Graph Neural Network for Turbidity Prediction in the Upper Esopus Creek Watershed

George H. Myers, Aidan Hayes, Kristen Underwood, John Kemper, Andrew Schroth

CERM 2025

10/23/2025

Background

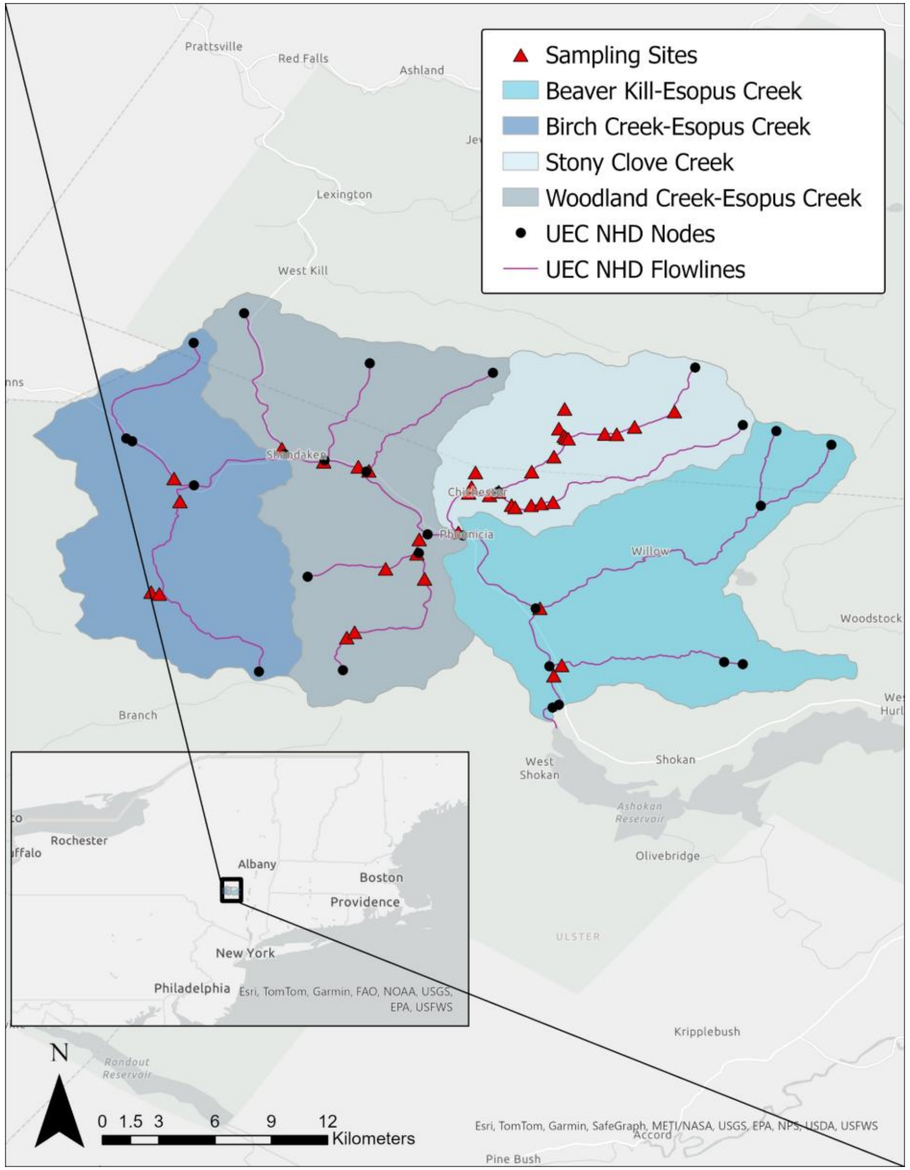
Suspended Sediment in the Upper Esopus Creek

- The Upper Esopus Creek (UEC) is impacted by chronic and acute turbidity (Tn) levels [3]
- Tn is generally a function of suspended sediment concentration (SSC) in the UEC [4]
- Heightened SSC have numerous negative environmental and economic impacts [1]
- Increased water treatment costs are of particular concern in the UEC, as the Ashokan reservoir provides ~40% of NYC's drinking water.

