PHYSICS-INFORMED INDUCTIVE BIASES FOR VOLTAGE PREDICTION IN DISTRIBUTION GRIDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Voltage prediction in distribution grids is a critical yet difficult task for maintaining power system stability. Machine learning approaches, particularly Graph Neural Networks (GNNs), offer significant speedups but suffer from poor generalization when trained on limited or incomplete data. In this work, we systematically investigate the role of inductive biases in improving a model's ability to reliably learn power flow. Specifically, we evaluate three physics-informed strategies: (i) power-flow-constrained loss functions, (ii) complex-valued neural networks, and (iii) residual-based task reformulation. Using the ENGAGE dataset, which spans multiple low- and medium-voltage grid configurations, we conduct controlled experiments to isolate the effect of each inductive bias and assess both standard predictive performance and out-of-distribution generalization. Our study provides practical insights into which model assumptions most effectively guide learning for reliable and efficient voltage prediction in modern distribution networks.

1 INTRODUCTION

Distribution networks are shifting from passive delivery systems to actively managed infrastructures shaped by high DER penetration, electrification, and frequent reconfigurations. These conditions create volatile conditions and tight real-time decision windows. Voltage prediction is a central task in this setting: maintaining voltages within operational limits is essential for efficiency, equipment safety, and system stability. Moreover, in distribution grids, voltage prediction is the main target as these networks typically contain only a slack bus and load buses. For an n-bus distribution system, the AC power-flow problem determines bus voltage magnitudes and phase angles from specified injections based on a coupled system of $2(n-1)$ nonlinear equations. Classical solvers can be accurate but scale poorly and may fail to converge under realistic operating conditions. Simplified linear methods are computationally efficient but unreliable for distribution grids, where higher resistance-to-reactance (R/X) ratios cause more voltage drops, which can lead to greater active power losses and impact voltage stability. These limitations have motivated increasing attention to machine learning approaches, which approximate the mapping from loads and network parameters to voltages without repeated numerical iterations. Recent studies, particularly those using Graph Neural Networks (GNNs), demonstrate that ML-based solvers can achieve orders-of-magnitude speedups while retaining strong accuracy and flexibility (Donon et al., 2020; Jeddi & Shafieezadeh, 2021). Unlike tabular models that must be retrained for each topology, GNNs naturally adapt to varying grid sizes and configurations with little structural modification. This adaptability makes them well-suited for distribution systems, which are characterized by heterogeneity and frequent reconfigurations. However, purely data-driven models suffer from poor generalization, especially under unseen configurations or operating conditions, a key barrier for deployment in real grids. This motivates a systematic investigation into how inductive biases can guide ML models toward more accurate, generalizable, and reliable predictions.

Inductive biases are assumptions embedded in a model's architecture, training process, or data representation that guide learning toward specific solutions. For voltage prediction, relevant inductive biases include the integration of graph-based models, physical laws (e.g., power flow equations), or operational constraints. By incorporating these biases, models can better capture the underlying structure of the problem, leading to improved performance and robustness. Among inductive biases, physics-informed approaches have received particular attention. These models embed prior knowledge of the system's governing equations to enforce physical consistency in a supervised or

unsupervised training regime. Nevertheless, most prior research has focused narrowly on equation-constrained losses, without investigating alternative ways of embedding physical knowledge. Moreover, little is known about the impact of these inductive biases on generalization performance, which is crucial for deployment in real-world grids where unseen configurations are the norm.

In this work, we investigate the role of inductive biases in learning voltage prediction for distribution grids. Specifically, we analyze three classes of physics-informed inductive biases:

1. **Power-flow-constrained loss functions**, which incorporate the physical equations directly into the optimization objective.
2. **Complex-valued neural networks**, which align model representations with the natural complex-valued structure of electrical quantities.
3. **Task reframing as residual prediction**, where the model learns to approximate deviations from a baseline solver rather than absolute voltages.

For each inductive bias, controlled experiments are conducted across a dataset of typical distribution grid topologies, evaluating standard predictive accuracy as well as generalization performance under unseen network configurations [1]. Due to their inherent generalization advantage, we focus on Graph Neural Network architectures in this study. By isolating and comparing the presented inductive biases rather than model complexity alone, we provide a systematic assessment of their effectiveness in guiding learning toward physically meaningful and generalizable solutions, highlighting new pathways for learning voltage prediction. Ultimately, our findings provide concrete guidance for the design of ML models that are not only fast and accurate but also reliable for safety-critical, real-world deployment in modern distribution grids. In this light, we make the following contributions:

- We provide a systematic assessment of physics-informed inductive biases for voltage prediction in distribution grids.
- We perform a comprehensive benchmarking of predictive accuracy and out-of-distribution generalization across heterogeneous low- and medium-voltage networks.
- We derive actionable design guidelines identifying which inductive biases most effectively enhance reliability and scalability, learning-based voltage prediction models.

2 Literature Review

2.1 Role of Inductive Biases in Machine Learning

Inductive biases are the foundational assumptions embedded in a machine learning model or algorithm. They encode prior knowledge, constraints, or assumptions to guide how a model generalizes from training data to unseen examples. Thus, the design and selection of inductive biases allow models to perform well in varied settings, boosting generalization and interpretability, while also imposing tradeoffs between efficiency and flexibility. These inductive biases are essential when learning from finite data, as they help constrain the hypothesis space and enable effective learning. Without assumptions, there are infinitely many functions that can fit the training data equally well, so machine learning models need a way to prioritize certain hypotheses over others.

