

PHYSICS-INFORMED BUT NOT PHYSICS-CONSISTENT: ERROR GEOMETRY AND SUBSPACE PROJECTION FOR NEURAL AC POWER FLOW

Changhun Kim¹, Timon Conrad², Redwanul Karim¹, Karan Pahlajani¹, Julian Oelhaf¹, David Riebesel²
Tomás Arias-Vergara¹, Andreas Maier¹, Johann Jäger², Siming Bayer¹

¹Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

²Institute of Electrical Energy Systems, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

ABSTRACT

Recent neural power-flow solvers, including emerging foundation models, achieve accurate voltage predictions, yet such accuracy does not necessarily imply physically consistent solutions. Even small complex voltage errors can yield large AC power-balance residuals. We study this accuracy–consistency gap across PIGNN-GC, GridSFM, gridfm-graphkit, and LUMINA on realistic 2224-bus Great Britain network (GBnetwork) scenarios, with cross-grid evaluation of GridSFM over 31 systems. Using a singular value decomposition (SVD) basis fitted to training AC power-flow solutions, we find that neural prediction errors contain substantial components outside the dominant solution subspace. To address this mismatch, calibrated solution-subspace projection (CSP) suppresses off-subspace prediction components after train-only bias calibration, reducing Mean PB by 67.0%, 37.8%, 40.5%, and 68.9% for PIGNN-GC, GridSFM, gridfm-graphkit, and LUMINA, respectively, relative to calibrated predictions, while improving voltage-magnitude accuracy in all four models. These results identify output-error geometry as an important factor in physics-consistent neural AC power flow. Code: github.com/Kimchangheon/neural-acpf-error-geometry

Index Terms— AC power flow, physics-informed neural network, graph neural network, error geometry, subspace projection

1. INTRODUCTION

AC power flow (AC-PF) is a fundamental tool for steady-state power-system analysis [1]. It is repeatedly solved in congestion management [2] and contingency studies [3], and underlies optimal power flow (OPF) [4, 5] and state-estimation workflows [6]. The need for both computational efficiency and physical consistency has motivated neural surrogates including direct data-driven approximation [7, 8], physics-informed training [9, 10, 11], feasibility-seeking mechanisms [12, 13], and learned projection-based correction [14].

More recent cross-grid and foundation-model efforts—including GridSFM [15], gridfm-graphkit and GENCO [16, 17], and LUMINA [18]—extend this direction with global context and topology-aware representations of heterogeneous grid components [19] designed to operate across diverse grids and operating conditions [20].

However, accurate voltage prediction does not necessarily imply physical consistency [21, 22]. Physics-informed objectives can

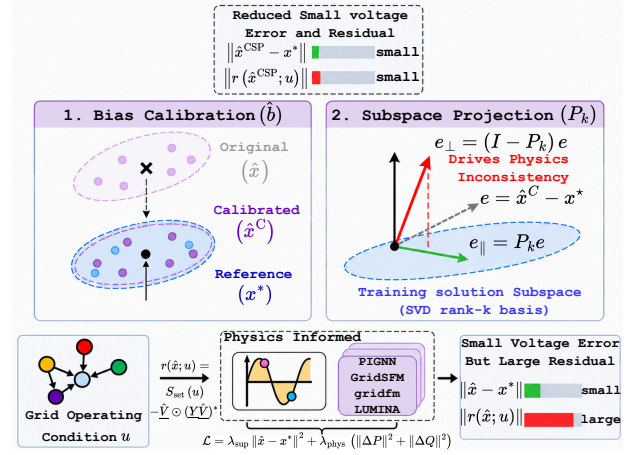


Fig. 1. A low-dimensional subspace is derived from the training solutions. CSP first removes train-estimated bias and projects the corrected prediction onto this subspace, suppressing off-subspace errors and improving voltage accuracy and AC power-balance consistency.

reduce power-flow residuals [23, 24], but remain soft training objectives and do not by themselves enforce exact satisfaction of the nonlinear AC equations [25]. Consequently, neural surrogates with small voltage errors can still exhibit non-negligible power-balance violations. We refer to this discrepancy as the *accuracy–consistency gap* and ask which properties of the remaining prediction error govern its physical impact.

