8. The comparisons among RLC models and PCEC. The upper two figures depict the output voltages at Phase B and a when the combined wave generator is connected to Phase A. The down two figures depict the output voltages at Phase A and b when the combined wave generator is connected to Phase a.

Table 2

Evaluation metrics among RLC models and PCEC (combined wave generator is connected to Phase A).

Model	Phase	MAE	MSE	R^2	MAPE
RLC	B	17.32	1592.41	0.83	2.46
	C	17.57	1748.59	0.81	3.07
	a	1.36	6.91	0.99	11.99
	b	0.97	3.66	0.99	5.18
	c	0.42	1.28	0.33	1515.64
OptRLC	B	16.37	1559.13	0.82	2.29
	C	16.66	1691.36	0.87	2.30
	a	1.32	6.82	0.99	6.79
	b	1.03	3.81	0.99	4.06
	c	0.42	1.29	0.46	327.52
PCEC	B	11.57	370.02	0.93	6.73
	C	8.82	283.76	0.94	6.61
	a	1.23	6.11	0.99	9.08
	b	0.69	2.18	0.99	3.83
	c	0.40	1.17	0.91	5.04

respectively. The upper two plots show that the voltage gains of both the RLC and OptRLC models are excessively high, deviating from the experimental data. While the OptRLC model outperforms the original RLC model, it still does not match the performance of PCEC. Additionally, the voltage drop rate following the gain is faster under PCEC than in the other two models. In the lower two plots, PCEC facilitates the adaptive parameter generation within the embedded RLC model, resulting in less bias compared to the other RLC models. Despite some deviations in the error compensation module, the voltage obtained using PCEC closely matches the experimental data.

Tables 2 and 3 present the evaluation metrics. Due to the effective integration of the RLC model with error compensation, the evaluation metrics for PCEC are notably lower than those of the other models, with the exception of R^2 . The deterministic capacitance, leakage, and resistance variables, which are either calculated or optimized based on the transformer parameters, fail to fully capture the dynamic behavior of transformers. As a result, error compensation is crucial for improving model performance. Interestingly, while some studies suggest that optimizing RLC variables can enhance modeling accuracy, this approach does not seem to outperform PCEC in the current comparison.

4.4.2. The comparisons among several machine learning methods using same dataset

To evaluate the efficiency of PCEC on the same dataset as in Section 4.4.1, several machine learning techniques are employed to

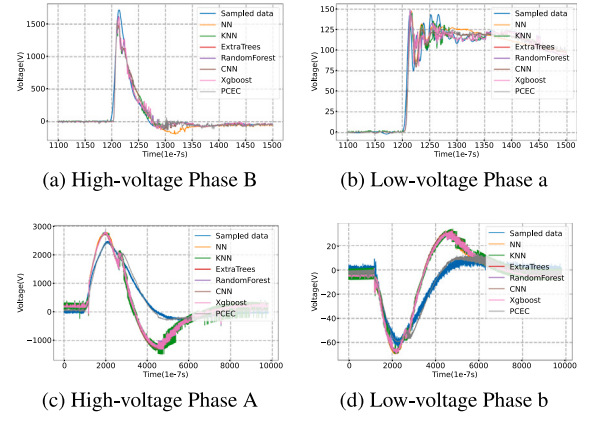


Fig. 9. The comparisons among machine learning models and PCEC. The upper two figures depict the enlarged output voltages at Phase B and a when the combined wave generator is connected to Phase A. The down two figures depict the output voltages at Phase A and b when the combined wave generator is connected to Phase a.

Table 3

Evaluation metrics among RLC models and PCEC (combined wave generator is connected to Phase a).