Inductive biases have been systematically applied via algorithmic design (i.e., neural network architecture), training strategies (e.g., regularization), priors in Bayesian inference, and even feature selection. For example, one of the simplest inductive biases often followed in machine learning is Occam's razor (Blumer et al., 1987), which suggests the preference for simpler functions over more complex ones. Occam's razor is often implemented via regularization techniques that penalize model complexity, such as L1 or L2 regularization, and is one of the first principles taught in machine learning to overcome overfitting. However, inductive biases can take many other forms, such as the smoothness bias in kernel methods (Schölkopf & Smola, 2002), sparsity bias in Lasso regression (Tibshirani, 1996), maximum margin in SVMs (Cortes & Vapnik, 1995), and locality assumptions in nearest neighbors algorithms (Cover & Hart, 1967). In deep learning, architectural

[1] All experiment models and code available at: [Anonymized]

choices such as convolution or recurrent layers impose structural priors, often enabling remarkable generalization without explicit specification of inductive biases. When the assumptions align with the structure of the domain, certain biases can accelerate model training or make algorithms more sample-efficient; without them, a model may simply memorize the training data (overfit) or fail to learn altogether. However, inappropriate bias can lead to poor performance if the assumptions do not match the underlying data distribution. This highlights the importance of carefully selecting and designing inductive biases to suit the specific problem and data at hand.

2.2 Physics-Informed Machine Learning in Power Systems

Physics-informed machine learning integrates physical laws and domain knowledge into machine learning models to enhance their performance, interpretability, and reliability (Karniadakis et al., 2021). This approach is particularly relevant in engineering domains like power systems, where physical principles govern system behavior. By embedding these principles into the learning process, this technique aims to provide more trustworthy models that can generalize better to unseen scenarios. In power systems, physics-informed neural networks (PINN) have been applied to various tasks, including state estimation, dynamics analysis, and optimal power flow (Huang & Wang, 2023). Here we review three strategies of physics-informed machine learning that are particularly relevant to our study of voltage prediction in distribution grids.

2.2.1 Power-flow-constrained Loss Function

This is one of the most common approaches to physics-informed machine learning in power systems and was highlighted as a core PINN paradigm by a recent review paper (Huang & Wang, 2023). This is typically achieved by assigning physical meanings to the variables in the output neural networks (NNs) and embedding the governing equations of these physical variables into the loss function. The loss function can thus be formally written as:

$$\mathcal{L} = L_{\text{pred}}(\hat{\mathbf{y}}, \mathbf{y}) + \lambda L_{\text{reg}}(\mathbf{W}) + \gamma L_{\text{phys}}(\mathbf{X}, \hat{\mathbf{y}})$$

where L_{pred} is a typical loss function (such as mean-squared error) which measures the difference between the predicted and target output ($\hat{y}$ and y respectively); L_{reg} is a regularization term (such as L1 or L2 norm) imposed on the weights W of the neural network; and L_{phys} is a physical regularization term that quantifies the degree to which the model's predictions violate the governing physical equations, with X representing the input features. λ and γ are thus hyperparameters that balance the contributions of the different loss components.

For power flow estimation, L_{phys} is generally composed of the power balance equations of the AC power flow formulation, or a linearized version of this. More specifically, using P as the active power, Q as the reactive power, V as the voltage magnitude, and θ as the voltage angle, we can express the AC power flow loss term as follows:

$$L_{P_i} = P_i - \sum_{k=1}^{N} |V_i||V_k| \left(G_{ik}\cos(\theta_i - \theta_k) + B_{ik}\sin(\theta_i - \theta_k)\right)$$

$$L_{Q_i} = Q_i - \sum_{k=1}^{N} |V_i||V_k| \left(G_{ik}\sin(\theta_i - \theta_k) - B_{ik}\cos(\theta_i - \theta_k)\right)$$

$$L_{\text{phys}} = \sum_{i=1}^{N} (L_{P_i}^2 + L_{Q_i}^2)$$

where i is a particular bus in an N-bus network, and G_{ik} and B_{ik} are the real and imaginary parts of the admittance between buses i and k, respectively.

Since this method applies to the training process itself, it is model-agnostic and can be adapted for arbitrary neural networks architectures. Due to the ubiquity of this approach, we only highlight the most relevant studies here. GraphNeuralSolver (Donon et al., 2020) was one of the first works to apply this technique for power flow estimation, using their own custom GNN architecture and serving

as a benchmark for many future works in this direction. (Jeddi & Shafieezadeh, 2021) integrates this method into a Graph Attention Network (GAT) model in order to learn the importance of neighbouring nodes, while (Böttcher et al., 2023) adopts a randomized, recurrent GNN architecture, so that its predictor learns how to solve the power flow problem from many different starting points. All three works solve the problem in a strictly unsupervised manner, completely skipping the need for labeled training data and the L_{pred} term. Other notable works include (Zamzam & Sidiropoulos, 2020) and (Habib et al., 2024) who successfully apply this technique to distribution grid state estimation using non-GNN architectures. However, many of these works focus their analysis on the transmission system context, and none of them evaluate the generalization performance across a structured suite of distribution grid configurations.

2.2.2 COMPLEX-VALUED NEURAL NETWORKS

Complex-Valued Neural Networks (CVNNs) extend traditional real-valued neural networks by allowing weights, biases, and activations to be complex numbers. This enables the model to capture phase relationships and interactions between real and imaginary components more naturally, potentially leading to improved performance in tasks involving complex-valued data. CVNNs have been shown to be effective in various domains, including computer vision, speech and signal processing, medical image reconstruction, and control systems (Lee et al., 2022).

