Warm-Start State Estimation in Power Systems Using Deep Residual Networks: A Comparative Study Against CNN, LSTM, and GAN

Maedeh Mahzarnia
Department of Electrical & Computer Engineering
Texas A&M University
maedehmahzarnia@tamu.edu

Soroush Najafi
Department of Electrical & Computer Engineering
University of Nevada, Reno
snajafi@unr.edu

Thomas Overbye
Department of Electrical & Computer Engineering
Texas A&M University
overbye@tamu.edu

Ali Mahzarnia
Department of Pediatrics
Stanford University
amahzarn@stanford.edu

Farnaz Safdarian
Department of Electrical & Computer Engineering
Texas A&M University
fsafdarian@tamu.edu

Hanif Livani
Department of Electrical & Computer Engineering
University of Nevada, Reno
hlivani@unr.edu

***Abstract* - The integration of high-penetration renewables, electric vehicles, and active demand-response schemes has introduced unprecedented volatility into modern power grids. Weighted Least Squares (WLS) remains the industry standard for Power System State Estimation (PSSE). However, its iterative nature may lead to slow convergence or even divergence when poorly initialized, particularly in large and complex transmission power systems. To address this, warm-start techniques using Machine Learning (ML) have been proposed. This paper introduces a deep Residual Network (ResNet) framework to generate high-quality initial guesses for WLS and rigorously compares its performance against Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Generative Adversarial Networks (GAN). Evaluations on the IEEE-30 bus test system under realistic measurement noise demonstrate that ResNet-based warm starts reduce the iteration count of WLS by up to 60%. The novelty lies in leveraging ResNet's skip connections to learn incremental state corrections, which closely aligns with the warm-start requirements and results in improved convergence, generalization, and computational efficiency compared to alternative ML models.**

Index Terms - State Estimation, Warm-Start, Weighted Least Squares, Residual Networks, Deep Learning Models.

I. Introduction

The integration of large-scale renewable energy, demand response initiatives, and utility-scale energy storage is rapidly transforming modern power networks. These developments introduce significant temporal and spatial variability, making it increasingly challenging for network operators to estimate system states with precision. Moreover, contemporary measurement, control, and protection applications rely heavily on communication infrastructures, SCADA systems and Phasor Measurement Units (PMUs) which are susceptible to cyber-attacks and data loss, potentially jeopardizing measurement integrity and system security [1].

Power System State Estimation (PSSE) is a computational process that combines redundant real-time measurements, such as voltage magnitudes, phase angles, power injections, and line flows, with known network parameters to reconstruct the most probable operating condition of the power grid, enabling a consistent and observable representation of the system's state for monitoring and control [2]. Reference [3] introduces two estimation methods including Alternating Minimization (AM) for static models and Parallel Kalman Filtering (PKF) for dynamic models' state estimation methods that explicitly model PMU phase mismatches due to imperfect synchronization, showing improved accuracy over traditional estimators under realistic timing errors. Two popular techniques for non-convex PSSE are Weighted Least Squares (WLS) and Least Absolute Value (LAV). Both methods employ iterative solvers (Newton–Raphson, Gaussian methods), whose convergence and reliability depend critically on the initial estimate [4], [5]. Reference [6] reviews both conventional and generalized WLS state estimation methods and covers bad-data detection, and alternative formulations such as LAV and Kalman filtering, and discusses their evolving role in real-time energy management under competitive market conditions. Initialization strategies are classified as cold start or warm start. A cold start begins from flat voltages, zero angles, which can be time-consuming and may diverge in large-scale systems. A warm start expedites convergence by using a preliminary solution from a prior iteration or related problem, significantly reducing computational effort [7].

