The Subtle Art of Problem Formulation: Fusing Physics into Graph Structures for GNN-Based Pressure Monitoring in Water Distribution Networks

Amin E. Bakhshipour
Department of Environmental and Civil Engineering
University of Kaiserslautern-Landau (RPTU)
67663 Kaiserslautern, Germany
amin.bakhshipour@rptu.de

Mahta Bakhshizadeh
German Research Center for Artificial Intelligence (DFKI)
Smart Data and Knowledge Services Department
67663 Kaiserslautern, Germany
mahta.bakhshizadeh@dfki.de

Sven O. Krumke
Optimization Department
University of Kaiserslautern-Landau (RPTU)
67663 Kaiserslautern, Germany
krumke@rptu.de

Ulrich Dittmer
Department of Environmental and Civil Engineering
University of Kaiserslautern-Landau (RPTU)
67663 Kaiserslautern, Germany
ulrich.dittmer@rptu.de

Ali Haghighi
Department of Environmental and Civil Engineering
University of Kaiserslautern-Landau (RPTU)
67663 Kaiserslautern, Germany
ali.haghighi@rptu.de

Abstract

Water Distribution Networks (WDNs) are vital for urban sustainability, but monitoring them is challenging due to sparse sensors and complex hydraulics. While traditional models handle forward problems well, reconstructing network states from limited measurements (the inverse problem) is less developed. Current machine learning methods often overlook physical laws. We introduce a novel framework that integrates hydraulic principles directly into the graph structure of a Graph Neural Network (GNN), enabling accurate reconstruction of nodal pressures from sparse data. Our approach includes a Dirichlet-based method for generating realistic demand and leakage scenarios and a new benchmarking protocol tailored to operational needs. Experiments on standard benchmarks demonstrate superior performance with fewer sensors compared to conventional

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

GNNs, and our evaluation framework links model errors to network characteristics, providing actionable insights for deployment. This physics-infused GNN paradigm advances WDN monitoring and leak detection.

1 Introduction

WDNs are critical urban infrastructures that supply clean water through interconnected pipes, pumps, and valves [1]. Essential for public health and urban resilience, WDNs face challenges from aging infrastructure, water scarcity, and rising maintenance costs [2]. Efficient monitoring for anomaly detection (e.g., leaks) and maintenance is vital but difficult due to limited sensor deployment, complicating real-time system understanding.

Operational tasks like pressure monitoring and leak detection rely on hydraulic modeling. Traditional models like EPANET [3] solve the forward problem, computing pressures from known inputs, but practical monitoring requires solving the inverse problem: inferring network states from sparse measurements. This underdetermined problem challenges traditional optimization-based methods, which are computationally intensive, require manual recalibration, and are sensitive to noise [4]. GNNs leverage WDNs' graph structure for modeling [5], but Physics-Informed Neural Networks (PINNs) focus on forward problems and face computational inefficiencies [6]. A gap exists in applying physics-informed methods to the inverse problem efficiently.

We propose a physics-infused GNN framework that embeds hydraulic laws into the graph structure, enabling accurate pressure reconstruction from sparse sensors. Our contributions include:

- A graph structure embedding hydraulic principles for GNNs.
- A method to transform heterogeneous WDN graphs for GNN compatibility.
- A benchmark suite for evaluating pressure reconstruction with sparse sensors.
- Superior performance over conventional GNNs with fewer sensors.
- A problem formulation aligned with operational needs for practical deployment.

Within the physics-driven machine learning literature, the term *physics-informed* typically refers to neural models that enforce physical constraints by introducing governing equations—such as conservation laws—directly in the loss function [7, 6]. In contrast, our use of *physics-infused* specifically denotes the embedding of hydraulic principles into the *graph structure and feature space* itself, rather than through the learning objective. This allows the GNN to encode network physics natively in its topology and message passing, distinguishing our approach from existing physics-informed GNNs and preventing ambiguity in terminology.

This work advances WDN monitoring and supports sustainable urban infrastructure management.

2 Related Work

GNN applications in WDNs address two key problems:

1- State estimation involves computing complete hydraulic states from known inputs. Traditional hydraulic models require calibration and are computationally intensive. GNN-based metamodels have emerged as faster alternatives [8, 9, 10], with physics-informed approaches showing particular promise. Physics-informed GNNs embed hydraulic laws into learning, achieving faster computation than simulators like EPANET while maintaining accuracy and robustness to varying demands [6].

2- State prediction (inverse problem) infers network states from sparse measurements, a critical challenge for urban water monitoring. Traditional optimization-based methods face computational bottlenecks and calibration issues. Recent GNN approaches include pressure reconstruction [11, 12], leak detection [13, 14], and sensor placement [15, 16]. However, among few studies that have explored GNN-based pressure reconstruction, the explicit integration of governing physical laws remains largely absent. Furthermore, while robust benchmarks exist for state estimation (forward problem), appropriate benchmarking for the inverse problem is still lacking.

3 Methodology

WDNs are modeled as undirected graphs $G = (V, E)$, where nodes V represent junctions, reservoirs, and tanks, and edges E represent pipes with hydraulic attributes (e.g., length, diameter, roughness) [6]. The physical behavior of WDNs is governed by conservation of mass and energy, ensuring consistent pressure and flow dynamics across the network.

Problem Definition: We address the inverse problem of reconstructing nodal pressures in a WDN from sparse sensor measurements, given known reservoir pressures and pipe attributes, but unknown consumer demands. This task is crucial for operational applications like leak detection and network control under limited sensor coverage. Our approach introduces three key innovations:

- **Realistic Demand and Leakage Modeling:** Unlike prior works using simplistic Gaussian noise, we model demand and leakage under night flow conditions, where leaks dominate due to low consumer demand, using a Dirichlet-based scenario generator to simulate diverse, realistic patterns (Appendix B).
- **Systematic Sensor Placement:** We propose the Conceptual Ranking Method (CRM) for efficient sensor placement, prioritizing nodes based on their sensitivity to leaks and network dynamics, outperforming random or optimization-based methods (Appendix A).
- **Physics-Infused Graph Structure:** We embed hydraulic laws directly into the graph topology and edge weights, enabling a GNN to learn physically consistent pressure relationships, enhancing accuracy and interpretability.