We hypothesize that error *direction*, rather than magnitude alone, is an important factor: similarly sized voltage errors can induce substantially different AC residuals depending on their orientation across buses. Recent work has examined directional effects of warm-start errors in the distinct context of Newton–Raphson (NR) convergence [26]. This motivates studying the *geometry* of neural-surrogate errors in the voltage-state space.

Goodwin et al. [27] characterized the differential geometry of the AC-PF solution manifold, while Park et al. [28] and Chevalier et al. [29] demonstrated low-dimensional structure in AC-OPF and PF solutions, respectively. Motivated by this evidence of low-dimensional solution structure, we ask whether the relationship between neural prediction errors and a *training solution subspace* can be quantified to explain physical inconsistency, and whether this structure can be exploited to reduce power-balance violations.

To address these questions, we make three contributions. (i) We decompose calibrated prediction errors of four neural surrogates into

Corresponding author: changhun.kim@fau.de. This work was supported by the GridAssist project under the “OptiNetD” initiative of the German Federal Ministry for Economic Affairs and Energy (BMWE) and by HPC resources from NHR@FAU, partially funded by the German Research Foundation (DFG).

in- and off-subspace components using a low-rank singular value decomposition (SVD) basis fitted to training AC-PF solutions, and show through fixed-norm interventions and local Jacobian analysis that off-subspace errors are substantially more residual-sensitive. (ii) Building on this finding, we propose *calibrated solution-subspace projection* (CSP), a lightweight post-hoc output correction that uses only training data, removes systematic bias, and suppresses off-subspace prediction components without retraining. (iii) We evaluate CSP on frozen PIGNN-GC, GridSFM, gridfm-graphkit, and LUMINA models on the 2224-bus GBnetwork, where it lowers AC power-balance residuals and voltage-magnitude error in all four models and provides a better initialization for one Newton update. We further assess its generality on 30 additional systems and its robustness under topology and load-distribution shifts.

2. ERROR GEOMETRY AND CSP

2.1. AC power-balance consistency and calibration

For an N -bus grid, let $x = [V, \theta] \in \mathbb{R}^{2N}$ denote the voltage state, with $V = V \odot e^{j\theta}$. Given operating condition u , a surrogate predicts $\hat{x} = \hat{f}_\phi(u)$, from which (P, Q) are recomputed using the network admittance model. For setpoints $(P^{\text{set}}, Q^{\text{set}})$, we define

$$r(\hat{x}; u) = \begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} P^{\text{set}} - P \\ Q^{\text{set}} - Q \end{bmatrix}, \quad \text{PB} = \sqrt{\Delta P^2 + \Delta Q^2}.$$

PB uses ΔP at PV/PQ and ΔQ at PQ buses, zero elsewhere; prescribed PV/slack magnitudes and the slack angle are restored before scoring. Mean/Max PB aggregate all bus-scenario pairs on a common 100-MVA per unit (p.u.) base.

Before geometric analysis, we remove train-only prediction bias:

$$\hat{b} = \mathbb{E}_{s \in \text{tr}} [\hat{x}_s - x_s^*], \quad \hat{x}^C = \hat{x} - \hat{b},$$

where s indexes scenarios, $\star$ denotes the converged NR reference, and C denotes calibration. $\hat{b}$ is frozen at test time; angle bias is estimated as the per-bus circular mean of wrapped prediction errors.

2.2. Error geometry and training solution subspace projection

From training Newton-Raphson solutions, let $X_V, X_\theta \in \mathbb{R}^{S_{\text{tr}} \times N}$ stack centered magnitudes and circularly centered angles:

$$X_V = W_V \Sigma_V U_V^\top, \quad X_\theta = W_\theta \Sigma_\theta U_\theta^\top, \quad P_k = U_k U_k^\top, \quad (1)$$

where $U_k = \text{diag}(U_{V,k}, U_{\theta,k})$ contains the leading k state-space directions per channel; $k = 16$ denotes 16 per channel.

For calibrated error $e = \hat{x}^C - x^*$,

$$e^\parallel = P_k e, \quad e^\perp = (I - P_k) e, \quad \hat{x} - x^* = \hat{b} + e^\parallel + e^\perp. \quad (2)$$

Here, $\hat{b}$ is the train-estimated mean bias, while $e^\parallel$ and $e^\perp$ partition the remaining error into in- and off-subspace components.

