Temporal Causal Graph Discovery in Complex HPC Network Traffic Simulations

Subha Ilamathy, Rafat Ashraf Joy, Matthew T. Dearing, Dr. Zhiling Lan

UNIVERSITY OF ILLINOIS CHICAGO
College of Engineering

Motivation

- Parallel Discrete Event Simulations (PDES) offers accurate HPC simulations but is **computationally intensive** and **slow** to scale.
- **Surrogate models** can accelerate simulations, and we explore if **causal insights** can improve their long-term forecasting stability.

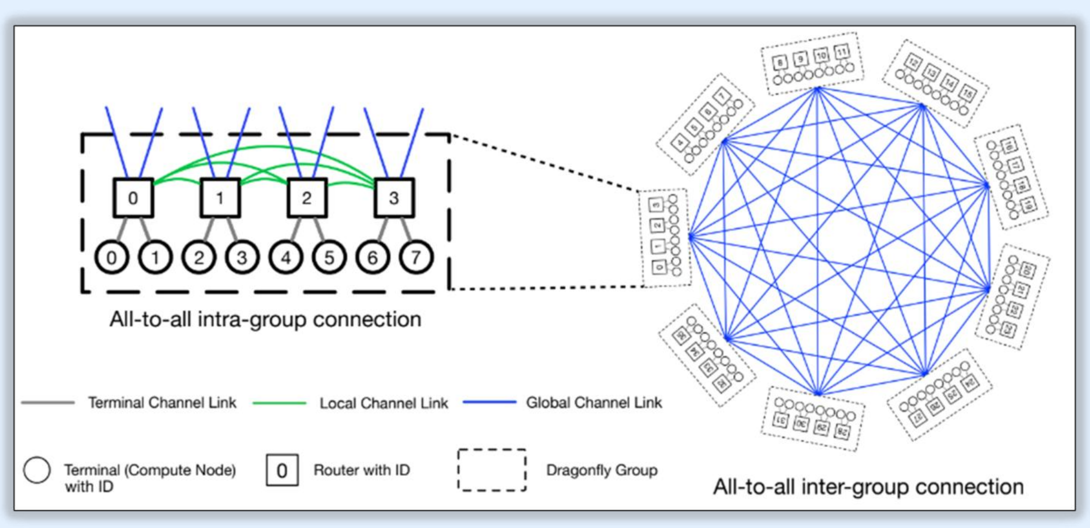
Research Questions

- Can **Causal Signals** hidden in HPC simulations **unlock better forecasting?**
- Can different **Causal Discovery methods** identify key drivers for accurate surrogate forecasting?

Features & Model

HPC Topology Simulated

Dragonfly [1] - a fully-connected graph
Routers: 36 **Compute Nodes:** 72



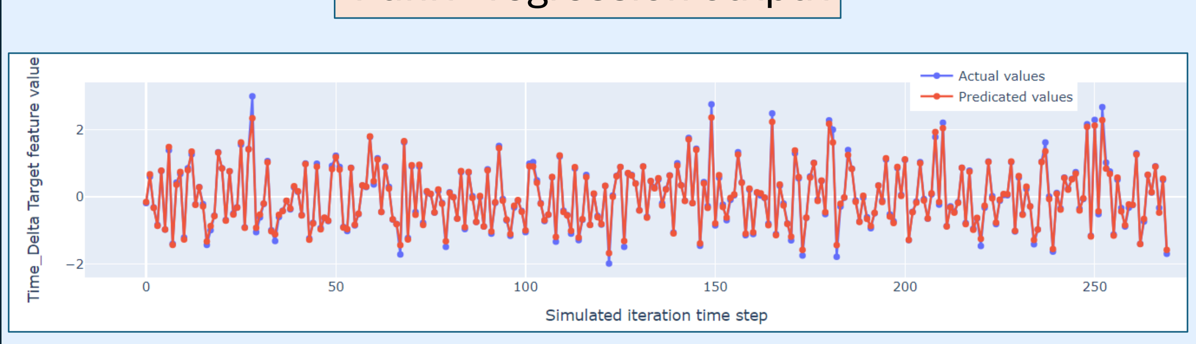
Data Features (per compute node)