Model	Phase	MAE	MSE	R^2	MAPE
RLC	A	285.15	187 765.47	0.76	10.62
	B	25.19	1066.79	0.60	3289.32
	C	323.86	214 620.78	0.73	8.89
	b	6.68	88.69	0.79	72 099.01
	c	7.94	135.88	0.69	78 574.56
OptRLC	A	283.29	184 114.05	0.77	4.63
	B	25.19	1066.84	0.61	7293.93
	C	319.18	209 028.25	0.75	5.25
	b	6.59	85.69	0.80	37 309.72
	c	7.88	132.36	0.70	58 938.81
PCEC	A	66.12	8302.21	0.98	2.29
	B	11.28	261.47	0.66	0.73
	C	65.03	7772.54	0.99	0.97
	b	2.42	10.51	0.97	4.81
	c	2.75	13.59	0.96	4.93

construct models as alternatives to the LSTM-based error compensation approach.

Fig. 9 compares the output voltages from the machine learning models and PCEC. As shown in the upper two plots, there are notable discrepancies between the machine learning models and PCEC. Specifically, KNN, RF, XGB and ET exhibit significant oscillations, leading to substantial deviations in the voltage on side a compared to the experimental values. Although the NN model produces a smoother voltage response, it shows a slower voltage drop, as seen in Fig. 9(a), resulting in simulation errors. In the lower two plots, these issues become even more apparent. The voltage fluctuations in KNN, RF, XGB and ET models are even more pronounced than in the experimental data, particularly on the high-voltage side, with noticeable voltage deviations. The NN-based simulation, while smoother, is still outperformed by the other methods. In contrast, PCEC delivers the best performance by effectively integrating the RLC model with error compensation.

Tables A.7 and A.8 present the evaluation metrics. Although oscillations are observed in Fig. 9(b), there is no significant difference in the overall evaluation metrics, as shown in Table A.7. The used machine learning models demonstrate the potential to capture high-voltage variations, provided that additional parameters are incorporated. However, when the combined wave enters the transformer from the low-voltage side (Phase a), the discrepancy between the machine learning models and PCEC becomes evident. As depicted in Figs. 9(c) and 9(d), PCEC consistently outperforms all other methods across all evaluation

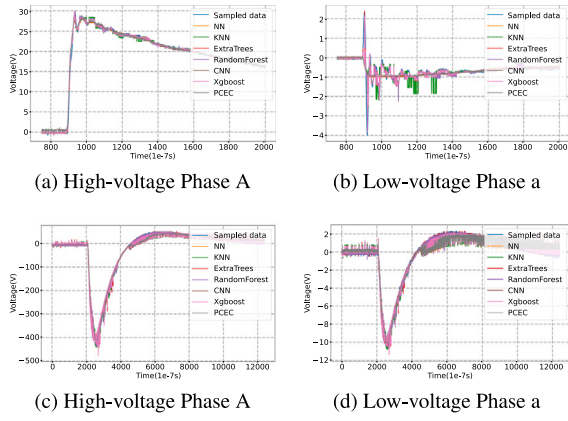


Fig. 10. The comparisons among machine learning models and PCEC. The upper two figures depict the enlarged output voltages at Phase A and a when the combined wave generator is connected to Phase C. The down two figures depict the output voltages at Phase A and a when the combined wave generator is connected to Phase c.

metrics. In situations where the low-voltage side of the transformer experiences a surge, traditional machine learning methods fail to accurately model the rapid response of the high-voltage side. By comparing the evaluation matrix in Table A.8 with PCEC, the PCEC outperforms other models at almost all phases, especially phase A.

4.4.3. The comparisons among several machine learning methods using additional dataset

To further verify the scalability and the priority of PCEC, we test PCEC using a new dataset in which the combination wave generator is connected to either the high-voltage or low-voltage side of phase C and the extra 400 Ω resistances are deleted in this experiment.

Fig. 10 shows the simulation results of all models. Different from the dataset from Section 4.4.1, the input signals in this section do not have a very high gain, which makes it easier to distinguish from the previous data. When the wave generator is connected to Phase C, all models perform equally well especially at phase A. In contrast, some extra oscillations in phase A are observed from other machine learning models occur when wave generator is connected to Phase c, resulting in the bias of simulation. Due to the low gain, the trend of all models are correct.