Around a converged state with fixed topology and bus types,

$$r(x^* + e; u) = r(x^*; u) + J(x^*; u) e + \mathcal{O}(\|e\|_2^2), \quad (3)$$

where $r(x^*; u) = 0$ and J is the AC residual Jacobian; hence equal-norm errors can yield different residuals by orientation.

Motivated by this orientational sensitivity, CSP first removes $\hat{b}$ and then projects onto the dominant training-solution directions:

$$\boxed{\hat{x}^{\text{CSP}} = \bar{x}_{\text{tr}} + P_k (\hat{x} - \hat{b} - \bar{x}_{\text{tr}})} \quad (4)$$

where $\bar{x}_{\text{tr}}$ is the training-solution mean; both $\bar{x}_{\text{tr}}$ and P_k remain fixed at test time. Projection can remove prediction error but also discard valid reference variation. Let $e^C = \hat{x}^C - x^*$ and $e^{\text{CSP}} = \hat{x}^{\text{CSP}} - x^*$. Since P_k is orthogonal,

$$\|e^{\text{CSP}}\|_2^2 - \|e^C\|_2^2 = \|(I - P_k)(x^* - \bar{x}_{\text{tr}})\|_2^2 - \|(I - P_k)e^C\|_2^2. \quad (5)$$

Hence, CSP improves state accuracy when the removed off-subspace prediction-error energy exceeds the discarded reference variation.

CSP is a lightweight, data-driven post-hoc output correction rather than a projection onto the nonlinear AC-feasible set; it therefore does not guarantee exact feasibility, but can improve physical consistency when the removed off-subspace components are residual-sensitive under the local AC Jacobian.

3. EXPERIMENTS AND RESULTS

Our primary study uses 37,022 converged pandapower Newton-Raphson operating points on the 2224-bus GBnetwork, split 1:1:1 for training/validation/test. The dataset combines global load/generation scaling with independent load P/Q and active-generation perturbations; the backbone topology and PV-voltage control schedule remain fixed. We evaluate PIGNN-GC, our global-context extension of PIGNN-Attn-LS [25] retaining its line-search, GridSFM [15], gridfm-graphkit [16], and LUMINA [18]; GridSFM is also evaluated cross-grid on 31 systems.

All models are trained from scratch using model-specific stable initialization, and CSP is evaluated within each frozen model. gridfm-graphkit uses its released architecture/configuration without a pretrained checkpoint. Both GridSFM and LUMINA retain their released backbones and are adapted from AC-OPF to AC-PF by fixing generator dispatch, predicting only voltage states with PF-specific residuals, and applying known-variable masking. All models use pre-resolution pandapower [1, 30] internal PPC-derived inputs. Training and inference use one NVIDIA H100 96-GB GPU; inference timing uses the largest memory-fitting batch to maximize GPU utilization. Calibration statistics and solution bases use training data only. From $k \in \{4, 8, 16, 32, 64\}$, validation Mean PB selects $k^* = 16$ without degrading calibrated voltage-magnitude RMSE and is frozen for test and distribution-shift experiments. Unless otherwise stated, mechanistic, ablation, and cross-grid analyses use model seed 42, whereas Tables 1 and 3 aggregate seeds 41–43.

Voltage RMSE alone can be small even when a model largely predicts a static per-bus profile rather than following scenario-dependent voltage variation [15]. To isolate this scenario-dependent response, we center each bus as $\tilde{V}_{is} = V_{is} - \bar{V}_i$ and $\tilde{\hat{V}}_{is} = \hat{V}_{is} - \bar{\hat{V}}_i$, and report

$$R_w^2 = \text{corr}^2(\tilde{V}_{is}, \tilde{\hat{V}}_{is}), \quad a_w = \frac{\text{Cov}(\tilde{V}_{is}, \tilde{\hat{V}}_{is})}{\text{Var}(\tilde{V}_{is})}. \quad (6)$$

R_w^2 is the squared correlation between the within-bus-centered prediction and reference across all bus-scenario pairs, while a_w measures the scale of the predicted variation relative to the true variation.