3.1 Physics-Infused Graph Construction

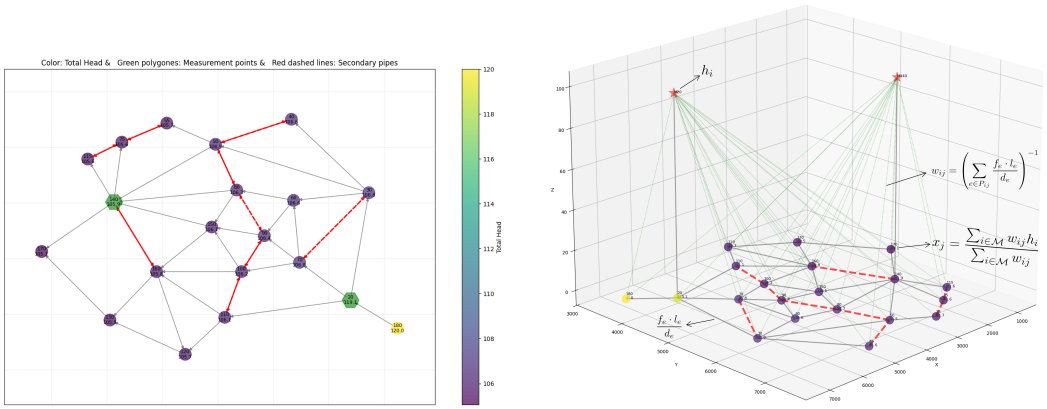
Our novel graph construction pipeline, illustrated in Figure 1 using the Kadu benchmark, transforms a WDN into a physics-infused representation through five steps:

1. **Network Import:** We import the WDN model using WNTR [17], computing edge weights as hydraulic resistance $f \cdot L/D$, where f is the friction factor, L is pipe length, and D is diameter. This encoding ensures that the graph represents the hydraulic resistance of each pipe, directly linking the edge attributes to the physical principles governing pressure loss.
2. **Directed Graph Construction:** We convert the undirected graph G into a directed graph $G' = (V, E')$ by simulating diverse demand scenarios (e.g., 1,000 realizations). Primary pipes, with consistent flow direction, are assigned single directed edges, while secondary pipes, with variable flow, are represented by bidirectional edges (Figure 1a).
3. **Sensor Placement via CRM:** Our CRM ranks nodes by their sensitivity to leak magnitude and hydraulic changes, ensuring uniform sensor distribution for robust monitoring (Appendix A). This computationally efficient method enhances diagnostic capability compared to prior approaches [15].
4. **Measurement Integration:** We connect measurement points to all nodes via shortest hydraulic paths, with weights reflecting cumulative resistance, propagating sensor data effectively across the network (Figure 1b).
5. **Node Feature Assignment:** Sensor nodes are assigned observed pressures, while unmeasured nodes receive features as weighted averages of sensor pressures, computed using hydraulic path weights. By incorporating a pressure feature that captures both the physical relationships between nodes and the influence of sensor locations, our method provides the GNN with richer, physics-informed input. Importantly, this formulation transforms the original heterogeneous graph-with distinct node types (consumer junctions and sensors) and edge types (pipes, weighted paths)-into a homogeneous graph structure. This conversion eliminates the need for specialized heterogeneous message passing layers, which can complicate the implementation of edge weights and increase model complexity [14].

3.2 GNN for Pressure Reconstruction

We develop a Graph Convolutional Network (GCN) operating on the physics-infused graph G'. The model uses three GCN layers with residual connections and batch normalization to predict nodal pressures, leveraging the embedded hydraulic laws for accuracy. We train with a combined mean

squared error and Huber loss for robustness, using the Adam optimizer with a learning rate scheduler. This approach achieves superior pressure reconstruction with fewer sensors, supporting operational tasks like leak detection.



(a) (b)

Figure 1: The data generation and graph construction pipeline. (a) Primary and secondary (red) pipes in the benchmark network. (b) 3D visualization of measurement points and connections.

4 Experiments

We evaluate our physics-infused GNN framework across eight benchmark WDNs: Alpervoits, Anytown, Baghmalek, BAK, Hanoi, Kadu, Modena, and Zhi Jiang. These span diverse scales for comprehensive assessment. Details on network structures and sensor placements are in Appendix D. Datasets are available in Supplementary Materials.

4.1 Data Generation and Baseline

Data generation emphasizes midnight flow conditions, where flat gradients challenge inverse reconstruction but enhance leak detectability. We use Dirichlet-based allocation for realistic, simultaneous anomalies across junctions (details in Appendix B). Each network yields 10,000 scenarios (60/20/20 split). The baseline is a standard GNN without physics-infused structure, trained identically for fair comparison.

4.2 Evaluation Metrics and Benchmarking

We use regression metrics like MAE, NMAE, MAPE, and thresholded accuracy (formulas in Appendix C). Our task-oriented protocol specifies accuracy thresholds (e.g., MAE $<$ 0.5m, Acc@2% $>$ 90%) and reports minimal sensors needed, aligning with operational needs.

4.3 Results and Discussion

Table 1 and A.1 in Appendix E compare our PhyInGNN with a standard GNN across eight benchmark WDNs under varying sensor masking ratios. PhyInGNN consistently outperforms the baseline, achieving significantly lower MAE (e.g., 0.250 m vs. 6.440 m on Anytown at 90% masking) and higher Acc@2% (e.g., 0.986 vs. 0.274), critical for operational tasks like leak detection. This advantage persists at high masking ratios, demonstrating robustness due to physics-informed graph structures embedding hydraulic principles. Compared to prior GNN-based pressure reconstruction studies, PhyInGNN excels: it achieves MAPE of 0.1–2.0% with a single sensor, versus 5% at a 5% sensor ratio in [11], and surpasses 87% accuracy at 10% error in [12] with over 90% at a stricter 2% threshold [12]. Our task-oriented benchmarking protocol, requiring MAE $<$ 0.5 m and Acc@2% $>$ 90%, aligns with real-world needs, optimizing sensor placement for cost-effective monitoring.

Table 1: Performance comparison for eight benchmark WDNs.