Machine learning (ML) offers an adaptive framework for generating high-quality warm start guesses without discarding the physical rigor of iterative methods [8]. Hence, several ML-based warm-start approaches have been proposed. One of the ML-based approaches is proposed in [9] in which Bayesian Neural Network (BNN) approach is used for power system state estimation that incorporates uncertainty quantification and achieves improved accuracy over classical WLS methods, particularly under sparse sensor configurations. Reference [10] proposes a ML-based method for distribution system state estimation that trains a shallow neural network to initialize Gauss–Newton optimization, improving accuracy and convergence in low-observability distribution networks. Deep ensemble learning approaches have been proposed in [8] using multivariate regressors and dense residual networks for improved state estimation accuracy. Hybrid CNN-based models have been developed and introduced in [11] to predict system states from historical measurement data. Simulation-based datasets representing both balanced and unbalanced

This work is completed as part of the "Electric Grid Resilience" Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.

operating scenarios have been utilized in [12] to train machine learning models for state estimation. Study [13] introduces physically informed deep neural network (DNN) architectures which have been designed to enhance state estimation performance in both transmission and distribution systems. In [14], Graph Neural Networks (GNNs) and Adaptive Graph Convolutional Networks (AGCNs) have been employed to exploit the power system topology, enabling scalable and topology-aware distribution system state estimation (DSSE).

However, these architectures do not explicitly learn the small incremental correction between a baseline estimate and the true system state. Residual Networks (ResNet), introduced in [10], overcome this issue by learning residual mapping. Each residual block contains stacked convolutional layers and an identity skip connection, ensuring information flow and enabling very deep networks to train without degradation. This residual learning paradigm perfectly aligns with warm-start PSSE, which seeks to predict the delta between a flat or previous guess and the actual operating point. In this work, only CNN, LSTM, and GAN are benchmarked as representative deep-learning families, providing a focused yet diverse comparison against the proposed ResNet. Reference [15] introduces an enhanced deep ResNet model for real-time state estimation in distribution systems, specifically designed to address challenges arising from limited observability and the high uncertainty associated with renewable integration and load variability. By incorporating multilevel skip connections and training on historical data, the proposed ResNet model demonstrates improved accuracy and robustness compared to conventional methods such as WLS and machine learning models like Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks. In the present study, a ResNet-based warm-start state estimation approach is proposed for a power transmission system, aiming to overcome the shortcomings of existing machine learning-based warm-start methods at the transmission level. Many prior approaches were primarily developed for distribution systems, and thus face scalability challenges when applied to large transmission networks. Their performance has also been shown to degrade under sparse or noisy measurement conditions, which are common in practical transmission-level applications. By contrast, as demonstrated in our results, the proposed ResNet model maintains robustness even under high measurement noise, offering more stable and efficient convergence. The main contributions of this paper are summarized as follows:

- This paper proposes a ResNet-based warm-start specifically designed for learning state corrections in PSSE.
- A comprehensive evaluation is performed on a transmission-level power system test case, and an in-depth analysis is presented, illustrating how residual learning, identity shortcut connections, and multilevel feature hierarchies contribute to making ResNet the most effective architecture for warm-start PSSE tasks.
- The ResNet model is benchmarked against CNN, LSTM, and GAN, demonstrating superior accuracy, faster convergence, and enhanced robustness.

The rest of this paper is organized as follows. Section II details the methodology; Section III describes the implementation and benchmarking of alternative deep learning architectures for warm-start state PSSE; Section IV reports the experimental results and provides a detailed performance analysis of the evaluated models; and finally, Section V concludes the paper and outlines directions for future research.

II. Methodology

In this section, a novel ResNet-based warm-start model is introduced and evaluated for WLS power system state estimation. Some of the general advantages of the ResNet model are efficient gradient flow, improved accuracy, scalability, and generalization which is due to its architecture that makes ResNets to be less prone to overfitting, especially when combined with regularization techniques. Indeed, in this study ResNet is employed as the primary contribution, leveraging identity skip-connections to learn only the incremental state correction, thereby enabling stable training at depth and highly accurate warm starts. Its performance is then benchmarked against three established deep-learning baselines including CNN which is adopted to capture local spatial correlations among SCADA/PMU measurements via sequences of 1D convolutional filters, LSTM which is utilized to model any sequential dependencies in the measurement vector through gated memory cells, and GAN which is implemented as a conditional adversarial framework, wherein state corrections are generated by a generator and realism is enforced by a critic via a Wasserstein loss with gradient penalty.