Turbid runoff in the Stony Clove Creek during the 12/25/2020 storm. NYC DEP

NYC DEP and the USGS initiated a study to understand where Tn is sourced in the UEC, and to evaluate the effectiveness of stream turbidity reduction projects (STRPs)

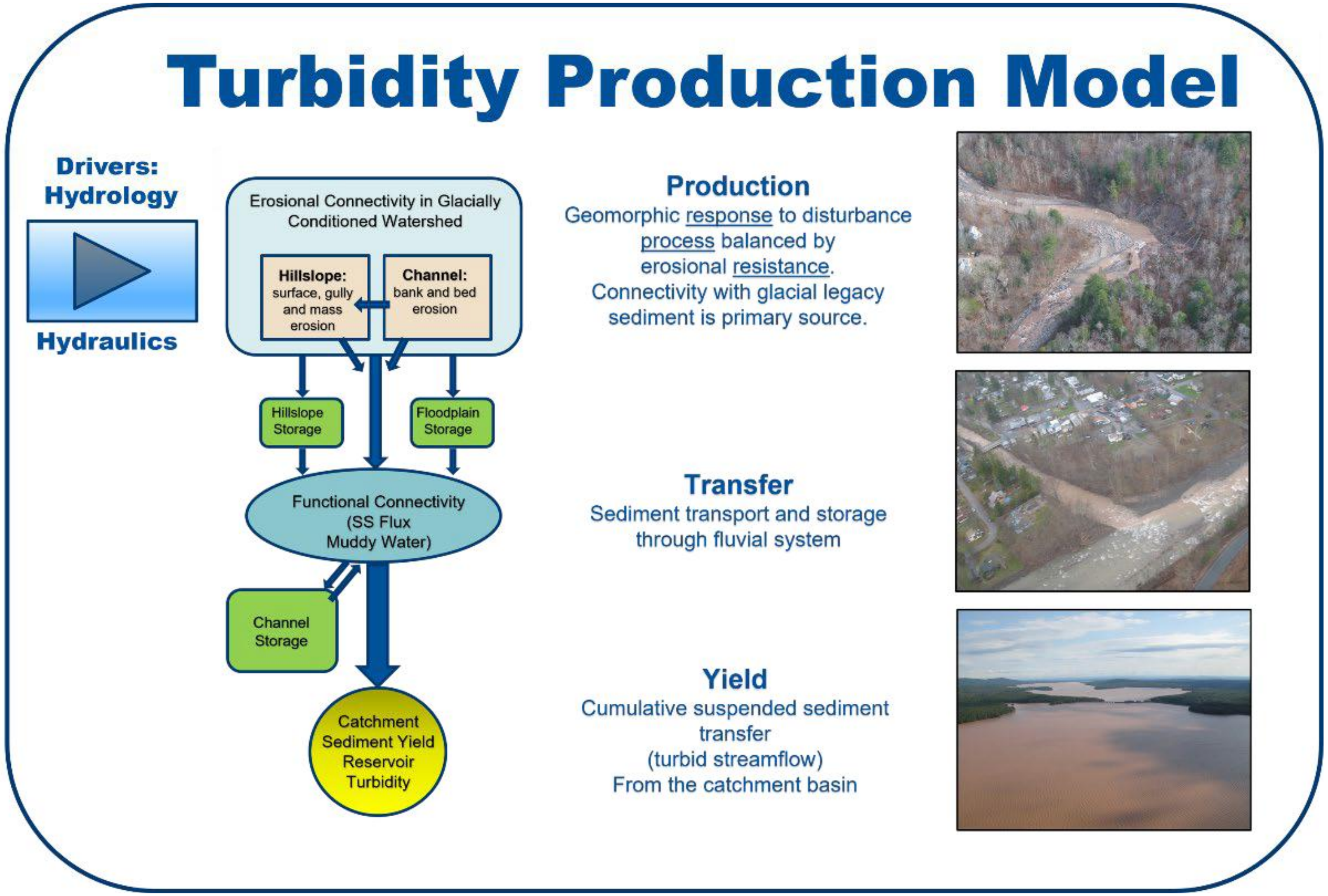
This nested monitoring network provides an amazing opportunity to apply data driven methods to model Tn levels and provenance in the UEC watershed.



Monitoring network in the UEC watershed.

Inland water networks and their water quality vary in space and time.

- Temporal drivers: precipitation, snow melt, river discharge, land use changes.
- Spatial drivers: sediment source variations, human interventions like dams and armoring, slope.
- We need methods that can integrate both spatial and temporal dynamics into their predictions.