Table 4 shows the evaluation metrics of PCEC using additional dataset when the wave generator is connected to Phase C and c. Tables A.9 and A.10 shows the evaluation metrics of other machine learning models. Due to the low gain, the evaluation metrics are very close among all the methods. When the wave generator is connected to high voltage side, the models perform similarly, which is consistent with the performance in the figure. However, when wave generator is connected to low voltage side, PCEC outperforms other models. The evaluation metrics has been slightly improved when we use LSTM-based error compensation module.

4.5. Ablation experiments

This subsection explores the importance of data reconstruction module and adaptive parameter generation module by deleting them separately in PCEC.

4.5.1. The importance of data reconstruction

The data reconstruction module is designed to filter potential noise from peripheral sources, thereby enhancing modeling accuracy. Table 5 presents the simulation results for both the high-voltage side (HV) and the low-voltage side (LV). When the combined wave is applied from Phase A, the data reconstruction module does not perform well,

Table 4

Evaluation metrics of PCEC using additional dataset (combined wave generator is connected to Phase C and c).

Model	Phase	MAE	MSE	R^2	MAPE
HV	A	0.21	0.10	0.99	0.25
	B	0.14	0.09	0.99	0.12
	a	0.04	0.02	0.92	4.45
	b	0.04	0.02	0.78	0.62
	c	0.08	0.12	0.97	0.92
LV	A	4.19	95.74	0.99	0.43
	B	5.72	92.92	0.98	7.21
	C	2.10	52.98	0.99	0.30
	a	0.20	0.10	0.99	1.20
	b	0.27	0.15	0.98	1.74

Table 5

Evaluation metrics of PCEC without data reconstruction module.

Model	Phase	MAE	MSE	R^2	MAPE
HV	B	9.02	369.76	0.91	3.52
	C	9.12	324.31	0.92	8.50
	a	1.28	6.32	0.99	6.78
	b	0.81	2.85	0.99	3.29
	c	0.69	1.93	0.82	11.68
LV	A	386.98	285706.26	0.70	2.56
	B	11.89	272.29	0.64	0.71
	C	386.12	283348.73	0.73	2.92
	b	9.24	166.10	0.71	2.77
	c	9.22	172.51	0.66	2.07

Table 6

Evaluation metrics of PCEC without adaptive parameter generation module.

Model	Phase	MAE	MSE	R^2	MAPE
HV	B	8.11	2254.09	0.13	2.71
	C	7.90	2196.24	0.19	2.80
	a	1.52	6.53	0.98	9.76
	b	1.17	7.93	0.98	30.30
	c	0.42	1.25	0.61	9.33
LV	A	269.11	188836.61	0.60	10.76
	B	17.98	494.58	0.15	51.80
	C	280.27	209193.22	0.62	6.26
	b	8.30	161.34	0.36	11.75
	c	8.62	160.67	0.28	1.39

particularly in Phase B. Despite some minor differences, the evaluation metrics for other voltages remain similar to those of the original PCEC model. However, another experiment reveals distinctly different results. Given that noise signals can be amplified by the transformer, data reconstruction proves to be crucial when surge signals enter the low-voltage side. PCEC with data reconstruction significantly outperforms the model without data reconstruction, particularly on the high-voltage side. Therefore, data reconstruction is a critical component of the PCEC framework.

4.5.2. The importance of adaptive parameter generation

By removing the adaptive parameter generation module, PCEC is reduced to a data-driven modeling approach, rather than a physics-consistent model.

Table 6 summarizes the performance of PCEC without the adaptive parameter generation module. Unlike the case with the data reconstruction module, the adaptive parameter generation module incorporates essential physical information about transformers, providing valuable prior knowledge to the model. As a result, it plays a critical role in the overall performance of PCEC. The table shows that both experiments exhibit poor performance when this module is excluded. Even when the combined wave is applied to the high-voltage phase A, the voltage fits for individual phases are suboptimal. Consequently, with the same number of parameters, the physics-consistent PCEC model demonstrates superior performance, primarily due to the presence of the adaptive parameter generation module.