3.1. Off-subspace error and directional sensitivity.

Off-subspace error. Off-subspace error is substantial across models and is positively associated with Mean PB. Although the rank-16 basis explains 98.8%/99.7% of training voltage-magnitude/angle variation, the median scenario-wise fractions of total squared error lying outside the respective subspaces are 26.3/55.0/56.8/27.1% for magnitude

Table 1. GBnetwork results at $k = 16$. All models report mean $\pm$ standard deviation over three model seeds. C denotes calibration, P_{16} projection, and CSP_{16} their composition. Inference time is GPU-synchronized forward execution plus the corresponding output transform, reported per scenario.

Model	Output	V RMSE $\downarrow$	θ RMSE ($^\circ$) $\downarrow$	$R_w^2 \uparrow$	a_w	Mean PB $\downarrow$	Max PB $\downarrow$	Inference (ms/scen.)
Per-bus mean	Baseline	.02341	6.847	.000	.000	.12084	5.88	–
PIGNN-GC	Raw	.02850 $\pm$.00232	1.108 $\pm$.175	.671 $\pm$.050	.546 $\pm$.072	.31592 $\pm$.03240	195.89 $\pm$ 52.54	10.034 $\pm$.023
	C	.01393 $\pm$.00105	.777$\pm$.161	.671 $\pm$.050	.546 $\pm$.072	.16661 $\pm$.02352	190.00 $\pm$ 37.53	10.028 $\pm$.023
	P_{16}	.01624 $\pm$.00254	1.091 $\pm$.166	.829$\pm$.045	.541 $\pm$.071	.07594 $\pm$.00422	5.07 $\pm$.85	10.036 $\pm$.023
	CSP_{16}	.01222$\pm$.00134	.835 $\pm$.141	.829$\pm$.045	.541 $\pm$.071	.05497$\pm$.00257	3.93$\pm$.07	10.034 $\pm$.032
GridSFM	Raw	.02868 $\pm$.01652	2.294 $\pm$.302	.768 $\pm$.151	.724 $\pm$.149	.26867 $\pm$.13004	9.10 $\pm$ 4.02	25.045 $\pm$.207
	C	.01086 $\pm$.00403	1.781$\pm$.055	.768 $\pm$.151	.724 $\pm$.149	.08251 $\pm$.01093	7.29 $\pm$ 3.53	25.015 $\pm$.096
	P_{16}	.01814 $\pm$.01145	2.244 $\pm$.325	.936$\pm$.032	.724 $\pm$.149	.07630 $\pm$.01942	5.08 $\pm$.49	25.087 $\pm$.279
	CSP_{16}	.00782$\pm$.00307	1.807 $\pm$.053	.936$\pm$.032	.724 $\pm$.149	.05136$\pm$.00143	4.94$\pm$.42	24.577 $\pm$.091
gridfm-graphkit	Raw	.00892 $\pm$.00154	.385 $\pm$.042	.905 $\pm$.030	.811 $\pm$.045	.09282 $\pm$.01763	19.82 $\pm$ 8.65	5.917 $\pm$.019
	C	.00753 $\pm$.00118	.367$\pm$.042	.905 $\pm$.030	.811 $\pm$.045	.07685 $\pm$.01307	18.88 $\pm$ 8.69	5.730 $\pm$.004
	P_{16}	.00617 $\pm$.00099	.503 $\pm$.018	.959$\pm$.008	.801 $\pm$.044	.04593 $\pm$.00156	5.36 $\pm$.56	5.723 $\pm$.011
	CSP_{16}	.00606$\pm$.00090	.496 $\pm$.021	.959$\pm$.008	.801 $\pm$.044	.04571$\pm$.00135	5.08$\pm$.18	5.779 $\pm$.012
LUMINA	Raw	.02653 $\pm$.00031	2.703 $\pm$.313	.443 $\pm$.014	.346 $\pm$.012	1.29896 $\pm$.85056	664.19 $\pm$ 448.00	7.548 $\pm$ 0.369
	C	.01780 $\pm$.00020	1.442 $\pm$.060	.443 $\pm$.014	.346 $\pm$.012	.21432 $\pm$.01186	20.65 $\pm$ 15.82	7.380 $\pm$ 0.155
	P_{16}	.01898 $\pm$.00049	2.160 $\pm$.268	.906$\pm$.012	.338 $\pm$.012	.07641 $\pm$.00128	5.47 $\pm$.56	7.981 $\pm$ 0.868
	CSP_{16}	.01570$\pm$.00024	1.282$\pm$.058	.906$\pm$.012	.338 $\pm$.012	.06662$\pm$.00074	4.61$\pm$.19	7.891 $\pm$ 0.567

and 10.3/2.7/39.6/31.7% for angle, for PIGNN-GC/GridSFM/gridfm-graphkit/LUMINA, respectively. After accounting for differences in total error norms, scenarios with larger off-subspace fractions have higher Mean PB, with partial Spearman correlations of .659/.635/.594/.833 for magnitude and .789/.906/.809/.860 for angle (all 95% CIs exclude zero).