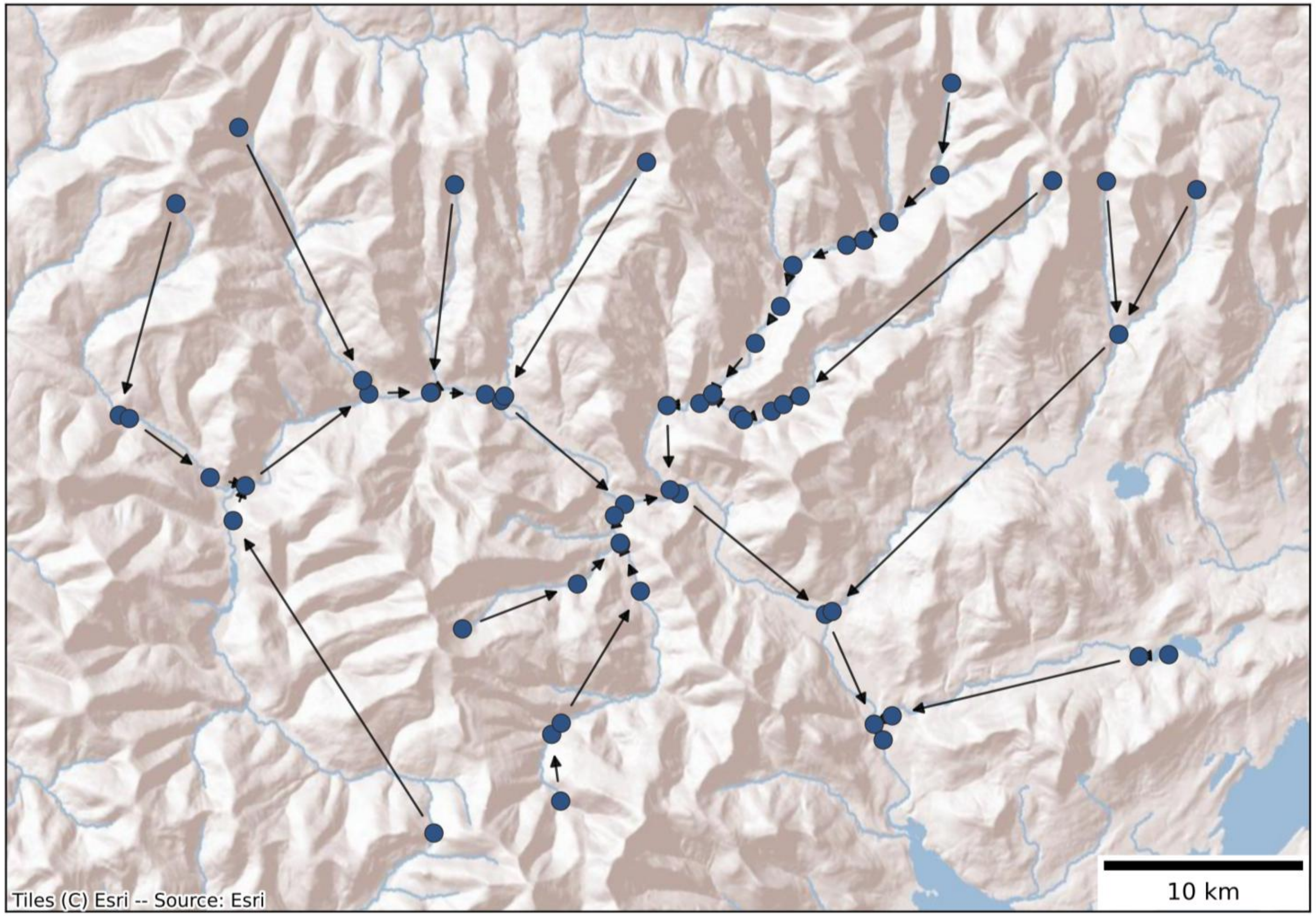
Conceptual model of Tn production. NYC DEP

Deep Learning for Water Quality Modeling

- Well suited to discovering complex patterns from high-dimensional data
- Designed to learn directly from spatiotemporal data and embed dynamic relationships
- Relatively quick to train and allow for near real-time predictions
- Outperform process-based models for water quality prediction [7]

Graph Neural Networks (GNNs)

- GNNs are trained on graph structured data which are comprised of a series of nodes connected by edges
- GNNs leverage the spatial, temporal, and hydrologic connectivity that governs nutrient and sediment flux
- GNNs have shown great promise for predicting water quality and quantity at gaged and ungaged locations in river networks [1, 5, 8]



A directed, spatial graph of the UEC watershed.