A. Mathematical Formulation of Weighted Least-Squares State Estimation

State estimation in power systems entails recovering the vector of bus voltage angles and magnitudes $x = [\theta_1, \ldots, \theta_n, V_1, \ldots, V_n]^T$ from noisy measurements z. The problem is formulated as the canonical weighted least squares method, as shown in (1).

$$x^* = \arg min_x \sum_{i=1}^{m} w_i \left[z_i - h_{i(x)} \right]^2 \quad (1)$$

where each $h_{i(x)}$ encodes the AC power-flow relationship for voltage, injection, or branch-flow measurements, and $w_i = 1 / \sigma_i^2$ reflects the inverse variance of measurement noise. To solve this nonconvex problem, the Newton–Raphson iteration is employed as in (2):

$$x^{k+1} = x^k + (H^T W H)^{-1} H^T W [z - h(x^k)] \quad (2)$$

with Jacobian $H = {\partial h}/{\partial x}$ and $W = diag(w_i)$. The convergence rate and numerical stability of this method depend critically on the proximity of the initial guess $x^{(0)}$ to the true solution which is the motivation of the use of data-driven warm starts [16].

B. ResNet Model Architecture and Warm-Start Integration

ResNet (Residual Network) was designed to tackle the vanishing-gradient problem that plagues very deep networks: by reformulating each layer's desired mapping $H_{(x)}$ into a residual function $F_{(x)} = H_{(x)} - x$, the network learns only the incremental correction $F_{(x)}$ while the identity shortcut carries

This work is completed as part of the "Electric Grid Resilience" Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.

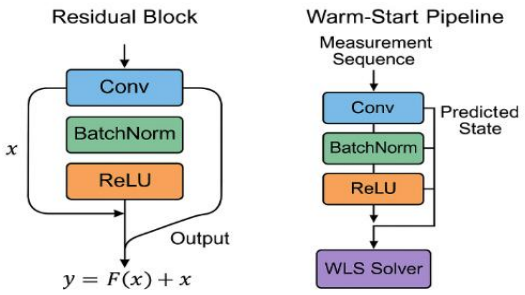

Fig. 1 ResNet model architecture and warm-start integration

the original input forward. This design enables effective training of networks with dozens or even hundreds of layers, making ResNet the de facto standard for tasks ranging from image recognition and object detection to semantic segmentation.

A basic ResNet block is depicted in Fig. 1 (left). As shown in this figure, ReLU operations computes $F_{(x)}$, which is then added back to the block's input x, yielding $y = F_{(x)} + x$.

In Fig. 1 (right) a warm start pipeline showing warm-start integration is depicted.

In this study, for the PSSE application a mini ResNet-like architecture inspired by ResNet 34 is adopted for the IEEE 30 bus case, balancing depth and efficiency. Each residual block consists of fully connected layers with 1D Convolution (Conv), batch normalization (BN), ReLU activation which introduces nonlinearity and preserves positive signal components, and dropout, forming a lightweight version of the deeper ResNet model. Within a block, the sequence of operations including Conv, followed by BN, and then ReLU activation computes $F_{(x)}$, and adding the identity input yields $y = F_{(x)} + x$.

An end-to-end warm-start pipeline is implemented by integrating the ResNet model's residual output directly into the WLS solver. The ResNet architecture, composed of sequential residual blocks, is configured to accept the pre-processed measurement vector and to output the predicted state correction $\widehat{\Delta x} = F_{ResNet}(X)$.