Fixed-norm intervention. At fixed error norm, suppressing off-subspace error consistently lowers Relative Mean PB across all four surrogates, whereas suppressing in-subspace error has the opposite effect (Fig. 2), isolating error orientation from error magnitude. We test this with an oracle diagnostic using the known reference state, varying the voltage-magnitude error orientation at fixed ℓ_2 norm:

$$\tilde{e}(\alpha) = \|e_V\|_2 \frac{e_V^\parallel + \alpha e_V^\perp}{\|e_V^\parallel + \alpha e_V^\perp\|_2}, \quad (7)$$

where phase angles remain at their calibrated predictions. Decreasing α suppresses off-subspace error; the control analogously suppresses the in-subspace component by exchanging $e_V^\parallel$ and $e_V^\perp$.

Jacobian sensitivity. Off-subspace directions are substantially more residual-sensitive than in-subspace directions under the local AC Jacobian, consistent with the fixed-norm intervention. Using directional gain $g(d) = \|Jd\|_2/\|d\|_2$, 32 random direction pairs at each of 128 held-out operating points yield a median paired off-/in-subspace gain ratio of 1,563 $\times$. The same gain contrast appears in the actual calibrated prediction errors: the ratios of median off- to in-subspace gains are 22.81 $\times$, 575.83 $\times$, 4.69 $\times$, and 6.51 $\times$ for PIGNN-GC, GridSFM, gridfm-graphkit, and LUMINA, respectively. The observed error regime is consistent with the first-order expansion in Eq. (3), with median linearized-to-nonlinear residual-change ratios between 0.996 and 1.009 across all four models.

3.2. Solution subspace specificity.

Our matched-basis comparison shows that preserving the same amount of reference-state variation is insufficient for low PB, indicating that the specific state-space directions retained by the basis

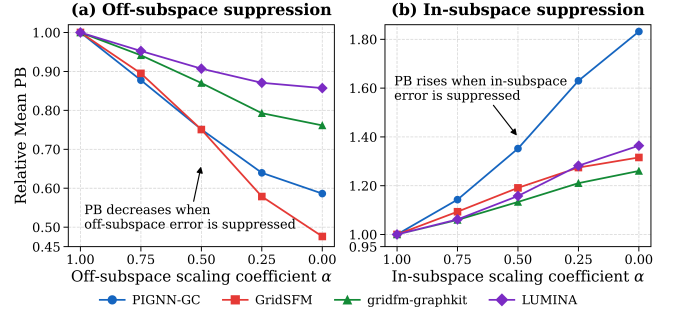


Fig. 2. Fixed-norm directional intervention on GBnetwork, with Relative Mean PB normalized at $\alpha = 1$.

Table 2. Matched basis comparison on GBnetwork for PIGNN-GC. Parentheses denote V/θ ranks or smoothing strengths.

Basis	Ref. var. retained (%)	Ref. PB	Pred. PB
SVD (16/16)	98.789/99.683	.0415	.0569
Laplacian matched (1750/2122)	98.819/99.688	26.15	26.22
Spatial matched ($\beta = .357/.208$)	98.801/99.684	6.20	6.22

are important. We compare SVD₁₆ with graph-Laplacian and spatial low-pass controls, each tuned using training references to match the V/θ variation retained by SVD₁₆. On the test set, all three preserve nearly identical reference variation (Table 2), yet the alternatives yield substantially larger PB for both projected references and corrected predictions.