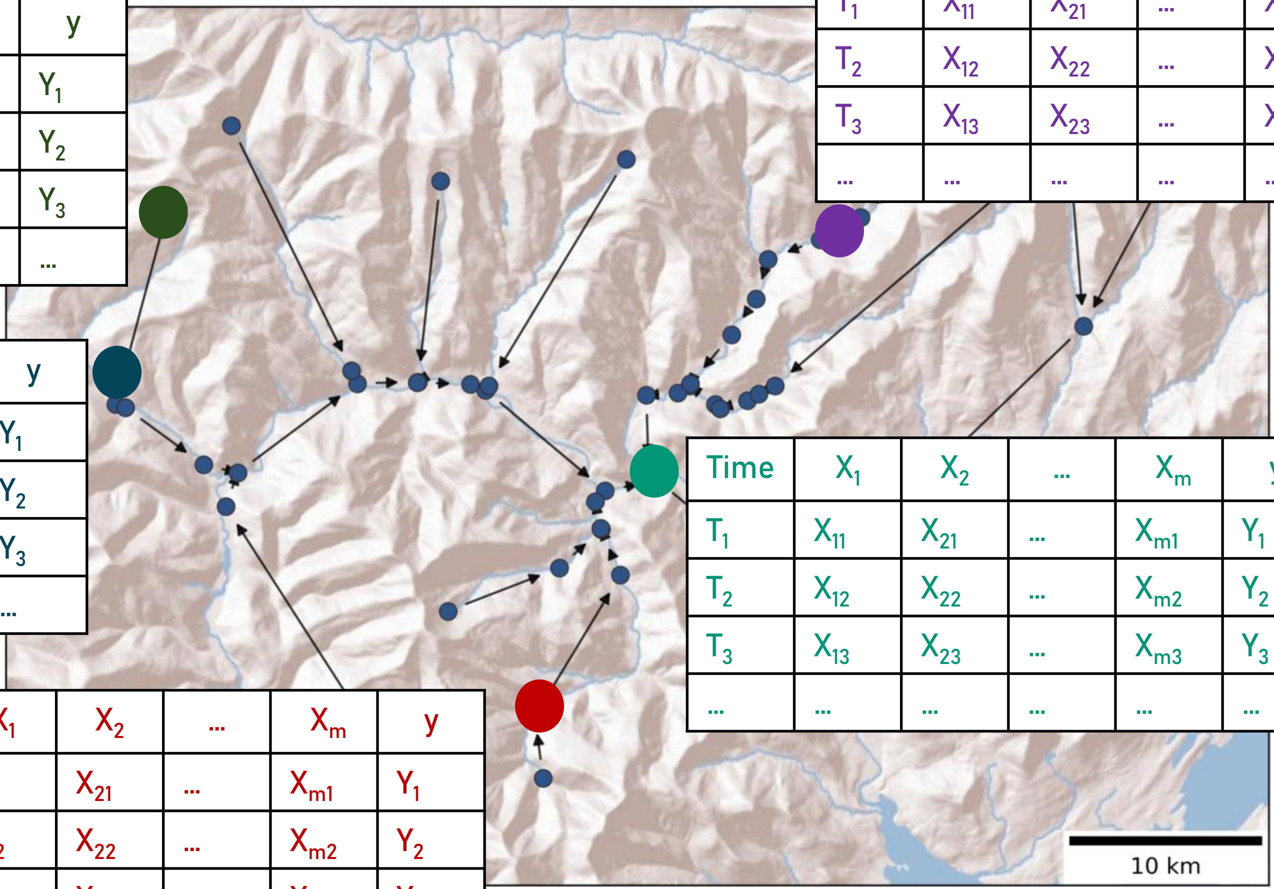
Graph data structure intuition

Site	**Time**	**X_1**	**X_2**	**…**	**X_m**	**y**
1	T_1	X_{111}	X_{211}	…	X_{m11}	Y_{11}
1	T_2	X_{112}	X_{212}	…	X_{m12}	Y_{12}
1	T_3	X_{113}	X_{213}	…	X_{m13}	Y_{13}
…	…	…	…	…	…	…
2	T_1	X_{121}	X_{221}	…	X_{m21}	Y_{21}
2	T_2	X_{122}	X_{222}	…	X_{m22}	Y_{22}
…	…	…	…	…	…	…
n	T_1	X_{1n1}	X_{2n1}	…	X_{mn1}	Y_{n1}
…	…	…	…	…	…	…
n	T_k	X_{1nk}	X_{2nk}	…	X_{mnk}	Y_{nk}

Graph Data Structure Intuition

Time	X_1	X_2	…	X_m	y
T_1	X_{11}	X_{21}	…	X_{m1}	Y_1
T_2	X_{12}	X_{22}	…	X_{m2}	Y_2
T_3	X_{13}	X_{23}	…	X_{m3}	Y_3
…	…	…	…	…	…

Time	X_1	X_2	…	X_m	y
T_1	X_{11}	X_{21}	…	X_{m1}	Y_1
T_2	X_{12}	X_{22}	…	X_{m2}	Y_2
T_3	X_{13}	X_{23}	…	X_{m3}	Y_3
…	…	…	…	…	…