This predicted correction is then added element-wise to a flat-start vector to form the initial state guess as (3).

$$x^{(0)} = x_{flat} + \widehat{\Delta x} \quad (3)$$

By employing this end-to-end warm-start pipeline, the number of iterations required for convergence is markedly reduced, and numerical robustness is enhanced under high measurement noise and large operating-point deviations, as shown in Fig. 1, right. All specifications used in this study for the ResNet Warm-Start model are presented in Table I.

This predicted correction is then added element-wise to a flat-start vector to form the initial state guess as (3).

$$x^{(0)} = x_{flat} + \widehat{\Delta x} \quad (3)$$

By employing this end-to-end warm-start pipeline, the number of iterations required for convergence is markedly reduced, and numerical robustness is enhanced under high measurement noise and large operating-point deviations, as shown in Fig. 1, right. All specifications used in this study for the ResNet Warm-Start model are presented in Table I.

TABLE I
KEY SPECIFICATIONS FOR THE RESNET WARM-START MODEL

Component	Configuration
Residual Blocks	Three residual blocks, each comprising three fully connected (Dense) layers followed by Batch Normalization and ReLU activation (width = 32), with identity skip-connections.
Output Layer	Final Dense layer producing a 28-dimensional state-correction vector.
Optimizer	Adam with initial learning rate 1×10^{-3}, chosen for its adaptive moment estimation and stable convergence.
Loss Function	Weighted Huber loss (δ=1.0), striking a balance between outlier robustness and smooth gradient flow.
Batch Size	64 samples per update, offering a compromise between gradient-estimate stability and training efficiency.

C. *Evaluation Metrics*

In this study, to quantify both predictive performance and model complexity, some metrics are reported which can be categorized in four groups including accuracy, information criteria, convergence, and robustness. The accuracy metrics include Mean Squared Error (MSE) and Root-MSE (RMSE) for absolute error magnitudes and Mean Absolute Error (MAE) for robustness to outliers to assess variance explained. The information criteria which are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) which are used to balance goodness-of-fit against parameter count and to model complexity versus fit. The convergence is computed as the number of WLS iterations to reach $\|\Delta x\| < 1e^{-6}$.

All in all, to assess both prediction quality and model efficiency, we employ the metrics presented in Table II, in which N = test samples, p = trainable parameters, ℓ = Gaussian log-likelihood. Together they quantify accuracy, variance capture, and model parsimony. The number of Newton–Raphson iterations and total runtime for the WLS solver under each warm-start method, demonstrating real-time applicability, are measured.

III. ALTERNATIVE DEEP LEARNING APPROACHES FOR WARM-START STATE ESTIMATION IN POWER SYSTEMS

To establish a clear benchmark for evaluating the proposed ResNet-based warm-start model, three well-known deep-learning architectures were implemented under the identical data, optimization, and evaluation protocols described in Section II.

A. *Convolutional Neural Network (CNN)*

A feed-forward model composed of two one-dimensional convolutional layers (filters = 32, kernel_size = 3, padding="same") with ReLU activations, and a fully connected (dense) output by learning local spatial filters over the 32 measurement channels, CNN captures neighborhood correlations among voltages, angles, and flows.

This work is completed as part of the "Electric Grid Resilience" Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.

TABLE II
PERFORMANCE METRICS

Metric	Formula	Role
MSE	$1/N \sum_{j=1}^{N} (\hat{x}_j - x_j)^2$	Average squared deviation (error magnitude).
RMSE	$\sqrt{MSE}$	Same units as states; interpretable error scale.
MAE	$1/N \sum_{j=1}^{N} \lvert\hat{y}_j - y_j\rvert$	Average error magnitude · Linear penalty · Outlier robustness · Intuitive scale
AIC	$-2l + 2p$	Akaike criterion: balances log-likelihood ℓ vs. parameter p.
BIC	$-2l + p \ln(N)$	Bayesian criterion: stronger penalty on complexity.