- **44 features** from a synthetic HPC network simulation of MILC [2]. Data generated using CODES [3] framework using base and augmented features.
  1. Iteration time (Time_Diff)
  2. Network traffic (QOS_Data)
  3. Virtual Channel occupancy (VC)
  4. Downstream credits (DC)

Time Series Model Architecture

From [4]. Trained on exogenous features
Input: 28 prior time steps **Shape**: (28, 43)
Output: Application iteration time

1D ConvNet + Regularization
LSTM + Regularization
Dense layer (features)
1 unit - regression output

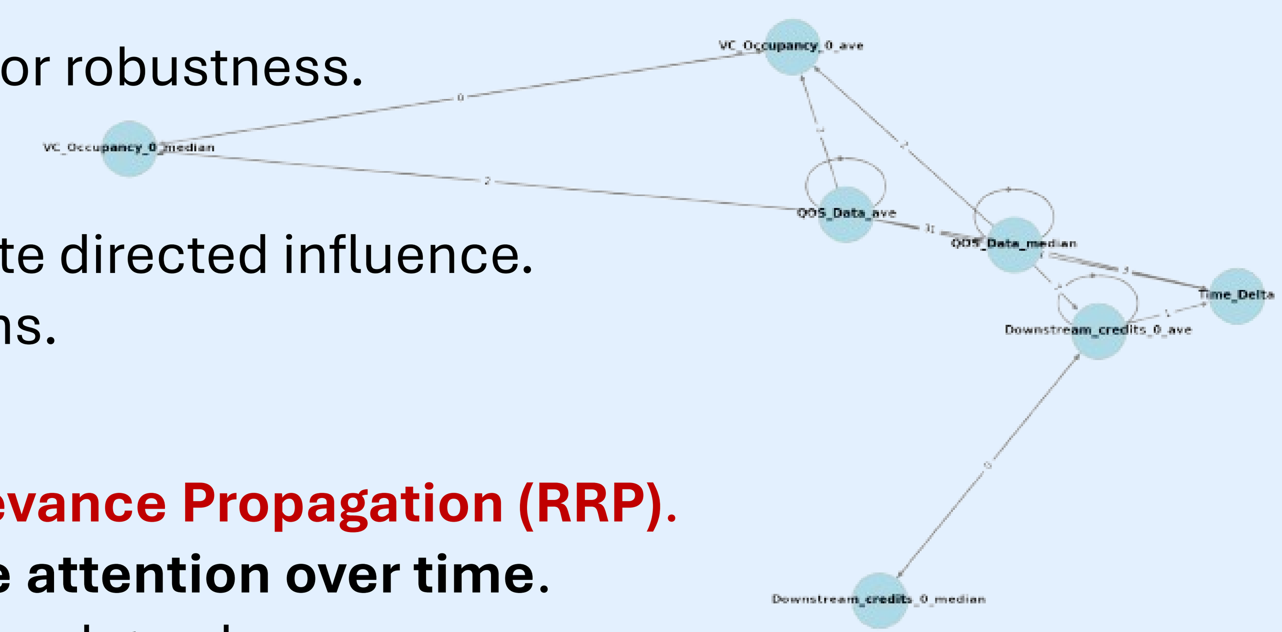


Experiments & Methods

1. Causal Discovery Methods

To identify the features most responsible for influencing the application iteration time, we applied **four Causal Discovery techniques** on multivariate time series data:

1. **Granger Causality -** Linear, lag-based causality detection.
2. **Enhanced Granger -** Weighted loss for robustness.
3. **NAVAR (Neural Additive VAR)**
   - Non-linear components to estimate directed influence.
   - Learns node-specific causal graphs.
4. **CausalFormer**
   - Transformer model **Relative Relevance Propagation (RRP).**
   - Captures **lag-aware multivariate attention over time**.
   - Interpretable attention-based causal graphs.