Network	Model	Mask Ratio	MAE (m)	Acc02	Network	Model	Mask Ratio	MAE (m)	Acc02
Alprovits	GNN	71.4%	0.360	0.905	**Baghmalek**	GNN	97.3%	1.230	0.413
	PhyInGNN	71.4%	**0.070**	**1.000**		PhyInGNN	97.3%	**0.320**	**0.905**
Anytown	GNN	90.0%	6.440	0.274	**Kadu**	GNN	84.0%	0.650	0.950
	PhyInGNN	90.0%	**0.250**	**0.986**		PhyInGNN	84.0%	**0.110**	**0.999**
BAK	GNN	88.6%	0.440	0.962	**Modena**	GNN	97.0%	1.080	0.626
	PhyInGNN	88.6%	**0.230**	**0.984**		PhyInGNN	97.0%	**0.190**	**0.988**
Hanoi	GNN	90.6%	0.500	0.972	**ZhiJiang**	GNN	96.5%	1.080	0.626
	PhyInGNN	90.6%	**0.100**	**0.998**		PhyInGNN	96.5%	**0.190**	**0.988**

To assess the contribution of each architectural and methodological component of our physics-infused framework, we conducted a detailed **ablation study** on the Kadu benchmark with two optimally placed sensors. We systematically varied the inclusion of physics-based edge weights, residual connections, GNN depth, measurement site selection, batch normalization, and dropout. Results demonstrate that each element, especially the physics-based edge weights and sensor placement by CRM, measurably improves predictive accuracy and robustness in sparse-sensing regimes. Removing any of these components led to increases in MAE, NMAE, and a reduction in R^2. For example, omitting physics-based edge weights increased MAE from 0.14 to 0.17 m and reduced R^2 from 0.982 to 0.976. A full quantitative breakdown of all configurations is provided as Table A.2 in the Appendix.

To conclude, PhyInGNN enables efficient leak detection by identifying pressure anomalies, streamlining rehabilitation planning, and delivering significant cost savings for water authorities. Its framework, applicable to gas and electricity networks governed by Kirchhoff's laws, advances smart infrastructure management. Visual R^2 maps reveal error patterns tied to network topology, guiding sensor deployment (Appendix E). Future work includes real-world validation and incorporating additional physical laws such as conservation of mass.

5 Conclusion

This work introduced a novel physics-infused GNN framework for pressure monitoring in WDNs, addressing critical challenges at the intersection of hydraulic modeling, machine learning, and practical utility. By embedding physical laws directly into the graph structure and grounding the problem formulation in real-world operational requirements, our approach achieves superior accuracy with minimal sensor coverage compared to conventional GNN method. Extensive experiments on diverse benchmark networks demonstrate the robustness and scalability of our method, while the proposed benchmarking protocol ensures that model evaluation remains closely aligned with industry needs and downstream tasks. Beyond advancing the state of the art in WDN monitoring, our findings highlight the importance of problem-driven research and the integration of domain knowledge into machine learning architectures, paving the way for more reliable, interpretable, and impactful solutions in smart infrastructure management.

Acknowledgments and Disclosure of Funding

This project is funded by the Deutsche Forschungsgemeinschaft (DFG) under Project number 544048327.

References

[1] Robert Sitzenfrei. Using complex network analysis for water quality assessment in large water distribution systems. *Water Research*, 201:117359, 2021.

[2] Amin Minaei, Ali Haghighi, and Hamid Reza Ghafouri. Computer-aided decision-making model for multiphase upgrading of aged water distribution mains. *Journal of water resources planning and management*, 145(5):04019008, 2019.

[3] Lewis A Rossman. The epanet programmer's toolkit for analysis of water distribution systems. In *WRPMD'99: Preparing for the 21st Century*, pages 1–10. 1999.

[4] Ali Haghighi and Helena M Ramos. Detection of leakage freshwater and friction factor calibration in drinking networks using central force optimization. *Water resources management*, 26:2347–2363, 2012.

[5] Zepeng Zhang and Olga Fink. Algorithm-informed graph neural networks for leakage detection and localization in water distribution networks. *arXiv preprint arXiv:2408.02797*, 2024.

[6] Inaam Ashraf, Janine Strotherm, Luca Hermes, and Barbara Hammer. Physics-informed graph neural networks for water distribution systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 21905–21913, 2024.

[7] Salvatore Cuomo, Vittorio S Di Cola, Francesco Giampaolo, Gianluigi Rozza, Maziar Raissi, Francesco Piccialli, and George Em Karniadakis. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):1–56, 2022.

[8] Ariele Zanfei, Andrea Menapace, Bruno M Brentan, Robert Sitzenfrei, and Manuel Herrera. Shall we always use hydraulic models? a graph neural network metamodel for water system calibration and uncertainty assessment. *Water Research*, 242:120264, 2023.

[9] Bulat Kerimov, Riccardo Taormina, and Franz Tscheikner-Gratl. Towards transferable metamodels for water distribution systems with edge-based graph neural networks. *Water Research*, 261:121933, 2024.

[10] Lu Xing and Lina Sela. Graph neural networks for state estimation in water distribution systems: Application of supervised and semisupervised learning. *Journal of Water Resources Planning and Management*, 148(5):04022018, 2022.

[11] Gergely Hajgató, Bálint Gyires-Tóth, and György Paál. Reconstructing nodal pressures in water distribution systems with graph neural networks. *arXiv preprint arXiv:2104.13619*, 2021.

[12] Huy Truong, Andrés Tello, Alexander Lazovik, and Victoria Degeler. Graph neural networks for pressure estimation in water distribution systems. *Water Resources Research*, 60(7):e2023WR036741, 2024.

[13] Ariele Zanfei, Andrea Menapace, Bruno M Brentan, Maurizio Righetti, and Manuel Herrera. Novel approach for burst detection in water distribution systems based on graph neural networks. *Sustainable Cities and Society*, 86:104090, 2022.

[14] Melanie Schaller, Michael Steininger, Andrzej Dulny, Daniel Schlör, and Andreas Hotho. Liquor-hgnn: A heterogeneous graph neural network for leakage detection in water distribution networks. In *LWDA*, pages 454–469, 2023.

[15] Weidong Zhang, Xiaoping Yang, and Juan Li. Sensor placement for leak localization in water distribution networks based on graph convolutional network. *IEEE sensors journal*, 22(21):21093–21100, 2022.

[16] Andrea Menapace, Ariele Zanfei, Manuel Herrera, and Bruno Brentan. Graph neural networks for sensor placement: A proof of concept towards a digital twin of water distribution systems. *Water*, 16(13):1835, 2024.