3.3. Post-training CSP correction

Table 1 compares Raw, C , P_{16} , and CSP_{16} across the four frozen surrogates. Relative to C , CSP reduces Mean PB by 67.0%, 37.8%,

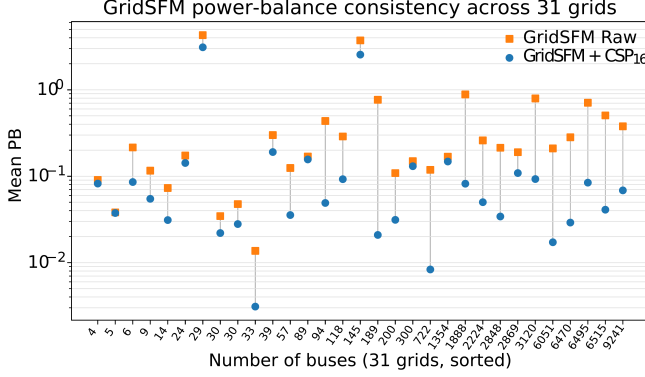


Fig. 3. Grid-by-grid physical consistency of GridSFM before and after CSP₁₆ across 31 systems ordered by bus count. Mean PB is shown on a logarithmic scale.

40.5%, and 68.9%, respectively, while lowering voltage-magnitude RMSE in all four models. Projection improves R_w^2 with little change in a_w ; angle accuracy is model dependent. Raw and CSP16 inference times are comparable.

Magnitude-only, angle-only, and full projection show model-dependent effects of angle projection. PIGNN-GC achieves the lowest PB with full CSP, whereas GridSFM performs best with magnitude-only projection. For gridfm-graphkit, adding angle projection reduces Mean/Max PB from .0612/18.23 to .0457/5.08, but degrades angle RMSE from .367° to .496°. For LUMINA, it reduces Mean/Max PB from .1939/38.53 to .0660/4.51 while improving angle RMSE from 1.374° to 1.215°. Thus, better physical consistency from angle projection need not imply better angle accuracy.

These model-dependent accuracy changes are explained by Eq. (5). For magnitude, the removed off-subspace prediction-error energy exceeds the discarded reference variation across all four models. For angle, the balance differs across models: in gridfm-graphkit, angle projection discards more valid reference variation than model error and therefore worsens angle RMSE, whereas the opposite holds for LUMINA.

3.4. Complementarity with one-step Newton refinement.

CSP provides a consistently better initialization for one-step Newton refinement, improving both state accuracy and post-update physical consistency across all four models. Starting from C or CSP₁₆, we apply one Newton update (NR₁) to non-slack angles and PQ voltage magnitudes, with validation-selected damping factor; $\eta = 1$ is selected for all models, with no linear-solve failures. Relative to C initialization, CSP lowers magnitude and angle RMSE in all models and reduces post-update Mean PB by 21.1–96.4% (Table 3).

The benefit is particularly pronounced for LUMINA: $C + \text{NR}_1$ shows a large one-seed overshoot despite a successful linear solve, whereas CSP₁₆ initialization remains consistent across seeds.

3.5. Cross-grid breadth and distribution shift

Cross-grid breadth. CSP reduces GridSFM Mean PB across all 31 systems (Fig. 3), spanning 4–9,241 buses and sub-kV to 750-kV levels. The set includes 28 pandapower/PYPOWER-format systems from IEEE, PEGASE, RTE, and GB benchmark families, plus three Common Information Model (CIM)-based networks. For each system, calibration and subspace statistics are fitted from its training data,

Table 3. One Newton update from calibrated or CSP₁₆ outputs on GBnetwork. V RMSE and Mean PB are reported in 10^{-3} . Values are means $\pm$ standard deviations over three model seeds.

Model	Start	V RMSE (10^{-3})	θ RMSE ($^\circ$)	Mean PB (10^{-3})
PIGNN-GC	C	$1.80 \pm .12$	$.119 \pm .013$	$2.84 \pm .46$
	CSP ₁₆	$1.17 \pm .10$	$.073 \pm .009$	$1.48 \pm .13$
GridSFM	C	11.86 ± 6.77	$.681 \pm .380$	2.50 ± 1.54
	CSP ₁₆	$.97 \pm .20$	$.059 \pm .013$	$1.22 \pm .29$
gridfm-graphkit	C	$.943 \pm .232$	$.0843 \pm .0236$	$.941 \pm .325$
	CSP ₁₆	$.496 \pm .085$	$.0277 \pm .0041$	$.742 \pm .080$
LUMINA	C	13.71 ± 9.23	$.671 \pm .038$	46.08 ± 64.15
	CSP ₁₆	$1.040 \pm .016$	$.0802 \pm .0031$	$1.653 \pm .036$

Table 4. PIGNN-GC with fixed-rank CSP₁₆ under topology and operating-distribution shifts. T transfers the control calibration and basis, whereas R refits both on condition-matched training data.