Table A.7

Evaluation metrics among machine learning models using same dataset (combined wave generator is connected to Phase A).

Model	NN					KNN				
Phase	B	C	a	b	c	B	C	a	b	c
MAE	9.13	9.25	1.25	0.57	0.47	9.14	10.07	1.29	0.69	0.40
MSE	404.87	364.61	6.77	2.61	1.71	372.46	313.74	5.12	2.06	1.16
R^2	0.91	0.92	0.98	0.99	0.87	0.93	0.93	0.98	0.99	0.89
MAPE	3.11	11.42	9.44	8.55	10.99	6.94	6.23	9.68	5.72	11.11
Model	ET					RF				
Phase	B	C	a	b	c	B	C	a	b	c
MAE	8.34	8.93	1.13	0.61	0.39	8.36	8.84	1.12	0.62	0.39
MSE	375.46	261.14	4.51	1.60	1.16	370.50	268.45	4.70	1.61	1.01
R^2	0.93	0.94	0.99	0.99	0.86	0.93	0.94	0.99	0.99	0.91
MAPE	6.76	11.76	2.98	2.98	8.31	2.56	6.91	38.89	2.98	8.30
Model	CNN					XGB				
Phase	B	C	a	b	c	B	C	a	b	c
MAE	8.53	9.64	1.26	0.72	0.43	9.31	8.94	1.21	0.62	0.42
MSE	373.49	318.82	6.23	2.65	1.31	372.60	352.32	6.67	2.63	1.47
R^2	0.91	0.92	0.99	0.99	0.89	0.92	0.93	0.98	0.98	0.91
MAPE	2.66	13.40	9.64	3.80	8.58	7.99	7.91	9.67	3.87	9.59

Table A.8

Evaluation metrics among machine learning models using same dataset (combined wave generator is connected to Phase a).

Model	NN					KNN				
Phase	A	B	C	b	c	A	B	C	b	c
MAE	392.12	12.43	391.12	9.24	9.10	403.15	11.99	399.18	9.60	9.34
MSE	280 899.72	282.70	280 181.41	160.84	160.93	307 714.96	346.25	301 704.24	174.29	182.18
R^2	0.71	0.59	0.74	0.72	0.68	0.69	0.60	0.72	0.70	0.65
MAPE	2.90	0.71	2.27	2.38	2.66	7.77	1.35	3.08	2.82	2.23
Model	ET					RF				
Phase	A	B	C	b	c	A	B	C	b	c
MAE	386.39	11.05	385.85	9.28	9.27	385.21	11.05	384.78	9.26	9.25
MSE	285 604.82	256.58	283 899.15	166.38	172.29	283 907.64	257.92	282 377.69	165.98	171.76
R^2	0.70	0.67	0.73	0.71	0.66	0.71	0.67	0.74	0.71	0.66
MAPE	3.39	0.72	2.43	2.08	3.00	2.50	0.71	15.93	2.06	2.48
Model	CNN					XGB				
Phase	A	B	C	b	c	A	B	C	b	c
MAE	381.08	12.19	385.15	10.11	9.99	387.20	11.30	386.51	9.29	9.28
MSE	280 589.80	285.26	281 763.99	193.07	192.35	287 372.58	265.80	285 517.30	166.86	173.04
R^2	0.71	0.67	0.74	0.68	0.63	0.71	0.67	0.73	0.71	0.66
MAPE	3.53	0.89	1.62	1.83	7.84	4.11	0.73	8.12	7.27	3.93

Table A.9

Evaluation metrics among machine learning models using additional dataset (combined wave generator is connected to Phase C).