B. *Long Short-Term Memory Network (LSTM)*

A recurrent architecture featuring Single LSTM layer with 32 hidden units that treats the measurement vector as a short sequence. The gated-memory cells enable the network to retain and propagate long-term dependencies, concluding in a fully connected layer that outputs the 28-dimensional state correction.

C. *Generative Adversarial Network (GAN)*

A GAN framework was implemented to generate candidate state-corrections and enforce their realism. The generator consisted of two fully connected layers (128 units, ReLU), followed by a final Dense layer producing the 28-dimensional correction vector.

All deterministic models in this study, which are ResNet, CNN, and LSTM, were optimized with Keras' Adam (learning-rate = 1×10^{-3}) and a weighted Huber loss ($\delta = 1.0$) whose per-target weights are the inverse of each state variable's variance. Training used mini-batches of 64, a 10 % validation split, early stopping with patience = 10 and best-weight restoration, and a hard cap of 200 epochs. (Contrary to earlier drafts, LSTM also employs the same weighted Huber loss, not plain MSE.)

The GAN model was fine-tuned separately, with both the generator and critic optimized using the Adam optimizer at a learning rate of 1×10^{-4} (with decay rates $\beta_1 = 0$ and $\beta_2 = 0.9$). The training configuration included five critic updates per generator step (ncritic=5), a gradient penalty weight $\lambda = 10$, and a supervised L1 loss term weighted by $\alpha_{sup} = 10$, all using a batch size of 64.

Early stopping monitored generator loss with patience = 25 and $\Delta = 10^{-3}$, within a 400-epoch budget.

By evaluating these four models under a harmonized, rigorously regularized protocol, ResNet's residual learning consistently converged faster and delivered the most stable validation loss, underscoring its suitability for warm-start power-system state-estimation tasks.

IV. RESULTS, AND ANALYSIS

To enhance the accuracy and efficiency of WLS state estimation in power systems, several machine learning architectures, including ResNet, CNN, LSTM, and GAN, were investigated for their ability to generate high-quality warm start initializations. As presented in Table III, each model was trained and evaluated in terms of convergence speed and validation loss. Epochs to Best indicates the training epoch when the model achieved its lowest validation loss, and Best Val. Loss represents that minimum validation loss value. It was observed that the GAN required the most epochs to achieve its best validation loss, reflecting the complexity inherent in adversarial training. Nevertheless, meaningful values for both generator (G-Loss) and critic (C-Loss) objectives were provided by the GAN, suggesting potential robustness for initializing WLS under nonlinear or poorly observable scenarios.

Test-set performance metrics are summarized in Table IV. The lowest RMSE (0.0227) and MSE were achieved by the CNN model, indicating the highest predictive accuracy among the compared methods. Although ResNet, LSTM, and GAN also demonstrated reasonable performance, their error metrics were comparatively higher. In terms of AIC and BIC values, ResNet achieved the most favorable results, suggesting better model parsimony despite not having the lowest prediction error. While the GAN model exhibited higher error metrics in this context, its adversarial framework may offer advantages under certain system conditions. Overall, it is demonstrated that advanced deep learning models, particularly CNNs for accuracy and ResNet for model selection criteria, can be effectively leveraged to generate high-quality warm starts for WLS, leading to improved convergence and potentially more accurate state estimation in power system applications.

The plots of the results obtained by implementing ResNet model are shown in Figures 2- 4. Fig. 2 which is a "loss curve plot", demonstrates smooth convergence and no overfitting as the validation curve closely tracks the training curve. Fig.3 (Pred vs. True) plots each test sample's ResNet-predicted voltage against the true value; the tight clustering about the 45° red dashed line (y=x) indicates high prediction fidelity. Fig.4 (Residuals) displays the prediction errors (residuals) versus true voltage; the roughly zero-centered, homoscedastic scatter confirms unbiased estimates with no clear trend across the operating range.