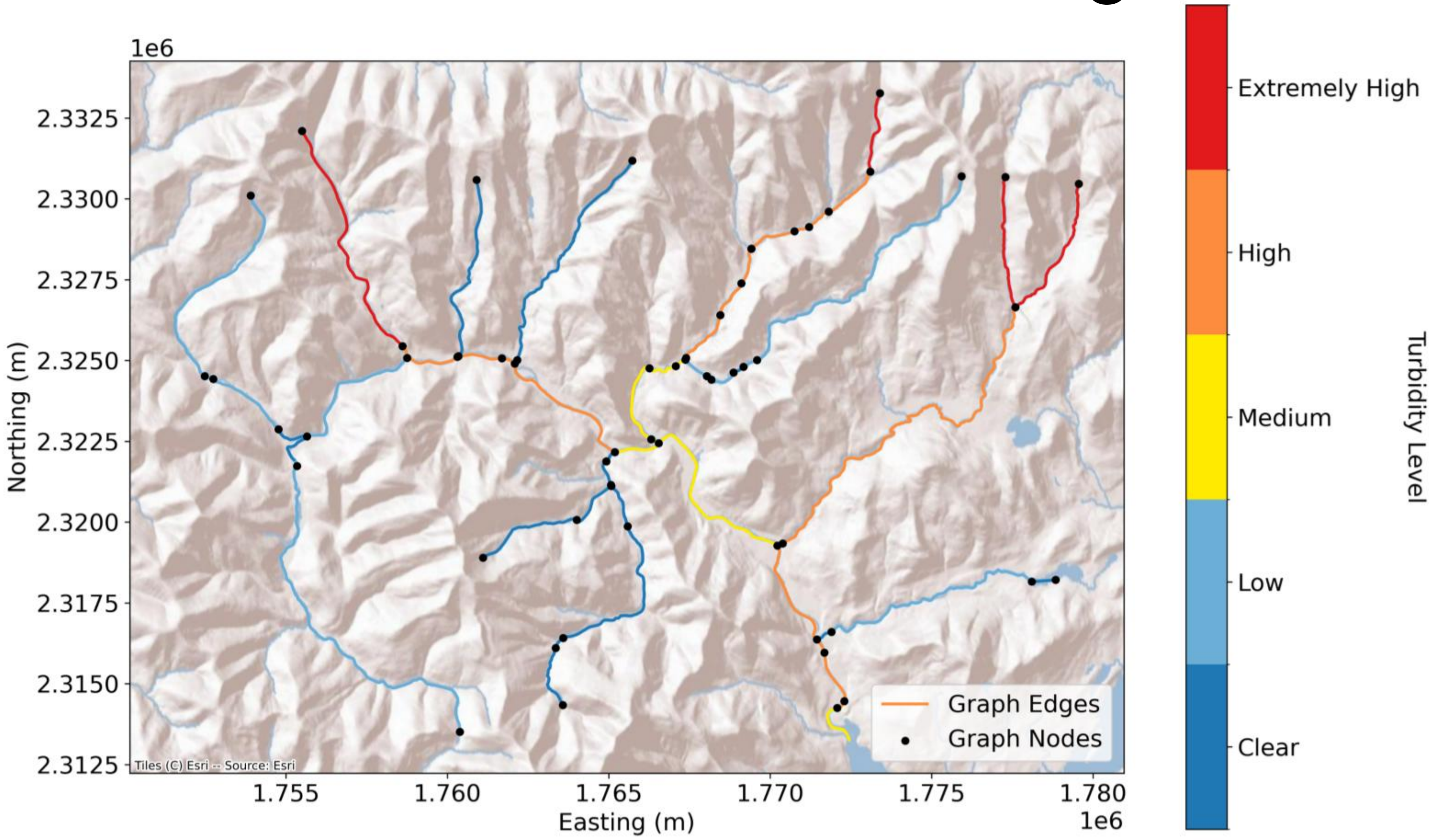
Time	X_1	X_2	…	X_m	y
T_1	X_{11}	X_{21}	…	X_{m1}	Y_1
T_2	X_{12}	X_{22}	…	X_{m2}	Y_2
T_3	X_{13}	X_{23}	…	X_{m3}	Y_3
…	…	…	…	…	…

Time	X_1	X_2	…	X_m	y
T_1	X_{11}	X_{21}	…	X_{m1}	Y_1
T_2	X_{12}	X_{22}	…	X_{m2}	Y_2
T_3	X_{13}	X_{23}	…	X_{m3}	Y_3
…	…	…	…	…	…

Time	X_1	X_2	…	X_m	y
T_1	X_{11}	X_{21}	…	X_{m1}	Y_1
T_2	X_{12}	X_{22}	…	X_{m2}	Y_2
T_3	X_{13}	X_{23}	…	X_{m3}	Y_3
…	…	…	…	…	…

Novel application of GNNs for source tracing

We're applying a GNN for source tracing of Tn throughout the UEC watershed to help guide targeted STRPs.