2. Feature Weighting Techniques

We evaluated **two feature** integration strategies to incorporate causal signals into surrogate model training.

1. **Direct Feature Scaling**
   - Causal features identified by NAVAR and CausalFormer were **amplified**.

$$x_i^{\text{scaled}} = 2 \cdot x_i, \quad \text{for all causal features } x_i.$$

2. **Attention-Based Weighting**
   - Attention weights learned during training using softmax, enabling **adaptive emphasis** on causal features.
   - Enhanced both **interpretability** and **forecasting robustness**.

$$\alpha_i^{\text{scaled}} = \alpha_i \cdot m_i$$

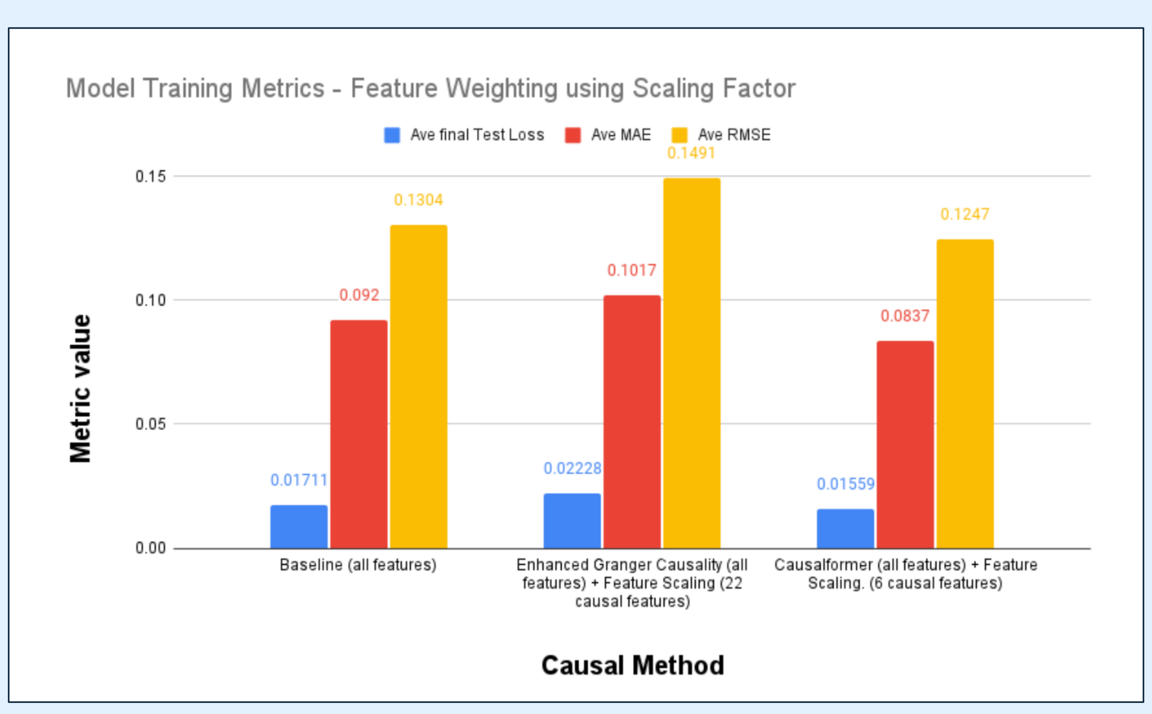
$$x_i^{\text{weighted}} = x_i \cdot \alpha_i^{\text{scaled}}$$