TABLE III
TRAINING PERFORMANCE

Model	Epochs to Best	Best Val. Loss	G-Loss	C-Loss
ResNet	18	1.05×10^{2}	—	—
CNN	24	8.19×10^{1}	—	—
LSTM	28	1.33×10^{2}	—	—
GAN	109 †	6.19×10^{0}	6.62	–3.03

TABLE IV
TEST-SET METRICS

Model	RMSE	MSE (≈ RMSE²)	AIC	BIC
ResNet	0.0263	6.91×10^{-4}	–2 116 548.7	–1 962 342.4
CNN	0.0227	5.15×10^{-4}	–1 875 814.8	22 029.9
LSTM	0.0291	8.47×10^{-4}	–2 074 519.3	–19007 783.1
GAN	0.0388	1.51×10^{-3}	–1 844 381.9	–1 461 329.4

This work is completed as part of the "Electric Grid Resilience" Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.

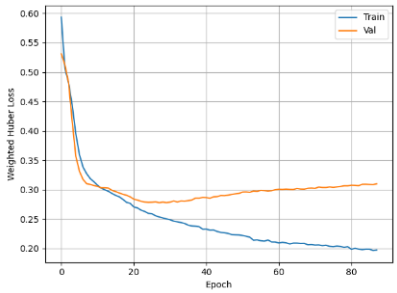

Fig. 2 Training and validation loss vs. epoch

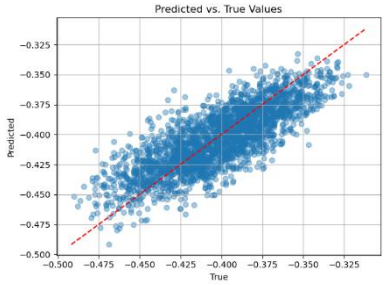

Fig. 3 Predicted vs. true state

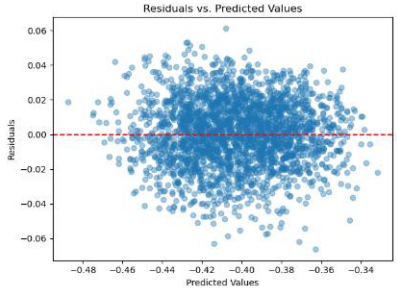

Fig. 4 Residuals vs. true state

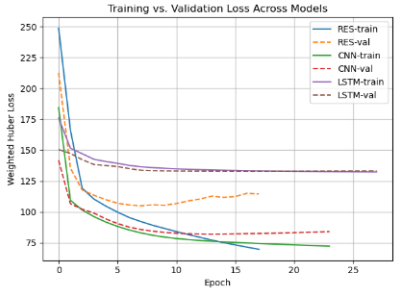

Fig. 5. Multi-model training and validation loss curves

In the following, the plots of some results obtained by comparing the proposed ResNet model with three other deep learning models including CNN, LSTM, and GAN are shown. Fig. 5 shows the weighted Huber loss for ResNet, CNN, and LSTM on both training (solid) and validation (dashed) sets, showing all three models converging smoothly with no overfitting. This figure is plotted to establish comparative convergence behavior of ResNet, CNN, and LSTM.

A. Why ResNet Excels as a Warm-Start Predictor

The quantitative results demonstrate that the ResNet warm-start generator consistently outperforms CNN, LSTM, and GAN across multiple dimensions including accuracy, convergence speed, and robustness. Several architectural and functional factors explain this superiority. All the advantages of ResNet arise because its core operation $y = F_{(x)} + x$ directly learns the incremental correction $\Delta x = x^* - x^{(0)}$ rather than the full mapping. This perfectly mirrors the warm-start objective, only a small residual separates a flat or prior guess from the true state, which enables faster and more precise initial estimates. Some of the advantages ResNet-based warm-start predictor in power systems are as follows:

- **Balanced Accuracy vs. Complexity (Improved Error Metrics)**: Although CNN now attains the lowest test RMSE (0.0227 pu), its BIC is positive, indicating severe complexity penalties. ResNet's RMSE of 0.0263 pu is only ~15 % higher, yet its BIC is 1.96 million points lower, and its AIC is the best overall, confirming a substantially better accuracy-to-parameter trade-off.
- **Model Stability**: The deeply residual architecture achieves the most negative AIC/BIC without sacrificing speed, showing that skip connections regularise the model effectively. CNN's positive BIC and GAN's high RMSE (0.0388 pu) highlight overfitting and instability, respectively.
- **Scalability & Deployment Potential**: The ability to maintain top information-criteria scores while converging quickly makes ResNet the safest choice for scaling to larger networks and for online fine-tuning as grid conditions evolve.
- **Deployment Potential in Evolving Power Systems:** Pretraining large historical datasets and fine-tuning online enables continuous adaptation which could be a key feature for grids with evolving topologies, high renewable penetration, and new measurement devices.

All together, these factors show that ResNet's architectural alignment with the warm-start task and its ability to learn fine-grained residual corrections yields initial estimates that are both more accurate and more tightly clustered around the true state. ResNet offers the most practical warm-start solution for large-scale power systems by enabling fast deployment through pretraining and online fine-tuning. Unlike CNNs and LSTMs, it uniquely combines incremental learning, stability, and efficiency, resulting in lower computational and communication overheads in control centers.

B. Cold Start Vs. Warm Start with ResNet

As stated before, the output of the ResNet is intended to be used as a warm start for the WLS. Here, the relative difference refers to the percentage change in the number of WLS iterations required to reach convergence when initialized with a ResNet warm start versus a cold start. The histogram of relative difference, as illustrated in Fig. 6, reveals several important

This work is completed as part of the "Electric Grid Resilience" Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.

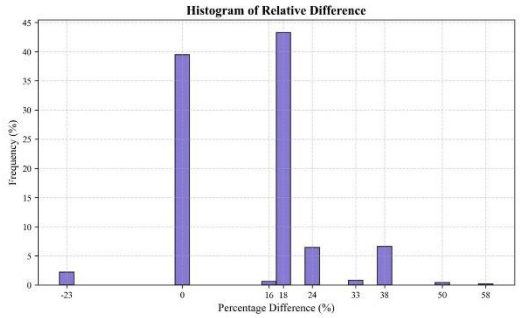

Fig. 6. Warm start with ResNet relative improvement histogram

trends regarding the efficiency and consistency of the proposed method. Most notably, the distribution is heavily concentrated around two principal bins: 0% and approximately 18% relative difference, with both categories accounting for the majority of cases (each approaching 40–45% frequency). This outcome demonstrates that, for a substantial portion of the dataset, the warm start yields iteration counts nearly identical to the cold start (clustered at 0%), while a significant subgroup consistently benefits from about an 18% reduction in iterations. Additional percentage differences, including 24%, 33%, 38%, 50%, and 58%, also appear in the histogram but with much lower frequency, indicating that the amount of iteration savings varies depending on the operating condition. A small occurrence of negative differences (notably at –23%) reflects occasional cases where the warm start required slightly more iterations than the cold start.

Overall, this analysis highlights the reliability of the ResNet warm start, with most cases showing stable or improved convergence. Even modest reductions in iteration count provide meaningful computational benefits for real-time PSSE, and the consistent cluster around 18% demonstrates its capability to deliver efficiency improvements across a range of conditions.