Condition	V RMSE	$R_w^2 \uparrow$	a_w	Mean	Med.	p95	p99	Max
Control	.01281	.8476	.4866	.0581	.0131	.2665	.5462	3.58
$N-1$ (T)	.01314	.8391	.4750	.0603	.0131	.2713	.5632	33.16
$N-2$ (T)	.01288	.8458	.4870	.0617	.0132	.2713	.5721	27.45
GBcorr (T)	.01966	.4819	.4025	.1003	.0231	.4512	.9746	7.00
GBcorr (R)	.01620	.5150	.4247	.0680	.0174	.3031	.6415	4.99

using nominal $k = 16$ per channel and the available SVD components on smaller systems. PV-voltage setpoints vary for most systems, but are fixed for GBnetwork with conflicting controllers.

Topology and spatial shifts. CSP remains stable in most statistics under $N-1/N-2$ topology shifts, with degradation concentrated in isolated worst-case violations, whereas GBcorr causes a broader performance drop. Under the topology shifts, voltage accuracy, tracking, and bulk PB remain close to control, while larger maxima motivate worst-case assessment beyond average error. GBcorr applies correlated, spatially varying load scaling across 14 zones while preserving total demand and generation settings. Transferring both calibration and basis gives Mean PB of 0.1003 p.u.; refitting calibration alone reduces this to 0.0944, basis alone to 0.0740, and both to 0.068. Although the transferred/refitted bases retain 93.55%/98.53% of GBcorr magnitude variation, projecting converged GBcorr references yields Mean PBs of 0.0868/0.0438 p.u., respectively. This reference-only diagnostic isolates the effect of basis transfer from surrogate error and helps explain the improvement from basis refitting.

4. CONCLUSION

This work identifies output-error geometry as an important factor in the gap between voltage accuracy and physical consistency in neural AC power-flow surrogates. Across four architecturally diverse neural surrogates, we show that small voltage errors can contain off-subspace components that induce large AC power-balance violations. Building on this insight, calibrated solution-subspace projection suppresses these components, reducing voltage-magnitude error and power-balance residuals without retraining; the residual reduction extends to 30 additional grids. These results suggest that error direction, not only magnitude, should inform surrogate evaluation and design. They also motivate extending physics-consistent neural surrogates toward broader power-system analysis, estimation, and optimization.