Model	NN					KNN				
Phase	A	B	a	b	c	A	B	a	b	c
MAE	0.24	0.20	0.05	0.05	0.08	0.18	0.19	0.04	0.04	0.08
MSE	0.14	0.10	0.02	0.02	0.12	0.11	0.10	0.02	0.02	0.11
R^2	0.99	0.99	0.88	0.63	0.90	0.99	0.99	0.96	0.72	0.95
MAPE	0.86	0.12	8.38	0.69	1.18	0.05	12.80	10.05	0.72	25.03
Model	ET					RF				
Phase	A	B	a	b	c	A	B	a	b	c
MAE	0.17	0.15	0.03	0.03	0.06	0.20	0.16	0.04	0.04	0.07
MSE	0.08	0.06	0.01	0.01	0.08	0.09	0.07	0.01	0.01	0.08
R^2	0.99	0.99	0.98	0.81	0.99	0.99	0.99	0.97	0.79	0.99
MAPE	0.34	0.52	6.22	0.90	1.34	0.17	0.12	10.79	0.54	1.21
Model	CNN					XGB				
Phase	A	B	a	b	c	A	B	a	b	c
MAE	0.23	0.20	0.05	0.05	0.08	0.29	0.17	0.04	0.03	0.07
MSE	0.13	0.10	0.02	0.02	0.12	0.08	0.07	0.01	0.01	0.08
R^2	0.99	0.99	0.84	0.65	0.87	0.99	0.99	0.93	0.81	0.98
MAPE	0.55	0.14	5.69	0.83	1.01	0.26	0.17	7.63	1.37	1.27

5. Conclusion

Accurate and efficient wide-bandwidth transformer modeling is crucial for power systems. This paper presents a novel physics-consistent

error compensation modeling approach for wide-bandwidth transformers, referred to as PCEC, designed to operate under limited computational resources. We propose a data reconstruction module and incorporate an additional maximum likelihood estimation (MLE) loss

Table A.10

Evaluation metrics among machine learning models using additional dataset (combined wave generator is connected to Phase c).

Model	NN					KNN				
Phase	A	B	C	a	b	A	B	C	a	b
MAE	9.27	10.41	7.33	0.43	0.42	6.61	5.84	5.30	0.30	0.32
MSE	212.99	200.45	152.13	0.33	0.30	174.53	127.38	125.94	0.25	0.22
R ²	0.97	0.92	0.98	0.95	0.94	0.98	0.96	0.98	0.96	0.96
MAPE	1.48	7.14	1.73	3.73	9.54	8.39	5.83	0.49	3.74	0.92
Model	ET					RF				
Phase	A	B	C	a	b	A	B	C	a	b
MAE	4.28	4.56	3.35	0.20	0.23	5.40	5.25	4.29	0.25	0.26
MSE	96.55	87.31	68.01	0.15	0.13	119.50	94.98	86.63	0.17	0.15
R ²	0.99	0.98	0.99	0.97	0.97	0.98	0.98	0.99	0.97	0.97
MAPE	0.55	2.93	1.38	3.64	13.88	0.75	9.75	0.74	1.91	2.70
Model	CNN					XGB				
Phase	A	B	C	a	b	A	B	C	a	b
MAE	9.40	10.31	7.04	0.44	0.42	6.27	5.60	4.90	0.27	0.28
MSE	212.72	194.74	154.82	0.33	0.31	114.36	93.75	79.44	0.17	0.16
R ²	0.97	0.94	0.98	0.95	0.94	0.98	0.97	0.99	0.97	0.97
MAPE	1.30	2.90	0.71	4.98	7.80	1.04	4.09	1.09	2.40	2.90

function to denoise the input signals. Following this, the adaptive parameter generation module is introduced to integrate the time-varying physical characteristics of transformers into the RLC models within the PCEC framework. The error compensation module is employed to address potential deviations in the RLC models. Analytical experiments demonstrate that PCEC effectively compensates for errors between experimental and simulated data—an outcome that is difficult to achieve solely through RLC parameter optimization or machine learning methods. Ablation studies emphasize the essential role of both the data reconstruction and adaptive parameter generation modules, particularly on the low-voltage side.

A key limitation of PCEC is its sensitivity to data, which reduces its generalizability compared to purely physics-based models. The approach also requires additional data for training the neural networks. Nevertheless, we believe that PCEC holds significant potential for achieving physics-consistent modeling in power systems.