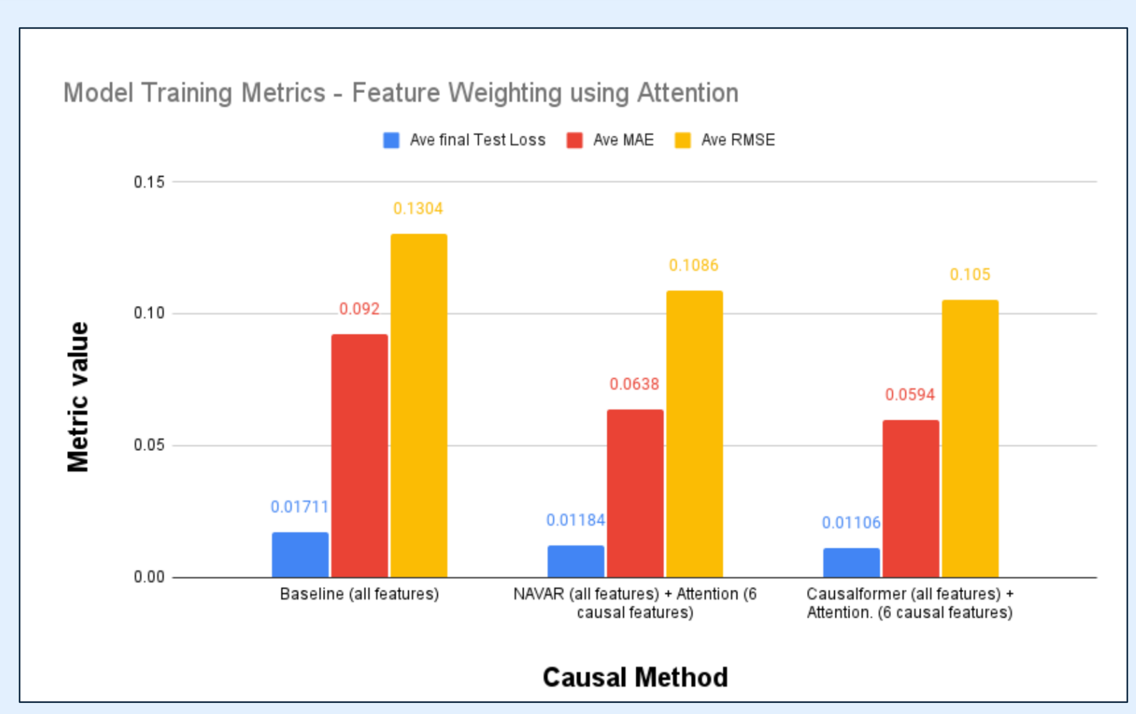
Results & Future Work

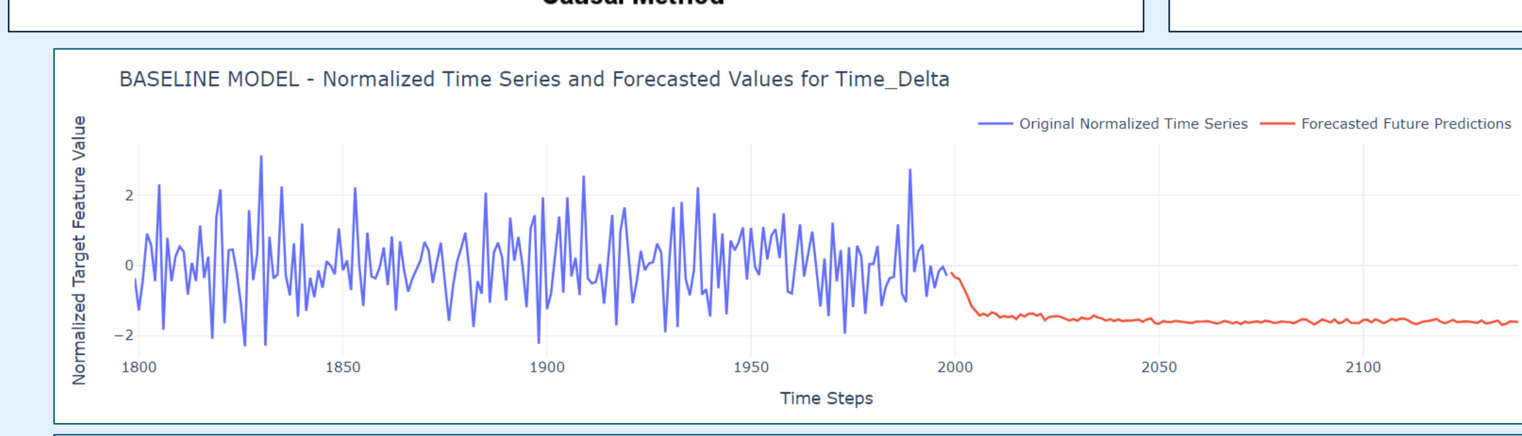
Which Causal Methods Worked Best?