A typical layer computes

$$z' = Wz + b, \quad W \in \mathbb{C}^{m\times n},\ z \in \mathbb{C}^n,$$

which in real–imaginary form expands as

$$z' = (W_r x - W_i y) + i(W_r y + W_i x), \quad (z = x + iy).$$

Nonlinearities are either applied directly component-wise, $\sigma(z) = \sigma(x) + i\sigma(y)$, or in polar form, $\sigma(z) = f(|z|)e^{ig(\angle z)}$, where f and g are nonlinear functions applied to the magnitude and phase, respectively. Training employs *Wirtinger derivatives*, which calculates partial derivatices based on the complex weights and their complex conjugates, allowing standard gradient-based optimization in $\mathbb{C}$. Normalization and loss functions are defined analogously to the real case, often by acting on real and imaginary parts separately or jointly via $|z|$. For more details on CVNNs, the reader is referred to (Lee et al., 2022). In power systems, CVNNs naturally accommodate phasor quantities such as complex voltages and power flows, preserving their inherent algebraic structure.

To date, there are relatively few works that explore complex-valued neural networks for power system applications, despite the fact that electrical quantities such as voltages and currents are inherently complex-valued. For distribution grid voltage prediction, this allows voltage magnitudes and angles to be modeled jointly, rather than separately as in real-valued models. An early work developed a CVNN for load flow prediction and noted improved reliability compared to a real valued neural network baseline (Ceylan et al., 2005). However, both models were very small MLPs, consisting of one hidden layer of maximum 8 neurons, and evaluation was performed on a toy 6-bus network. Thus, the results seem hardly tractable for a real-world setting. A more comprehensive study is presented in (Wu et al., 2023), which introduces a complex-valued spatio-temporal graph convolutional network (Cplx-STGCN) for the tasks of power system state forecasting and false data injection detection, and highlight the advantages of this model against several baselines. However, none of the analyzed baselines include a real-valued counterpart of Cplx-STGCN. Moreover, the second complex-valued model presented, CplxFNN, performs poorly. Additionally, the evaluation examines a single IEEE 118-bus transmission system and the topology change experiment is limited to the removal of a single line. Therefore, the results remain largely inconclusive of the direct effects of CVNNs for distribution grid voltage prediction. Though not directly a machine learning model, the recently-proposed Physics-Informed Symbolic Regression (PISR) method leverages complex-valued representations to learn analytical approximations of the power flow equations (Eichhorn et al., 2025). The model embeds this complex information directly into the symbolic regression process, drastically elevating the performance of the non-physics-constrained variant, and allowing the model to learn more reliably with less data samples compared to the MLP baseline. Similarly, this study is limited to two small-scale systems, but the results are indicative of the promise of complex-valued representations for power system tasks.

2.2.3 Residual Prediction

Residual prediction reframes the learning task from predicting absolute values to predicting deviations from a known baseline or approximate solution. This can simplify the learning problem, as the model only needs to learn the residuals, which are often smaller in magnitude and may exhibit simpler patterns than the original target variable. The approach has been popularized in models such as ResNet (He et al., 2016) and XGBoost (Chen & Guestrin, 2016), which have had significant impacts in domains like computer vision and time-series forecasting.

Due to the complex physics of power systems, machine learning training can be unstable and heavily reliant on initializations and hyperparameters. However, voltage magnitudes and angles are typically close to a constant baseline set by the slack bus, thus motivating the use of residuals. By focusing on capturing deviations, the redundant learning of the trivial baseline is eliminated thereby reducing complexity. Currently, state of the art power-flow methods typically predict the absolute voltage targets directly. Residual learning was employed for probabilistic power flow (Chen & Zhang, 2023) and simulated on standard transmission system test cases. Particularly, both model-based and purely data-driven versions of weights initialization were utilized for the residual layer.

3 Methodology

3.1 Experiment Setup

For controlled comparison of the three inductive biases, we conduct a series of experiments using a consistent dataset, model architecture, and evaluation metrics. The ENGAGE dataset (Okoyomon, 2025) is a collection of low and medium voltage grids based on the SimBench networks (Meinecke et al., 2020). This test suite was introduced by Okoyomon & Goebel (2025) to evaluate the generalization capabilities of power flow models across several distribution grid configurations and promote robust grid planning and operation. With 3000 sample networks, derived based on structural and data variations of 10 base distribution grids, the dataset contains test cases with varying sizes, topologies, and electrical characteristics, making it ideal for evaluating generalization performance. We first establish a baseline to identify how well each model performs under known network configurations and learned data distributions. We randomly shuffle the data with all grids and adopt a 75/15/10 train/validation/test split for model training. To evaluate generalization performance, we conduct an Out-of-Distribution (OOD) experiment. We use a leave-one-out approach where we train on all but one of the base grid configurations and test on the held-out configuration. We thus apply the same train/validation/test split, with 300 samples (10%) in the held-out configuration. This approach allows us to assess how well each model can generalize to unseen grid topologies and operating conditions, which is critical for real-world deployment.

To evaluate model performance, we employ the standard metrics of Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) between the predicted and true voltage magnitudes and angles across all buses in the network. Additionally, we report the model training time to assess computational efficiency. To evaluate generalization performance, the RMSE of each held-out configuration is compared against the RMSE on the baseline test set. A smaller increase in RMSE/MAPE on the held-out configuration indicates better generalization.