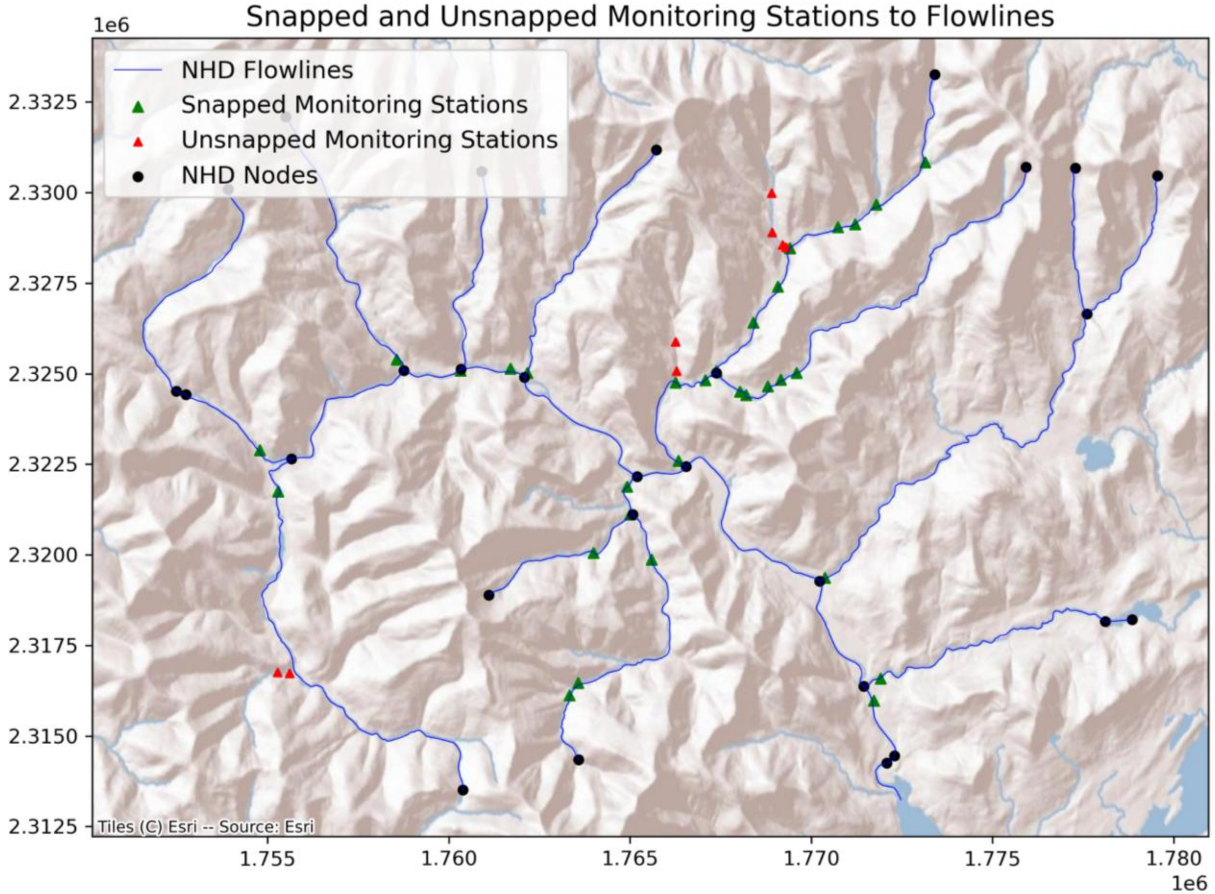
Theoretical source tracing of Tn in the UEC watershed