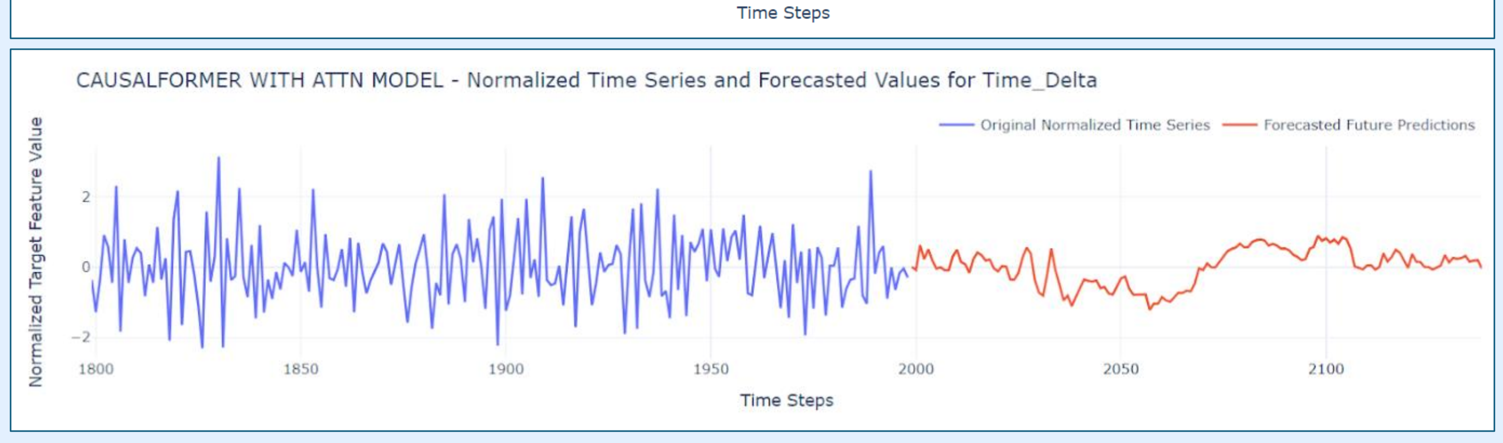
Forecasting Results by Causal Method

Method	Type	Result
Granger	Statistical	22 features selected, Poor model performance
Enhanced Granger	Statistical (WDS)	Slightly better than Granger, but still suboptimal
NAVAR	Neural VAR	Best results when used with attention weighting
CausalFormer	Transformer-based	**Top performer with Causal Attention**









We suggest that considering causal features **opens the door** to enhance forecasting capability through future research:

- Train causally-informed models for *each* exogenous variable.
- Use these independent predictions in the next-step predictors of the target variable model.

[1] John, Kim et al., 2008. "Technology-driven highly-scalable dragonfly topology", *ACM SIGARCH Computer Architecture News*, vol. 36, no. 3.
[2] This work was in part based on the MILC collaboration's public lattice gauge theory code. See http://physics.utah.edu/~detar/milc.html.
[3] Ross, et al. 2017. "Enabling parallel simulation of large-scale HPC network systems", In IEEE Transactions on Parallel and Distributed Systems, vol. 28. 87–100.
[4] Dearing, M. T. 2024 "Deep Learning Surrogate Models for Network Simulation," in *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation ACM*, pp. 65-66.