5. REFERENCES

- [1] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, “MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
- [2] Timon Conrad, Marco Pröll, David Riebesel, and Johann Jäger, “Framework for congestion management using particle swarm optimisation,” in *Proc. 61st Int. Universities Power Engineering Conf. (UPEC)*, 2026.
- [3] V. Bolz, J. Rueß, and A. Zell, “Data-driven graph convolutional neural networks for power system contingency analysis,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2023, pp. 1–5.
- [4] S. M. Frank, I. Steponavice, and S. Rebennack, “Optimal power flow: A bibliographic survey I: Formulations and deterministic methods,” *Energy Systems*, vol. 3, no. 3, pp. 221–258, 2012.
- [5] D. Owerko, F. Gama, and A. Ribeiro, “Optimal power flow using graph neural networks,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2020, pp. 5930–5934.
- [6] G. Cheng, Y. Lin, A. Abur, A. Gómez-Expósito, and W. Wu, “A survey of power system state estimation using multiple data sources: PMUs, SCADA, AMI, and beyond,” *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 1129–1151, 2024.
- [7] B. Donon, R. Clément, B. Donnot, A. Marot, I. Guyon, and M. Schoenauer, “Neural networks for power flow: Graph neural solver,” *Electric Power Systems Research*, vol. 189, pp. 106547, 2020.
- [8] N. Lin, S. Orfanoudakis, N. Ordonez Cardenas, J. S. Giraldo, and P. P. Vergara, “PowerFlowNet: Power flow approximation using message passing graph neural networks,” *International Journal of Electrical Power & Energy Systems*, vol. 160, pp. 110112, 2024.
- [9] D. Owerko, F. Gama, and A. Ribeiro, “Unsupervised optimal power flow using graph neural networks,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 6885–6889.
- [10] B. Jiang, C. Qin, and Q. Wang, “An unsupervised physics-informed neural network method for AC power flow calculations,” *IEEE Transactions on Power Systems*, vol. 40, no. 5, pp. 4407–4410, 2025.
- [11] T. Wen, W. Wang, J. Chen, and Z. Wang, “Power flow surrogate for power systems with high renewable penetration via a physics-informed graph attention network,” *Energies*, vol. 19, no. 13, pp. 2972, 2026.
- [12] H. T. Nguyen and P. L. Donti, “FSNet: Feasibility-seeking neural network for constrained optimization with guarantees,” in *Advances in Neural Information Processing Systems*, 2025, vol. 38, pp. 44366–44404.
- [13] P. L. Donti, D. Rolnick, and J. Z. Kolter, “DC3: A learning method for optimization with hard constraints,” in *International Conference on Learning Representations*, 2021.
- [14] M. Chzhen and P. L. Donti, “Improving feasibility via fast autoencoder-based projections,” in *International Conference on Learning Representations (ICLR)*, 2026.
- [15] W. Yang, A. Britto Mattos Lima, T. V. Spina, S. Fowers, B. Zhang, and C. White, “GridSFM: A foundation model for AC optimal power flow,” May 2026, Microsoft Research technical white paper.
- [16] GridFM Team, “gridfm-graphkit,” Python package, version 0.9.0, 2026.
- [17] A. Puech et al., “GENCO—a unified neural solver embedded in a development framework for steady-state grid analysis,” *arXiv preprint arXiv:2608.09921*, 2026.
- [18] Y. Li et al., “LUMINA: Foundation models for topology transferable ACOFF,” in *ICLR 2026 Workshop on Foundation Models for Science: Real-World Impact and Science-First Design*, 2026, arXiv:2603.04300.
- [19] S. de Jongh, F. Gielnik, F. Mueller, L. Schmit, M. Suriyah, and T. Leibfried, “Physics-informed geometric deep learning for inference tasks in power systems,” *Electric Power Systems Research*, vol. 211, pp. 108362, 2022.
- [20] R. Karim et al., “Parameter-efficient domain adaptation of physics-informed self-attention based GNNs for AC power flow prediction,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2026, pp. 1941–1945.
- [21] A. Ma, S. Ghamizi, A. Bojchevski, P. Zhang, and J. Cao, “Safe-PowerGraph: A safety-aware benchmarking framework for scalable graph neural networks in power flow and optimal power flow,” *Energy and AI*, vol. 25, pp. 100819, 2026.
- [22] M. Shamseldin, “A hybrid GNN-LSE method for fast, robust, and physically-consistent AC power flow,” *Electric Power Systems Research*, vol. 252, pp. 112380, 2026.
- [23] R. Nellikkath and S. Chatzivasilieiadis, “Physics-informed neural networks for AC optimal power flow,” *Electric Power Systems Research*, vol. 212, pp. 108412, Nov. 2022.
- [24] X. Hu, H. Hu, S. Verma, and Z.-L. Zhang, “Physics-guided deep neural networks for power flow analysis,” *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2082–2092, May 2021.
- [25] C. Kim et al., “Physics-informed GNN for medium-high voltage AC power flow with edge-aware attention and line search correction operator,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2026.
- [26] Shourya Bose, Helgi Hilmarsson, and Dhruv Suri, “Newton’s lantern: A reinforcement learning framework for fine-tuning ac power flow warm start models,” *arXiv preprint arXiv:2605.11102*, 2026.
- [27] A. Goodwin, J. Maack, and D. Sigler, “Power flow geometry and approximation,” *IEEE Transactions on Power Systems*, vol. 41, no. 2, pp. 982–993, 2026.
- [28] S. Park, W. Chen, T. W. K. Mak, and P. Van Hentenryck, “Compact optimization learning for AC optimal power flow,” *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4350–4359, Mar. 2024.
- [29] Samuel Chevalier, Luca Schenato, and Luca Daniel, “Accelerated probabilistic power flow in electrical distribution networks via model order reduction and neumann series expansion,” *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2151–2163, 2022.
- [30] L. Thurner et al., “pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, Nov. 2018.