3.2 Model Design

We provide a baseline GNN architecture to serve as a control model for our experiments. We employ a GraphConv-based message-passing architecture with batch normalization, following the standard GNN framework established by Morris et al. (2019). This provides strong representational capacity while maintaining theoretical grounding in GNN expressivity research. As input, the model accepts the active and reactive power injections and topological distance to slack bus as node features, and the line resistances and reactances as edge features. Additionally, every graph has access to the slack bus information (slack voltage magnitude, angle, resistance, and reactance) as global features. The model's task is then to predict the voltage magnitude and angle at every bus.

The model utilizes two-stage preprocessing and postprocessing layers, implemented using fully connected networks with ReLU activations and batch normalization. The core processing consists of 7 message-passing layers using the GraphConv update scheme:

$$x_i^{'} = W_1 x_i + W_2 \sum_{j \in N(i)} e_{ij} x_j$$

For all other models, we keep the architecture as similar as possible to the baseline, only changing the components necessary to implement each inductive bias. This approach allows us to isolate the effects of each inductive bias while minimizing confounding factors. We also apply uniform training procedures to ensure a fair comparison, using a mean squared error as a supervised loss function, with 3000 epochs, a batch size of 128, and an Adam optimizer with an adaptive learning rate starting from 10^{-3} until 10^{-5}. Using a hyperparameter tuning, we determine the optimal model configuration and ensure strong performance on standard accuracy metrics before proceeding with the experiment variants.

3.3 Power-flow-constrained Loss Model

We implement a loss function that combines the standard supervised loss with a physics-based regularization that penalizes violations of the AC power flow equations. More specifically, we train the baseline GNN as usual, and then fine-tune the model using the physics informed loss for 20 epochs. This scheme proved to be much more reliable than training completely unsupervised or with a weighted physics loss, since these methods are highly reliant on hyperparameters, with initial loss values beginning as high as 10^8. Furthermore, since the data is derived using pandapower (Thurner et al., 2018), we replicate their AC power balance equations and model assumptions to implement our loss function. We ensure that the physics informed loss function provides near-zero values for the ground truth voltage targets before using it for model training.

3.4 Complex-Valued Model

We implement a GNN architecture that operates on complex-valued inputs and outputs, using complex-valued layers and activation functions. This approach allows the model to naturally capture the relationships between voltage magnitudes and angles, which are inherently complex-valued quantities in power systems. Where applicable, we use the complexPyTorch library (Popoff, 2021) and implement custom complex-valued GNN layers such as complex batch normalization and complex GraphConv to ensure compatibility with the base GNN framework.

3.5 Residuals Model

We implement a GNN architecture that predicts the residuals as deviations from the slack bus voltage, rather than predicting absolute voltage values directly. This approach simplifies the learning task, as the model only needs to learn the differences from a known baseline (the slack bus voltage), which are often smaller in magnitude and may exhibit simpler patterns than the original target variable. We implement the residuals model by enforcing a single skip connection in the model, that connects the neural network's output to the slack bus reference value (Donti et al., 2017). This way, the model learns to output residuals, which are added to the slack bus reference values to produce the final voltage predictions. This allows us to keep the input data format consistent across all models, simplifying data handling and preprocessing.

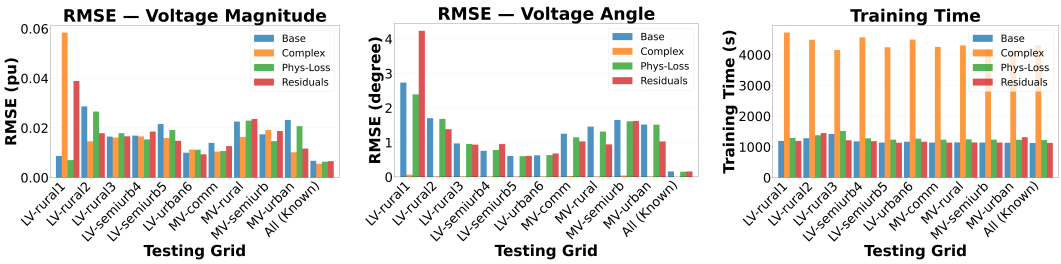
4 Results

We evaluate three inductive-bias variants for voltage prediction in distribution grids compared to a purely data-driven baseline model. We report performance for predicting grids with familiar structures, and use out-of-distribution experiments to examine model generalization. The evaluation networks are SimBench low-voltage (LV) and medium-voltage (MV) systems, consisting of a mixture of rural, semiurban, urban, and commercial networks. In this section, we present general model performance results across all experiments using RMSE and time metrics. For a complete view of individual benchmarking results, we refer the reader to Appendix A.