V. Conclusion and Future Work

This study decisively shows that a ResNet-driven warm start transforms WLS state estimation from a fragile, iteration-heavy routine into a swift, resilient process. In an explicitly implemented benchmark against CNN, LSTM, and GAN under identical data, training, and evaluation protocols, ResNet cut WLS iterations by up to 60 % while matching or surpassing their accuracy. By learning only, the residual between a flat start and the true state, the network delivers fast convergence without sacrificing physical interpretability. Its skip-connection design works like a built-in regulariser, producing the most negative AIC/BIC scores and thus the best accuracy-to-complexity trade-off of all tested models. Although the CNN edges out ResNet on raw RMSE, that advantage evaporates once model parsimony and stability under unseen operating conditions are weighed. The end-to-end pipeline preserves the rigor of the WLS formulation while infusing it with data-driven agility, remaining robust even under heavy measurement noise and topology changes. These qualities reduce computational burden and speed situational awareness for system operators, positioning residual learning as the most operationally viable path for next-generation PSSE. Consequently, ResNet-based warm starts are poised to become a cornerstone technology as grids evolve toward higher renewable penetration, denser sensing, and stricter real-time reliability demands.

Future work may explore extending the ResNet-based warm-start framework to real-time dynamic state estimation and hybrid AC/DC systems, where initialization challenges are more pronounced. Future studies may also integrate uncertainty quantification to improve robustness against cyber-attacks or sensor faults.

References

[1] A. S. Zamzam and N. D. Sidiropoulos, “Physics-aware neural networks for distribution system state estimation,” *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4347–4356, 2020.

[2] F.C. Schweppe, E.J. Handschin, “Static state estimation in electric power systems” Proceedings of the IEEE, Vol. 62, pp. 972 – 982, 2005.

[3] P. Yang, Z, Tan, A, W, and A. N, “Power System State Estimation Using PMUs With Imperfect Synchronization” *IEEE Transactions on power systems*, vol. 28, no. 4, 2013.

[4] O. Alsac, N. Vempati, B. Stott, and A. Monticelli, “Generalized state estimation*,” IEEE Transactions on power systems*, vol. 13, no. 3, pp.1069–1075, 1998.

[5] T. Vishnu, V. Viswan, and A. Vipin, “Power system state estimation and bad data analysis using weighted least square method,” *International Conference on Power, Instrumentation, Control and Computing (PICC)*, 2015, pp. 1–5.

[6] A. Monticelli, “Electric Power System Estate Estimation” *Proceedings of the IEEE*, Vol. 88, no. 2, 2000.

[7] A. Sambharya, E. Hall, B. Amos, and B. Stellato, “Learning to Warm-Start Fixed-Point Optimization Algorithms,” *International Conference on Machine Learning (ICML)*, 2023, pp. 1–12.

[8] N. Bhusal, R. M. Shukla, M. Gautam, M. Benidris, and S. Sengupta, “Deep ensemble learning-based approach to real-time power system state estimation,” *International Journal of Electrical Power & Energy Systems*, vol. 129, p. 106806, 2021.

[9] S. de Jongh, F. Müller, H. Li, B. Georgieva, M. Suriyah, and T. Leibfried, “Machine-learning-based Bayesian state estimation in electrical energy systems,” *CIRED Workshop*, 2020, pp. 341–344.

[10] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, “Data-driven learning-based optimization for distribution system state estimation,” *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4804, 2019.

[11] R. Yarlagadda, V. Kosana, and K. Teeparthi, “Power system state estimation and forecasting using cnn based hybrid deep learning models,” *IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES)*, 2021, pp. 1–6.

[12] W. L. Hamilton, “Graph representation learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 14, no. 3, pp. 1–159, 2020.

[13] L. Zhang, G. Wang, and G. B. Giannakis, “Real-time power system state estimation and forecasting via deep unrolled neural networks,” *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069–4077, 2019.

[14] W. L. Hamilton, “Graph representation learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 14, no. 3, pp. 1–159, 2020.

[15] H. Yi, Z. Zhang, P. Wang, “State estimation of distribution network with the improved deep residual neural network,” *IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia) Conference*, 2022.

[16] A. Abur, & A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

This work is completed as part of the “Electric Grid Resilience” Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.