[17] Katherine A Klise, David Hart, Dylan Michael Moriarty, Michael Lee Bynum, Regan Murray, Jonathan Burkhardt, and Terra Haxton. Water network tool for resilience (wntr) user manual. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017.

Appendix

A Conceptual Ranking Method (CRM) for Measurement Site Selection

The Conceptual Ranking Method (CRM) is a computationally efficient approach for optimal sensor placement in water distribution networks. CRM prioritizes candidate measurement nodes based on their sensitivity ranks to key decision variables (e.g., leak parameters, pipe friction factors), rather than absolute sensitivity values. This rank-based approach mitigates the influence of extreme sensitivities and promotes a uniform spatial distribution of sensors.

Pseudocode for CRM:

1. Compute the sensitivity matrices S^a (to leak areas) and S^f (to pipe friction factors) for all candidate nodes.
2. For each row (parameter) in S^a and S^f, rank the candidate nodes by sensitivity (most sensitive gets rank 1).
3. For each node, sum its ranks across all parameters to obtain Sum_a and Sum_f.
4. Select the node with the lowest total rank (highest priority) as the next measurement site.
5. Remove all rows (parameters) from the rank matrices for which this node had rank 1 (i.e., is most sensitive).
6. If any parameters remain uncovered, repeat steps 3–5; otherwise, stop.
7. Return the ordered list of selected measurement sites.

CRM Pseudocode (compact form):

```
Input: Sensitivity matrices S^a (leak), S^f (friction), candidate nodes
Output: Ordered list of measurement sites

1. For each parameter in S^a and S^f, rank nodes by sensitivity (rank 1 = most sensitive)
2. While uncovered parameters remain:
    a. For each node, sum ranks across all parameters (Sum_a, Sum_f)
    b. Select node j* with lowest total rank
    c. Add j* to measurement site list
    d. Remove parameters for which node j* has rank 1
3. Return ordered measurement site list
```

Notes:

- CRM avoids mathematical programming or iterative optimization, reducing computational complexity.
- For frequency-domain analysis, CRM can be adapted to rank nodes by sensitivity in the frequency response diagram.
- Nodes selected earlier have higher diagnostic priority.

B Mathematical Details of Dirichlet-Based Scenario Generation

Our data generation process is designed to rigorously benchmark inverse pressure reconstruction in WDNs, with a particular focus on the operationally critical midnight flow regime. During these low-demand hours, legitimate consumption is minimal and system pressure is high, creating quasi-steady-state conditions that amplify the hydraulic signature of leaks. However, this regime also introduces significant numerical challenges: the hydraulic gradient across the network flattens, making pressure less sensitive to local withdrawals and rendering the inverse problem highly ill-posed. As a result, traditional solvers often converge poorly or yield non-unique solutions, especially in large-scale or looped networks. In contrast, peak demand periods feature more pronounced hydraulic gradients and improved pressure observability, but the dynamic and heterogeneous nature of legitimate consumption blurs the distinction between leaks and normal usage. This trade-off underscores the strategic importance of midnight flow-based analysis for robust and interpretable leakage detection.

To realistically capture the unpredictability of leaks and demand anomalies under these challenging conditions, we employ a principled scenario generation approach based on the Dirichlet distribution. For each scenario, the total extra demand, representing the aggregate of all possible leaks and anomalies, is allocated across all junctions using a uniform Dirichlet distribution. This enables simulation of multiple, simultaneous anomalies of arbitrary size and location, exposing the model to a broad and physically plausible range of hydraulic conditions.

Statistical testing confirms the uniformity and robustness of this allocation. For each benchmark network, we generate 10,000 scenarios, split into training (60%), validation (20%), and test (20%) sets. This domain-informed and physically grounded data generation strategy is essential for developing and evaluating inverse methods that are robust to the ill-posedness and operational realities of real-world WDNs.

Let n denote the number of junctions in the network. For each scenario, we allocate a total extra demand d_{total} across the n junctions. The allocation vector $\mathbf{p} = (p_1, \ldots, p_n)$ is sampled from a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = \mathbf{1}_n$:

$$\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_n)$$

The Dirichlet distribution $\text{Dirichlet}(\boldsymbol{\alpha})$ is defined over the $(n-1)$-dimensional simplex:

$$\Delta^{n-1} = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; x_i \geq 0,\ \sum_{i=1}^{n} x_i = 1 \right\}$$

with probability density function

$$f(p_1, \ldots, p_n; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{n} p_i^{\alpha_i - 1}$$

where $B(\boldsymbol{\alpha})$ is the multivariate Beta function. Setting $\alpha_i = 1$ for all i yields a uniform distribution over the simplex, ensuring that every possible partition of d_{total} among the junctions is equally likely.

The extra demand at each junction is then given by:

$$d_{\text{extra},i} = p_i \cdot d_{\text{total}}$$

To verify the uniformity of our demand generation, we performed statistical tests over n_{sim} scenarios. The average extra demand per junction approaches $\frac{d_{\text{total}}}{n}$, and the minimum and maximum demands reflect the expected spread of the Dirichlet distribution:

$$\text{avg\_demand} = \frac{d_{\text{total}}}{n}, \quad \text{min\_demand} = \min_i d_{\text{extra},i}, \quad \text{max\_demand} = \max_i d_{\text{extra},i}$$

As n_{sim} increases, the empirical distribution of extra demands converges to the theoretical uniform distribution over the simplex.

C Evaluation Metric Formulas

We report standard regression metrics including mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and R^2. In addition, we focus on several case-specific metrics that are particularly relevant for pressure monitoring in WDN:

Thresholded Accuracy. We report the proportion of nodes where the predicted pressure deviates from the true value by less than a specified percentage threshold. To facilitate both benchmarking and operational relevance, we consider two thresholds: 0.1 (enabling comparison with existing literature) and 0.05 (reflecting practical requirements for downstream tasks such as leak detection and control). This metric directly captures the operational reliability of the model under real-world accuracy constraints.

Maximum Percentage Error (MaxPE). MaxPE quantifies the largest relative error observed among all nodes in a scenario. This is especially important for leak scenarios, where pressures near leakage points can deviate substantially from normal conditions. Accurate prediction of these extremes is crucial for reliable leak detection and network safety.