Table 1 provides a statistical summary of the generalization experiments, while Figure 1 displays RMSE and training time results for every test case in the study. Voltage magnitude RMSE is presented using p.u. values, meaning that results are normalized with respect to the reference voltage

Table 1: Model Performance Comparison: OOD Experiments Statistical Summary

Model	RMSE VM (p.u.)				RMSE VA (deg)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Base	0.0087	0.0287	0.0179	0.0062	0.6049	2.7343	1.3271	0.6458
Complex	0.0102	0.0584	0.0189	0.0142	**0.0093**	**0.0651**	**0.0212**	**0.0179**
Phys-Loss	**0.0070**	**0.0266**	**0.0166**	**0.0060**	0.6013	2.3928	1.2628	0.5595
Residuals	0.0094	0.0389	0.0183	0.0083	0.6101	4.2366	1.3419	1.0598



(a) Voltage Magnitude (b) Voltage Angle (c) Training time

Figure 1: Summary of selected voltage prediction performance across all experiments.

of the network, allowing for interpretation as relative errors (e.g., 0.01 p.u. is equivalent to 1% of the nominal voltage). In terms of voltage magnitude, the supervised physics-informed (PI) variant achieves the highest reliability across the simulated topologies, indicating greater robustness as a magnitude predictor. The complex-valued model attains the best angle accuracy, with an improvement of two orders of magnitude across all experiments, at the cost of increased computation time (x4). Although the residual learning model performs the best in certain cases, we observe that it does not provide any significant advantage, and exhibits low performance consistency with higher standard deviation. Generally, both the PI and complex-valued variants are the top performers, with relative advantage depending on the specific grid under consideration.

4.1 Baseline Model

When predicting voltage magnitude on known grid configurations, the baseline GNN attains an average RMSE of $\mathbf{0.0067}$ p.u. for voltage magnitude, and $\mathbf{0.1573}^\circ$ for voltage angle. For the OOD experiments, the baseline yields a mean performance of $\overline{\mathrm{RMSE}}(|V|) = \mathbf{0.0179}$ p.u. and $\overline{\mathrm{RMSE}}(\angle V) = \mathbf{1.3271}^\circ$. Notably, this model proves to be quite stable in its OOD prediction of voltage magnitude. Mean training time is **1187**s over all experiments.

4.2 Complex-Valued Learning

The complex-valued model excels on both magnitude and angle prediction in across test cases. It achieves $\mathbf{0.00239}^\circ$ angle RMSE and $\mathbf{0.00555}$ p.u. magnitude RMSE when predicting on known grid configurations, proving to be the best among all methods. Across the generalization experiments, we observe a continued stable accuracy, achieving $\overline{\mathrm{RMSE}}(\angle V) = \mathbf{0.0212}^\circ$ and $\overline{\mathrm{RMSE}}(|V|) = \mathbf{0.0189}$ p.u. These results underscore the remarkable angle predictive performance of this model for both familiar and unseen networks. However, while the angle predictive accuracy is the best among the variants, the complex-valued model does not achieve the top result for voltage magnitude prediction. Additionally, this performance gain comes with the highest computational burden, producing an average training time of **4346**s across all experiments.

4.3 Physics-Informed Learning

The supervised physics-informed variant exhibits an increase in performance compared to the baseline model for almost all experiments, illustrating the impact of this PI fine-tuning mechanism. For

the known grids setting, it reaches RMSE($|V|$) = $\mathbf{0.00638}$ p.u. and RMSE($\angle V$) = $\mathbf{0.1488}^\circ$. Across OOD benchmarks, the method exhibits the most accurate and stable performance for voltage magnitude prediction, with $\overline{\text{RMSE}}(|V|) = \mathbf{0.0166}$ p.u., and the smallest standard deviation. When considering voltage angle, its mean performance ($\overline{\text{RMSE}}(\angle V) = \mathbf{1.263}^\circ$) is only outperformed by the complex-valued model. The mean training time is **1283**s across all experiments.

4.4 Residual Learning

The residual approach provides performance comparable to the baseline model. When predicting on familiar grids, it attains RMSE($|V|$) = $\mathbf{0.00662}$ p.u. and RMSE($\angle V$) = $\mathbf{0.1573}^\circ$, showing nearly identical performance to the baseline. Across generalization experiments, accuracy and stability are also similar to the baseline, with $\overline{\text{RMSE}}(|V|) = \mathbf{0.01828}$ p.u. and $\overline{\text{RMSE}}(\angle V) = \mathbf{1.342}^\circ$. However, it is worth noting that, unlike the baseline, the residuals model proved to be the best performing model in certain situations, as made evident in the LV-semiurb5 and LV-urban6 OOD experiments. Training times are comparable to the baseline (**1200**s).

5 Discussion

5.1 Inductive Biases for Voltage Prediction

When estimating voltage magnitude and angle for networks similar to those seen in training, the complex-valued model provides the most accurate predictions. Voltage angle is particularly impressive, significantly outperforming the other variants by approximately two orders of magnitude. Generally, all physics-informed inductive biases seem to improve the results of the already-performant baseline model when they have had prior exposure to all grid types. This demonstrates that all three inductive biases can be used to assist in data-driven learning of power grid characteristics.

When looking at the results of the generalization experiments, the presence of unseen grid topologies clearly influences model accuracy, with average performance dropping by one order of magnitude on average. In this setting, the physics-informed variant attains the best performance on average for voltage magnitude and the lowest variability across individual OOD grids. The complex-valued model continues to thrive in voltage angle prediction. In fact, the complex-valued model's results on the OOD tests greatly surpass those of the other models under known grid configurations, highlighting its fidelity for voltage angle prediction.