CRediT authorship contribution statement

Qianchao Wang: Writing – original draft, Validation, Software, Methodology, Data curation. **Chakhung Yeung:** Data curation. **Quan Zhou:** Investigation, Data curation. **Yuxuan Ding:** Validation, Software, Investigation, Data curation. **Yaping Du:** Writing – review & editing, Validation, Supervision. **Lei Jia:** Validation. **Song Zhang:** Validation.

Declaration of competing interest

No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. The author would like to declare on behalf of the co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

Acknowledgment

This work was supported by the China Southern Power Grid, China under Grant GZKJXM20222352.

Appendix. Comparison tables

See Tables A.7–A.10.

Data availability

The authors do not have permission to share data.

References

- [1] Li S, Li X, Kang Y, Gao Q. Load capability assessment and enhancement for transformers with integration of large-scale renewable energy: A brief review. *Front Energy Res* 2022;10.
- [2] Cao J, Du Y, Ding Y, Li B, Qi R, Zhang Y, et al. Lightning surge analysis of transmission line towers with a hybrid FDTD-peec method. *IEEE Trans Power Deliv* 2022;37(2):1275–84. <http://dx.doi.org/10.1109/TPWRD.2021.3083324>.
- [3] Sobouti M, Azizian D, Bigdeli M, Gharehpetian G. Multi-conductor transmission line model of split-winding transformer for frequency response and disk-to-disk fault analysis. *Int J Eng* 2021;34(6):1486–92.
- [4] Cheng B, Wang Z, Crossley P. Using lumped element equivalent network model to derive analytical equations for interpretation of transformer frequency responses. *IEEE Access* 2020;8:179486–96. <http://dx.doi.org/10.1109/ACCESS.2020.3027798>.
- [5] Mondal M, Kumbhar GB. Partial discharge localization in a power transformer: Methods, trends, and future research. *IETE Tech Rev* 2017;34(5):504–13.
- [6] Zhao X, Yao C, Abu-Siada A, Liao R. High frequency electric circuit modeling for transformer frequency response analysis studies. *Int J Electr Power Energy Syst* 2019;111:351–68.
- [7] Baravati PR, Hosseini SMH, Moazzami M. Comparing the new improved RLC and CMTL models for measuring partial discharge in transformer winding. *IEEE Trans Instrum Meas* 2021;70:1–10. <http://dx.doi.org/10.1109/TIM.2021.3088428>.
- [8] Hosseini SH, Baravati PR. New high frequency multi-conductor transmission line detailed model of transformer winding for PD study. *IEEE Trans Dielectr Electr Insul* 2017;24(1):316–23. <http://dx.doi.org/10.1109/TDEI.2016.005694>.
- [9] Lai Z, Mu H, Hua X, Zhou C, Lin H, Shao X, et al. A winding-core high-frequency equivalent circuit model for exactly simulating partial discharge signals propagation in transformer. *IEEE Trans Dielectr Electr Insul* 2024. <http://dx.doi.org/10.1109/TDEI.2024.3464581>, 1–1.
- [10] Liu J, Yao C, Yu L, Dong S, Liu Y. Using MLP to locate transformer winding fault based on digital twin. *Front Energy Res* 2023;11:1175808.
- [11] Kashani-Gharavi D, Faraji-Dana R, Mirzaei HR. A machine learning based hybrid algorithm for partial discharge localization in power transformers. *IEEE Trans Electromagn Compat* 2024;1–10. <http://dx.doi.org/10.1109/TEMC.2024.3482432>.
- [12] Hong K, Jin M, Huang H. Transformer winding fault diagnosis using vibration image and deep learning. *IEEE Trans Power Deliv* 2021;36(2):676–85. <http://dx.doi.org/10.1109/TPWRD.2020.2988820>.
- [13] Dey D, Chatterjee B, Dalai S, Munshi S, Chakravorti S. A deep learning framework using convolution neural network for classification of impulse fault patterns in transformers with increased accuracy. *IEEE Trans Dielectr Electr Insul* 2017;24(6):3894–7. <http://dx.doi.org/10.1109/TDEI.2017.006793>.
- [14] Moradzadeh A, Moayyed H, Mohammadi-Ivatloo B, Gharehpetian GB, Aguiar AP. Turn-to-turn short circuit fault localization in transformer winding via image processing and deep learning method. *IEEE Trans Ind Inform* 2022;18(7):4417–26. <http://dx.doi.org/10.1109/TII.2021.3105932>.
- [15] Xing Z, He Y, Wang X, Chen J, Du B, He L, et al. Vibration-signal-based deep noisy filtering model for online transformer diagnosis. *IEEE Trans Ind Inform* 2023;19(11):11239–51. <http://dx.doi.org/10.1109/TII.2023.3245193>.