University of Vermont

Methods

Building the graph

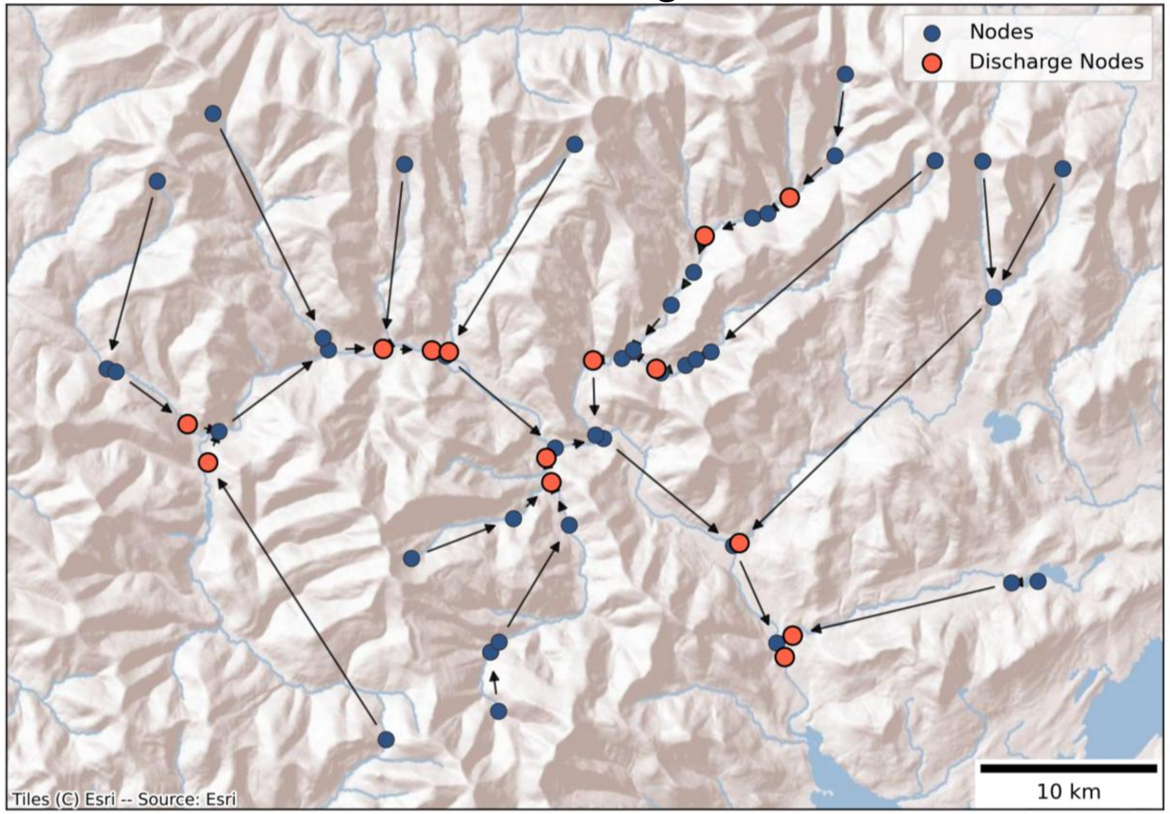
- A directed, spatial graph of the UEC was built from NHD MR
- High-frequency Tn and discharge data were down-sampled to an hourly mean
- Missing discharge data was filled from downstream nodes and applying a drainage area correction
- Hourly meteorological data were downloaded at each node from the AORC dataset
- Static reach attributes were compiled from the NHD and NLDI datasets



Map of the spatial graph of the UEC watershed showing monitoring stations that were included as nodes (green) and monitoring stations that were not (red).