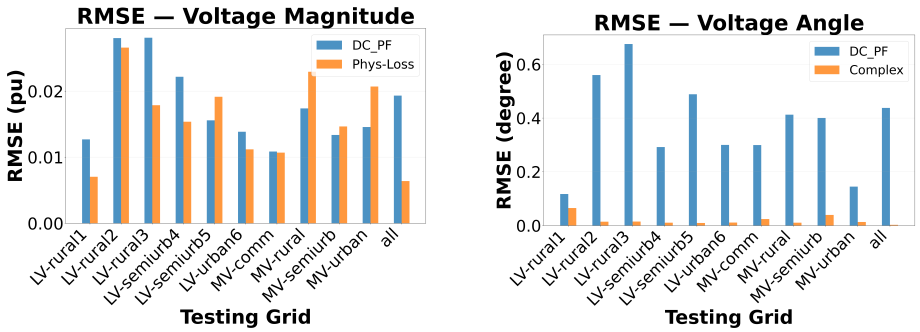
In light of these results, we propose that machine learning models leverage complex-valued neural networks to truly capture the interdependence between the real and imaginary components of the voltage phasors. *Fine-tuning* using physics-informed losses can also be a worthwhile add-on to enhance model accuracy, so long as one has accurate physical equations and relatively accurate initial predictive performance.

5.2 Real-world application

The complex-valued variant demonstrates noteworthy performance overall, achieving both high accuracy and stability across most tasks. This model surpasses the ML-based power-flow implementations of previous comparable studies, particularly in voltage angle prediction (Lin et al., 2024; Suri & Mangal, 2025; Okhuegbe et al., 2024; Hu et al., 2020). Contrary to prevalent data-driven approaches, the complex-valued variant naturally captures the inherent coupling between voltage magnitude and angle, treating them as interdependent variables. This precision helps satisfy a critical requirement in distribution system voltage prediction, where real power flows depend on angle differences and incorrectly estimated angles can lead to instability, congestion, or cascading failures.

The proposed physics-informed variant demonstrates particularly robust performance in predicting voltage magnitudes across diverse grid topologies. The incorporation of the AC power flow equations as a regularization term leads the learning process to remain consistent with Kirchhoff's laws, thereby reducing the likelihood of systematic deviations that might exist otherwise when extrapolating to unseen grid configurations.

As an additional interpretation, we compare the best performing models for voltage prediction to a commonly used solving method for voltage prediction: DC Power Flow. The DC power flow



(a) Voltage Magnitude

(b) Voltage Angle

Figure 2: Comparison of Best-Performing models to DC PF.

approximation linearizes the full AC method such that a non-iterative, convergent solution can be found. However, this simplification assumes small angle differences and voltages close to nominal, and thus struggles in situations with low voltage drops. Looking at Figure 2, we see that using GNN-based solvers, we are able to match the performance of DC PF for both known and unseen grids by adding physics informed loss functions for model fine-tuning. Additionally, by using complex-valued neural networks, we are able to outperform DC PF by one order of magnitude in voltage angle prediction. This is a significant result because it improves on a commonly observed limitation that GNN-based power flow methods perform well on known-configurations but are ineffective in predicting PF from unseen grids during training, and are usually outperformed by simplified models such as DC PF (Hansen et al., 2023; Yaniv et al., 2023; Okoyomon & Goebel, 2025).

5.3 Conclusion and Future Work

This study evaluates three inductive-bias strategies for voltage prediction in distribution grids. Using a dataset of heterogeneous low- and medium-voltage networks, we assess performance across familiar topologies and out-of-distribution evaluation cases. Our findings highlight complementary strengths. The complex-valued model achieves state-of-the-art angle prediction for both known and unknown grids, surpassing even DC power flow approximations, though at higher computational cost. The physics-informed variant consistently improves magnitude prediction and exhibits the lowest variability across unseen topologies, demonstrating robustness rooted in physical consistency. In contrast, the residual learning approach provided limited overall gains, though it excelled in select cases. Collectively, these results show that carefully chosen inductive biases substantially enhance the reliability of machine learning models for power system analysis.

Future research should extend these insights along three directions. Firstly, a deeper examination of physics-constrained training regimes is needed, including theoretical analysis and open benchmarking, to clarify when physics-informed losses improve generalization. For our particular study, this variant was the most sensitive to hyperparameters and required significant knowledge of the pandapower modeling library to produce an effective solution. Even then, the physics-informed loss could not be used competitively in a purely unsupervised manor (nor as a weighted loss addition), with performance results two orders of magnitude worse than the other variants at best. This finding comes at no surprise when considering the non-linear, non-convex nature of the AC power balance equations, but is nonetheless in conflict with several other works that tout its success. Secondly, evaluation on non-European grid datasets (e.g. Smart-DS) would test the applicability of these methods in unbalanced, three-phase networks with different structural characteristics. Finally, exploring hybrid models that combine multiple inductive biases could yield situation-dependent predictors, though early results suggest such benefits are non-additive and grid-specific.

By systematically studying inductive biases, this work contributes to the development of machine learning models that are not only accurate but also robust enough for deployment in real-world power system operations. Ultimately, our results indicate that physics-informed and complex-valued approaches can move ML-based solvers beyond academic benchmarks toward practical tools for reliable voltage prediction in distribution grids.

REFERENCES

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Occam's razor. *Information Processing Letters*, 24(6):377–380, April 1987. ISSN 0020-0190. doi: 10.1016/0020-0190(87)90114-1.

Luis Böttcher, Hinrikus Wolf, Bastian Jung, Philipp Lutat, Marc Trageser, Oliver Pohl, Andreas Ulbig, and Martin Grohe. Solving AC Power Flow with Graph Neural Networks under Realistic Constraints. In *2023 IEEE Belgrade PowerTech*, pp. 1–7, June 2023. doi: 10.1109/PowerTech55446.2023.10202246. URL http://arxiv.org/abs/2204.07000. arXiv:2204.07000 [cs].