Normalized Mean Absolute Error by Pressure Range (NMAE). To enable fair comparison across WDNs with varying pressure ranges and topographies, we introduce a normalized mean absolute error metric. NMAE computes the mean absolute error normalized by the range of predicted pressures (i.e., the difference between the maximum and minimum predicted pressure values). This scale-invariant metric is especially useful when comparing performance between flat networks (with little pressure variation) and networks with significant elevation differences.

Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Coefficient of determination ($\mathbf{R}^2$):

$$\text{R}^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

Thresholded Accuracy:

$$\text{Acc}(@\delta_{\text{thresh}}) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(|y_i - \hat{y}_i| \leq \delta_{\text{thresh}} \cdot |y_i| \right)$$

Maximum Percentage Error (MaxPE):

$$\text{MaxPE} = \max_i \left(\frac{|y_i - \hat{y}_i|}{|y_i|} \times 100 \right)$$

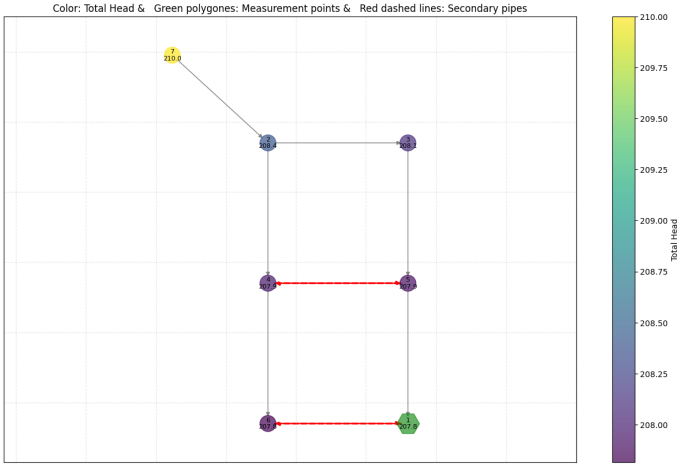
Normalized Mean Absolute Error by Pressure Range (NMAE):

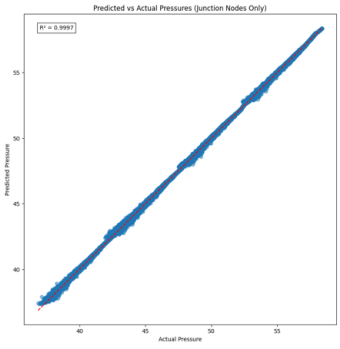
$$\text{NMAE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{\max_j \hat{y}_j - \min_j \hat{y}_j}$$

where N is the number of predictions, y_i and $\hat{y}_i$ are the true and predicted values, $\bar{y}$ is the mean of the true values, and $\mathbb{I}(\cdot)$ is the indicator function.

D Case Studies: Benchmark Networks, Sensor Placement, and Reconstruction Results

For each benchmark network, we present (a) the base graph topology with optimal sensor placement and (b) the spatial distribution of R^2 scores for nodal pressure reconstruction using our physics-infused GNN.