- [16] Zhou Y, He Y, Xing Z, Wang L, Shao K, Lei L, et al. Vibration signal-based fusion residual attention model for power transformer fault diagnosis. *IEEE Sensors J* 2024;24(10):17231–42. <http://dx.doi.org/10.1109/JSEN.2024.3382811>.
- [17] Wu X, He Y, Duan J, Li Z, Zhao Y. Fault diagnosis method for transformer windings based on residual neural network and state code. In: 2021 international conference on advanced electrical equipment and reliable operation. 2021, p. 1–5. <http://dx.doi.org/10.1109/AEERO52475.2021.9708263>.
- [18] Arivamudhan M, Santhi S. Model based approach for fault detection in power transformers using particle swarm intelligence. In: *Recent advancements in system modelling applications: proceedings of national systems conference 2012*. Springer; 2013, p. 287–300.
- [19] Jamshidi S, Doostmohammadi A, Jamshidi A. Determining the intensity and occurrence location of faults in transformers using frequency response analysis (FRA) with novel multistage optimization algorithm and SVM decomposition technique. In: 2024 28th international electrical power distribution conference. 2024, p. 1–10. <http://dx.doi.org/10.1109/EPDC62178.2024.10571706>.
- [20] Chanane A, Houassine H. A novel methodology for the transformer winding equivalent ladder network circuit parameters identification by employing the frequency-domain and population based method. *Russ J Nondestruct Test* 2024;60(8):921–34.
- [21] Abu-Siada A, Mosaad MI, Kim D, El-Naggar MF. Estimating power transformer high frequency model parameters using frequency response analysis. *IEEE Trans Power Deliv* 2020;35(3):1267–77. <http://dx.doi.org/10.1109/TPWRD.2019.2938020>.
- [22] Alwady AA, Al-Ameri SM. Application of ANN and equivalent electrical circuit for understanding the power transformer frequency response analysis. In: 2024 international conference on green energy, computing and sustainable technology. 2024, p. 50–4. <http://dx.doi.org/10.1109/GEOST60902.2024.10475087>.
- [23] Kingma DP. Auto-encoding variational bayes. 2013, arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- [24] Qixian D, Pengbo S, Gang L, Richang X, Zaiming W. Lightning stroke model and protection analysis of distribution transformer with different wiring groups. *Electr Eng* 2023;(13):196–200.
- [25] Sun Y, Zhang L, Hou M, Gao H. Establishment and simulation of distribution transformer transient model under lightning stroke. In: 2015 5th international conference on electric utility deregulation and restructuring and power technologies. 2015, p. 1757–62. <http://dx.doi.org/10.1109/DRPT.2015.7432526>.
- [26] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. 2015, [arXiv:1502.01852](https://arxiv.org/abs/1502.01852). URL <https://arxiv.org/abs/1502.01852>.
- [27] Dozat T. Incorporating nesterov momentum into adam. In: *ICLR workshop*. 2016.
- [28] An X, Zhao F, Dou Z, Lu F, Wang Q. A RLC parameter identification method for synchronous buck converters based on simulated annealing coupled cuckoo algorithm. *Electr Eng* 2024;1–14.