Murat Ceylan, Nurettin Çetinkaya, Rahime Ceylan, and Yüksel Özbay. Comparison of complex-valued neural network and fuzzy clustering complex-valued neural network for load-flow analysis. In *Turkish Symposium on Artificial Intelligence and Neural Networks*, pp. 92–99. Springer, 2005.

Kejun Chen and Yu Zhang. Physics-guided residual learning for probabilistic power flow analysis. *IEEE Access*, 11:90309–90321, 2023.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

Balthazar Donon, Rémy Clément, Benjamin Donnot, Antoine Marot, Isabelle Guyon, and Marc Schoenauer. Neural networks for power flow: Graph neural solver. *Electric Power Systems Research*, 189:106547, December 2020. ISSN 0378-7796. doi: 10.1016/j.epsr.2020.106547. URL https://www.sciencedirect.com/science/article/pii/S0378779620303515.

Priya Donti, Brandon Amos, and J Zico Kolter. Task-based end-to-end model learning in stochastic optimization. *Advances in neural information processing systems*, 30, 2017.

Sebastian Eichhorn, Anurag Mohapatra, and Christoph Goebel. Pisr: Physics-informed symbolic regression for predicting power system voltage. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, pp. 92–107, 2025.

Benjamin Habib, Elvin Isufi, Ward van Breda, Arjen Jongepier, and Jochen L. Cremer. Deep statistical solver for distribution system state estimation. *IEEE Transactions on Power Systems*, 39(2):4039–4050, March 2024. ISSN 1558-0679. doi: 10.1109/TPWRS.2023.3290358.

Jonas Berg Hansen, Stian Normann Anfinsen, and Filippo Maria Bianchi. Power Flow Balancing With Decentralized Graph Neural Networks. *IEEE Transactions on Power Systems*, 38(3):2423–2433, May 2023. ISSN 1558-0679. doi: 10.1109/TPWRS.2022.3195301. URL https://ieeexplore.ieee.org/abstract/document/9847037. Conference Name: IEEE Transactions on Power Systems.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Xinyue Hu, Haoji Hu, Saurabh Verma, and Zhi-Li Zhang. Physics-guided deep neural networks for power flow analysis. *IEEE Transactions on Power Systems*, 36(3):2082–2092, 2020.

Bin Huang and Jianhui Wang. Applications of Physics-Informed Neural Networks in Power Systems - A Review. *IEEE Transactions on Power Systems*, 38(1):572–588, January 2023. ISSN 1558-0679. doi: 10.1109/TPWRS.2022.3162473. URL https://ieeexplore.ieee.org/document/9743327. Conference Name: IEEE Transactions on Power Systems.

Ashkan B. Jeddi and Abdollah Shafieezadeh. A Physics-Informed Graph Attention-based Approach for Power Flow Analysis. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1634–1640, December 2021. doi: 10.1109/ICMLA52953.2021.00261. URL https://ieeexplore.ieee.org/document/9680165.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

ChiYan Lee, Hideyuki Hasegawa, and Shangce Gao. Complex-valued neural networks: A comprehensive survey. *IEEE/CAA Journal of Automatica Sinica*, 9(8):1406–1426, 2022.

Nan Lin, Stavros Orfanoudakis, Nathan Ordonez Cardenas, Juan S. Giraldo, and Pedro P. Vergara. Powerflownet: Power flow approximation using message passing graph neural networks. *International Journal of Electrical Power & Energy Systems*, 160:110112, September 2024. ISSN 0142-0615. doi: 10.1016/j.ijepes.2024.110112.

Steffen Meinecke, Džanan Sarajlić, Simon Ruben Drauz, Annika Klettke, Lars-Peter Lauven, Christian Rehtanz, Albert Moser, and Martin Braun. SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis. *Energies*, 13(12):3290, January 2020. ISSN 1996-1073. doi: 10.3390/en13123290. URL https://www.mdpi.com/1996-1073/13/12/3290. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 4602–4609, 2019.

Samuel N Okhuegbe, Adedasola A Ademola, and Yilu Liu. A machine learning initializer for newton-raphson ac power flow convergence. In *2024 IEEE Texas Power and Energy Conference (TPEC)*, pp. 1–6. IEEE, 2024.

Ehimare Okoyomon. Engage dataset: Evaluating network generalization for ac grid estimation, May 2025. URL https://doi.org/10.5281/zenodo.15464235.

Ehimare Okoyomon and Christoph Goebel. A framework for assessing the generalizability of gnn-based ac power flow models. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, pp. 476–488, 2025.

Sébastien M. Popoff. complexPyTorch: A high-level toolbox for complex-valued neural network modules in pytorch, 2021. URL https://pypi.org/project/complexPyTorch/. PyPI release; MIT License; supports complex-valued operations like ComplexLinear, Conv2d, BatchNorm, CRelu, etc.

Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

Dhruv Suri and Mohak Mangal. Powergnn: A topology-aware graph neural network for electricity grids. *arXiv preprint arXiv:2503.22721*, 2025.

Leon Thurner, Alexander Scheidler, Florian Schäfer, Jan-Hendrik Menke, Julian Dollichon, Friederike Meier, Steffen Meinecke, and Martin Braun. pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Transactions on Power Systems*, 33(6):6510–6521, 2018.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Tong Wu, Anna Scaglione, and Daniel Arnold. Complex-value spatiotemporal graph convolutional neural networks and its applications to electric power systems ai. *IEEE Transactions on Smart Grid*, 15(3):3193–3207, 2023.