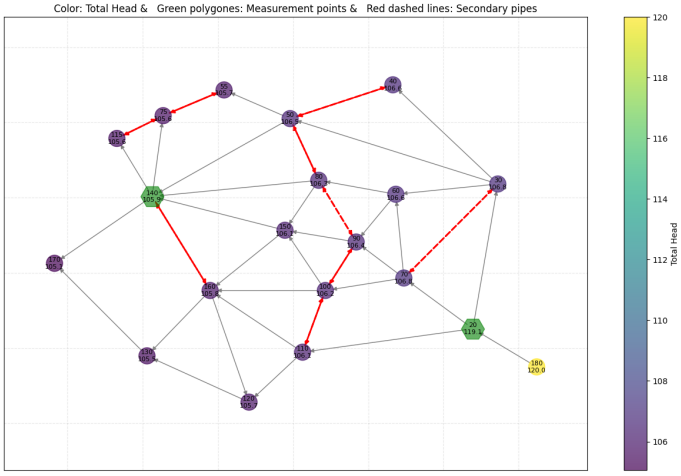


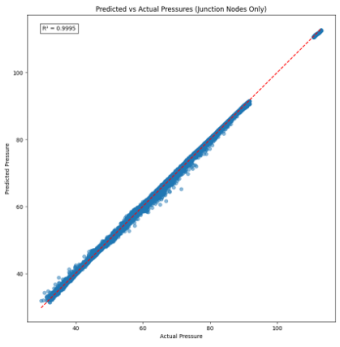


(a) Topology and optimal sensor placement

(b) Spatial R^2 map

Figure A.1: Alprovits benchmark

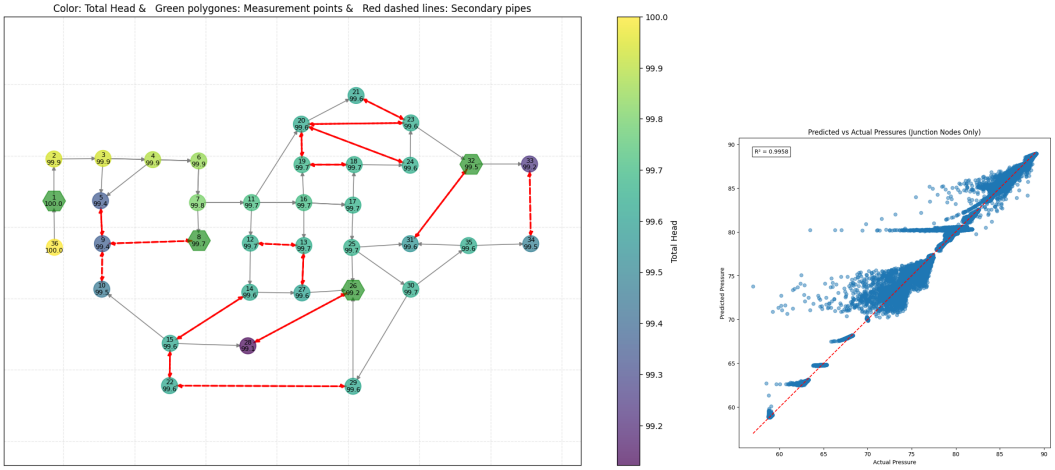




(a) Topology and optimal sensor placement

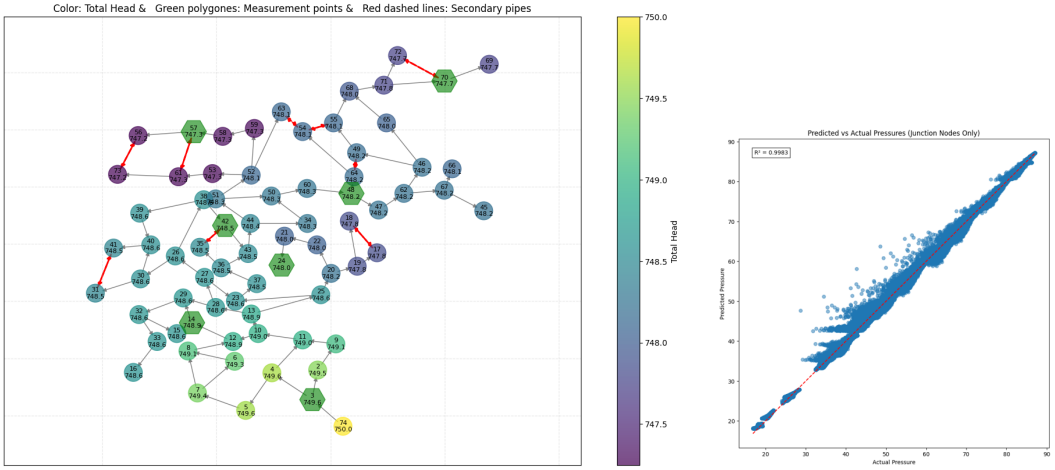
(b) Spatial R^2 map

Figure A.2: Anytown benchmark



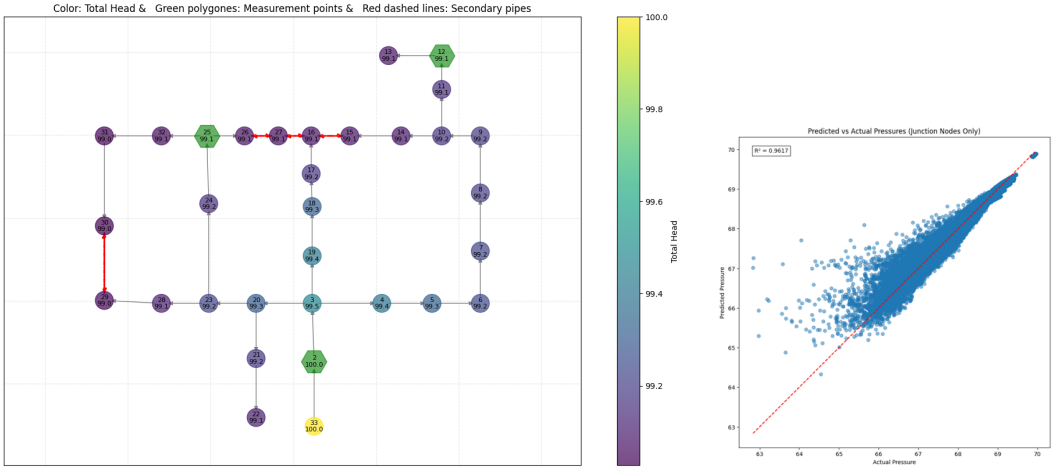
(a) Topology and optimal sensor placement (b) Spatial R^2 map

Figure A.3: BAK benchmark



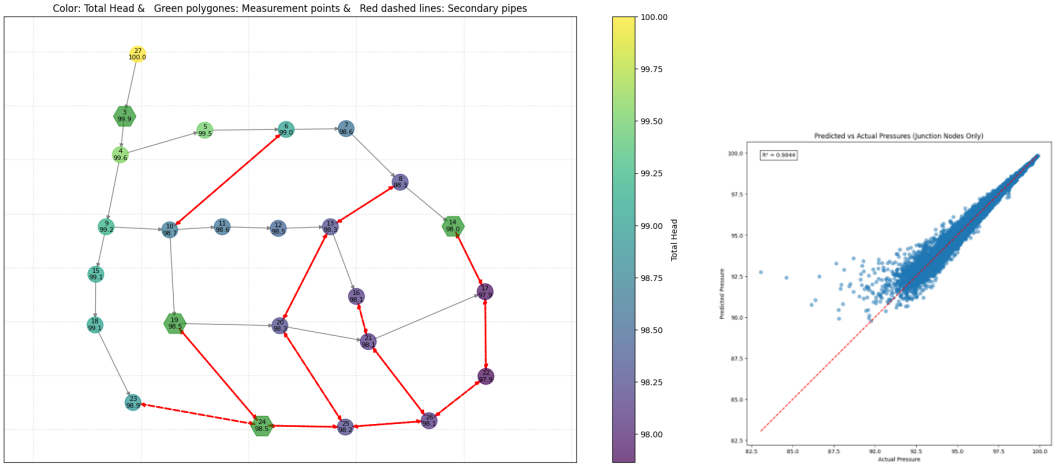
(a) Topology and optimal sensor placement (b) Spatial R^2 map

Figure A.4: Baghmalek benchmark



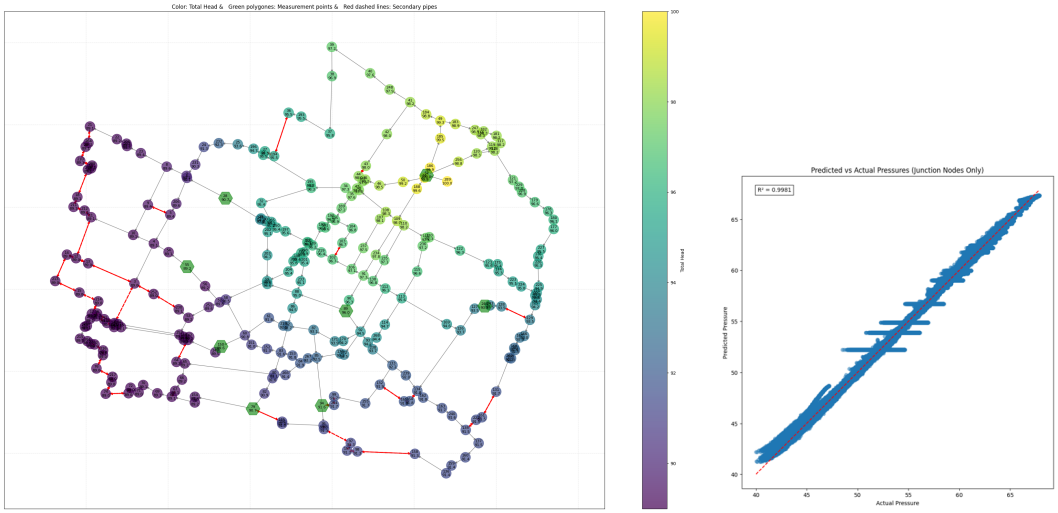
(a) Topology and optimal sensor placement (b) Spatial R^2 map