Arbel Yaniv, Parteek Kumar, and Yuval Beck. Towards adoption of GNNs for power flow applications in distribution systems. *Electric Power Systems Research*, 216:109005, March 2023. ISSN 0378-7796. doi: 10.1016/j.epsr.2022.109005. URL https://www.sciencedirect.com/science/article/pii/S0378779622010549.

Ahmed Samir Zamzam and Nicholas D. Sidiropoulos. Physics-aware neural networks for distribution system state estimation. *IEEE Transactions on Power Systems*, 35(6):4347–4356, November 2020. ISSN 1558-0679. doi: 10.1109/TPWRS.2020.2988352.

A Complete Benchmarking Results

We present the full benchmarking results in Table 2, which provides the raw performance metrics for all models across all testing grids, using RMSE, MAPE and training time as evaluation criteria.

Table 2: Raw Model Performance Results by Testing Grid

Testing Grid	Model	RMSE VM	RMSE VA	MAPE VM	MAPE VA	Time (s)
LV-rural1	Base	0.00868	2.73428	0.69669	1.82194	**1191.9**
LV-rural1	Complex	0.05842	**0.06514**	51.61096	1.89022	4722.3
LV-rural1	Phys-Loss	**0.00703**	2.39278	**0.56361**	**1.59738**	1287.7
LV-rural1	Residuals	0.03893	4.23657	2.36995	2.62264	1192.3
LV-rural2	Base	0.02866	1.70101	2.02971	1.06537	**1280.7**
LV-rural2	Complex	**0.01459**	**0.01416**	**1.10493**	**0.46439**	4486.7
LV-rural2	Phys-Loss	0.02658	1.68133	1.91425	1.04877	1376.7
LV-rural2	Residuals	0.01786	1.38314	1.49394	0.80328	1447.8
LV-rural3	Base	0.01650	0.97081	1.29962	0.58687	1419.6
LV-rural3	Complex	**0.01612**	**0.01459**	**1.27683**	**0.51593**	4153.4
LV-rural3	Phys-Loss	0.01787	0.95436	1.45133	0.57275	1517.4
LV-rural3	Residuals	0.01662	0.93536	1.30616	0.56816	**1213.8**
LV-semiurb4	Base	0.01690	0.75439	1.35473	0.49090	**1179.3**
LV-semiurb4	Complex	0.01657	**0.01053**	1.46401	**0.37068**	4564.2
LV-semiurb4	Phys-Loss	**0.01536**	0.78076	**1.29568**	0.50940	1275.9
LV-semiurb4	Residuals	0.01854	0.95156	1.48208	0.61797	1186.3
LV-semiurb5	Base	0.02154	0.60486	1.65013	0.35095	1137.9
LV-semiurb5	Complex	0.01590	**0.00933**	1.23808	**0.29761**	4244.1
LV-semiurb5	Phys-Loss	0.01914	0.60126	1.48068	0.34774	1233.5
LV-semiurb5	Residuals	**0.01481**	0.61009	**1.13968**	0.35005	**1134.1**
LV-urban6	Base	0.01000	0.62547	0.82257	**0.37637**	1165.8
LV-urban6	Complex	0.01129	**0.01110**	0.91247	0.37875	4493.8
LV-urban6	Phys-Loss	0.01118	0.63158	0.93967	0.38302	1265.2
LV-urban6	Residuals	**0.00937**	0.67909	**0.73814**	0.41331	**1164.8**
MV-comm	Base	0.01396	1.25189	1.13317	0.77431	**1138.5**
MV-comm	Complex	**0.01045**	**0.02395**	**0.82312**	0.78623	4253.3
MV-comm	Phys-Loss	0.01068	1.14607	0.87095	0.70025	1234.6
MV-comm	Residuals	0.01269	1.02931	1.08896	**0.57397**	1138.9
MV-rural	Base	0.02256	1.45830	1.85638	0.63937	1145.9
MV-rural	Complex	**0.01637**	**0.01055**	**1.36248**	**0.32539**	4303.7
MV-rural	Phys-Loss	0.02293	1.31355	1.90477	0.57902	1242.5
MV-rural	Residuals	0.02361	0.94275	1.69951	0.40661	**1143.2**
MV-semiurb	Base	0.01742	1.65200	1.45069	1.08598	**1138.9**
MV-semiurb	Complex	0.01917	**0.03934**	**1.12894**	1.20783	4189.4
MV-semiurb	Phys-Loss	**0.01465**	1.61223	1.13349	**1.06429**	1235.0
MV-semiurb	Residuals	0.01872	1.62399	1.44035	1.07307	1139.8
MV-urban	Base	0.02322	1.51757	1.73477	0.57886	**1131.7**
MV-urban	Complex	**0.01017**	**0.01314**	**0.82621**	**0.33376**	4083.5
MV-urban	Phys-Loss	0.02069	1.51453	1.49884	0.59561	1227.6
MV-urban	Residuals	0.01166	1.02719	0.93858	0.45834	1313.9
All (Known)	Base	0.00674	0.15726	0.50655	0.07627	**1125.2**
All (Known)	Complex	**0.00555**	**0.00239**	**0.38862**	**0.06419**	4311.9
All (Known)	Phys-Loss	0.00638	0.14876	0.48300	0.06915	1221.8
All (Known)	Residuals	0.00662	0.15726	0.49951	0.07154	1128.9