Figure A.5: Hanoi benchmark



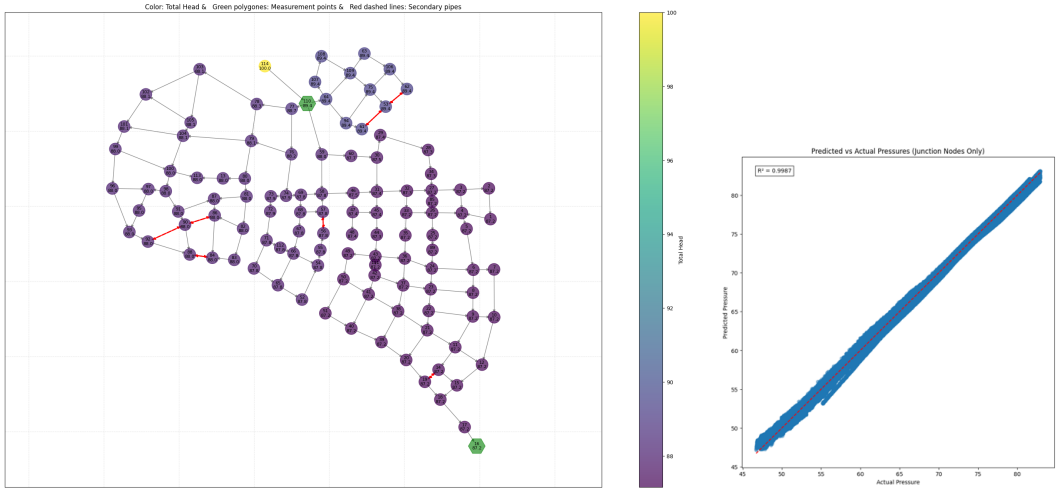
(a) Topology and optimal sensor placement (b) Spatial R^2 map

Figure A.6: Kadu benchmark



(a) Topology and optimal sensor placement (b) Spatial R^2 map

Figure A.7: Modena benchmark



(a) Topology and optimal sensor placement (b) Spatial R^2 map

Figure A.8: Zhijiang benchmark

E Comprehensive Benchmark Results

This appendix provides the complete quantitative results for all evaluated models across the eight benchmark water distribution networks. The table below reports detailed performance metrics for both the standard GNN and the proposed physics-infused GNN (PhyInGNN) under various sensor configurations.

F Ablation Studies

Interpretation. The ablation results demonstrate that each component of our framework is critical for robust and accurate pressure reconstruction under sparse sensing:

- **Physics-based edge weights:** Removing them increases MAE and reduces R^2, confirming the value of embedding hydraulic knowledge in the graph.
- **Residual connections and GNN depth:** Excluding residuals or using fewer layers significantly degrades performance, highlighting their necessity for stable and expressive learning.
- **Measurement site selection:** Randomizing sensor placement nearly doubles the error, underlining the importance of systematic, physically informed sensor selection.
- **Batch normalization and dropout:** Removing normalization or dropout leads to substantial performance drops, showing their role in generalization and training stability.

Overall, the full model achieves the best or near-best results across all metrics, validating the synergistic effect of combining physics-based graph construction, architectural innovations, and principled measurement design for the inverse problem in WDNs.

Table A.1: Full Performance comparison of standard GNN and PhyInGNN across eight benchmark WDNs. ("#Juncs" and "#Sens" denote the number of junctions and sensors, respectively.)

Network	Model	#Juncs	#Sens	Max Ratio	MSE	MAE (m)	RMSE	NMAE	MAPE	MaxPE	Acc10	Acc05	Acc02	R2	Actual Range	Predicted Range
Alprovits	**GNN**	7	1	0.857	1.2863	0.7721	1.1341	0.041	0.017	0.094	**1.000**	0.933	0.669	0.964	[36.94 , 58.40]	[40.40 , 57.84]
	GNN	7	2	0.714	0.4202	0.3611	0.6482	0.019	0.007	0.063	**1.000**	0.994	0.905	0.988	[36.94 , 58.40]	[37.44 , 57.97]
	PhyInGNN	7	1	0.857	0.0384	0.1482	0.1959	0.008	0.003	0.037	**1.000**	**1.000**	0.996	0.999	[36.94 , 58.40]	[38.15 , 58.37]
	PhyInGNN	7	2	0.714	**0.0117**	**0.0744**	**0.1081**	**0.004**	**0.002**	**0.020**	**1.000**	**1.000**	**1.000**	**1.000**	[36.94 , 58.40]	[37.35 , 58.37]
Anytown	**GNN**	20	1	0.950	94.9939	8.2041	9.7465	0.137	0.134	0.134	0.426	0.235	0.130	0.669	[29.41 , 113.11]	[54.38 , 111.96]
	GNN	20	2	0.900	70.8266	6.4367	8.4159	0.107	0.104	0.695	0.571	0.399	0.274	0.753	[29.41 , 113.11]	[35.42 , 111.25]
	PhyInGNN	20	1	0.950	2.4318	1.2725	1.5594	0.021	0.021	0.321	0.994	0.928	0.590	0.992	[29.41 , 113.11]	[36.03 , 113.27]
	PhyInGNN	20	2	0.900	**0.1313**	**0.2460**	**0.3623**	**0.004**	**0.004**	**0.117**	**1.000**	**0.999**	**0.986**	**1.000**	[29.41 , 113.11]	[31.38 , 112.60]
BAK	**GNN**	35	1	0.971	0.7496	0.4379	0.8658	0.009	0.006	0.379	0.998	0.992	0.961	0.989	[56.99 , 89.13]	[59.04 , 88.45]
	GNN	35	4	0.886	0.7385	0.4432	0.8593	0.009	0.006	0.759	0.998	0.992	0.962	0.990	[56.99 , 89.13]	[58.93 , 88.41]
	PhyInGNN	35	1	0.971	0.4640	0.3422	0.6812	0.008	0.004	0.229	**0.999**	0.994	0.971	0.993	[56.99 , 89.13]	[58.97 , 88.72]
	PhyInGNN	35	4	0.886	**0.2951**	**0.2257**	**0.5432**	**0.005**	**0.003**	**0.293**	**0.999**	**0.996**	**0.983**	**0.996**	[56.99 , 89.13]	[58.89 , 88.89]
Hanoi	**GNN**	32	1	0.969	0.5199	0.5869	0.7210	0.049	0.009	0.080	**1.000**	0.999	0.961	0.272	[62.84 , 69.96]	[67.78 , 69.87]
	GNN	32	3	0.906	0.4143	0.4972	0.6437	0.041	0.007	0.087	**1.000**	**1.000**	0.972	0.419	[62.84 , 69.96]	[66.63 , 69.92]
	PhyInGNN	32	1	0.969	0.0592	0.1652	0.2433	0.016	0.002	0.120	**1.000**	**1.000**	0.997	0.918	[62.84 , 69.96]	[66.05 , 69.87]
	PhyInGNN	32	3	0.906	**0.0269**	**0.0960**	**0.1640**	**0.010**	**0.001**	**0.070**	**1.000**	**1.000**	**0.998**	**0.962**	[62.84 , 69.96]	[64.33 , 69.89]
Baghmalek	**GNN**	73	1	0.986	2.4467	1.1970	1.5642	0.018	0.027	**0.662**	0.986	0.905	0.430	0.990	[16.36 , 87.12]	[18.43 , 83.42]
	GNN	73	2	0.973	2.4615	1.2255	1.5689	0.019	0.027	0.742	0.987	0.906	0.413	0.990	[16.36 , 87.12]	[18.55 , 83.16]
	PhyInGNN	73	1	0.986	0.3997	0.4134	0.6322	0.006	0.010	0.664	0.998	0.987	0.880	0.998	[16.36 , 87.12]	[18.14 , 87.21]
	PhyInGNN	73	2	0.973	**0.2729**	**0.3224**	**0.5224**	**0.005**	**0.008**	0.730	**0.999**	**0.992**	**0.905**	**0.999**	[16.36 , 87.12]	[17.69 , 86.57]
Kadu	**GNN**	25	1	0.960	1.6133	1.0021	1.2702	0.048	0.010	0.172	**1.000**	0.998	0.876	0.477	[83.06 , 99.87]	[95.21 , 99.70]
	GNN	25	4	0.840	0.8473	0.6496	0.9205	0.031	0.007	0.119	**1.000**	0.999	0.950	0.728	[83.06 , 99.87]	[91.10 , 99.67]
	PhyInGNN	25	1	0.960	0.1144	0.1994	0.3382	0.010	0.002	**0.074**	**1.000**	**1.000**	0.997	0.962	[83.06 , 99.87]	[92.19 , 99.70]
	PhyInGNN	25	4	0.840	**0.0452**	**0.1084**	**0.2127**	**0.005**	**0.001**	0.116	**1.000**	**1.000**	**0.999**	**0.984**	[83.06 , 99.87]	[89.62 , 99.84]
Modena	**GNN**	269	1	0.996	2.1398	1.1201	1.4628	0.040	0.021	0.219	0.990	0.916	0.604	0.955	[40.02 , 67.85]	[44.08 , 67.17]
	GNN	269	8	0.970	1.9036	1.0769	1.3797	0.038	0.021	0.208	0.995	0.928	0.626	0.960	[40.02 , 67.85]	[44.61 , 66.98]
	PhyInGNN	269	1	0.996	2.0184	1.0452	1.4207	0.037	0.020	0.213	0.992	0.917	0.626	0.957	[40.02 , 67.85]	[44.59 , 67.37]
	PhyInGNN	269	8	0.970	**0.0902**	**0.1935**	**0.3004**	**0.007**	**0.003**	**0.076**	**1.000**	**1.000**	**0.988**	**0.998**	[40.02 , 67.85]	[41.24 , 67.32]
ZhiJiang	**GNN**	113	1	0.991	80.4139	7.4941	8.9674	1.499	0.117	0.393	0.466	0.270	0.145	0.102	[46.72 , 82.90]	[50.82 , 81.14]
	GNN	113	4	0.965	71.0938	6.6677	8.4317	1.334	0.105	0.421	0.540	0.350	0.231	0.189	[46.72 , 82.90]	[47.31 , 81.31]
	PhyInGNN	113	1	0.991	**0.1038**	**0.2436**	**0.3221**	**0.049**	**0.004**	0.042	**1.000**	**1.000**	**0.995**	**0.999**	[46.72 , 82.90]	[46.58 , 82.77]
	PhyInGNN	113	4	0.965	0.1168	0.2672	0.3418	0.053	**0.004**	**0.038**	**1.000**	**1.000**	0.993	**0.999**	[46.72 , 82.90]	[47.16 , 83.12]

Table A.2: Ablation study on the Kadu network with two sensors. Best results in **bold**.

Configuration	**MAE**	**NMAE**	**MAPE**	**MaxPE**	**Acc@10%**	**Acc@5%**	**Acc@2%**	$\mathbf{R}^2$
Full Model	**0.140**	**0.0052**	**0.0015**	**0.0527**	**1.000**	**1.000**	**0.999**	**0.982**
No edge weights	0.169	0.0063	0.0018	0.0536	**1.000**	0.9999	0.998	0.976
No residual connection	0.388	0.0144	0.0040	0.0538	**1.000**	0.9999	0.991	0.898
Random measurement site	0.312	0.0116	0.0032	0.0670	**1.000**	0.9998	0.995	0.934
2 GNN layers	0.180	0.0067	0.0019	0.0536	**1.000**	**1.000**	**0.999**	0.975
1 GNN layer	0.422	0.0156	0.0044	0.0543	**1.000**	0.9999	0.988	0.881
No batch normalization	0.710	0.0263	0.0074	0.0576	**1.000**	0.9998	0.954	0.703
No dropout	0.421	0.0156	0.0044	0.0539	**1.000**	0.9999	0.986	0.878