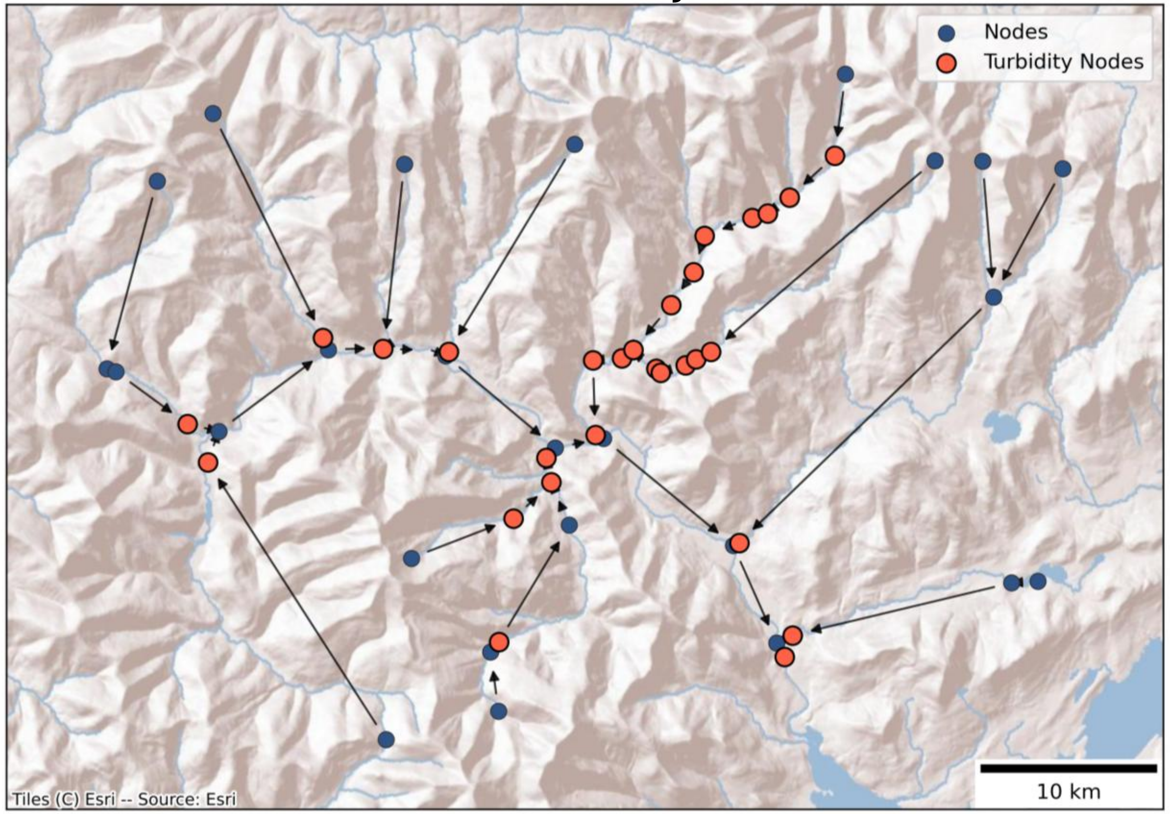
Compiling the available data

Discharge



Nodes with discharge data highlighted in red

Turbidity



Nodes with turbidity data highlighted in red

Preliminary Results

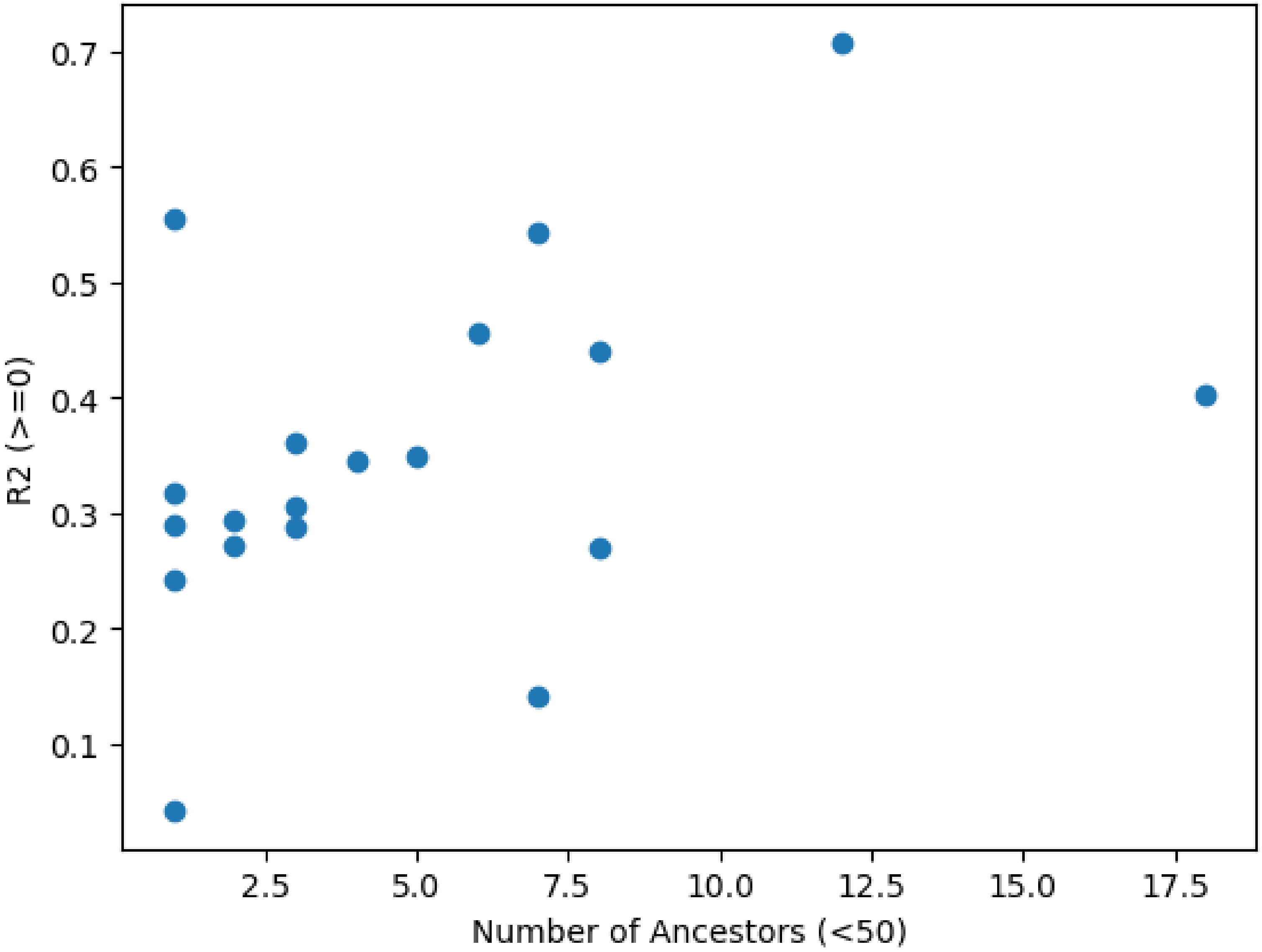
Overall model performance

Median MAE: 6.3 FNU

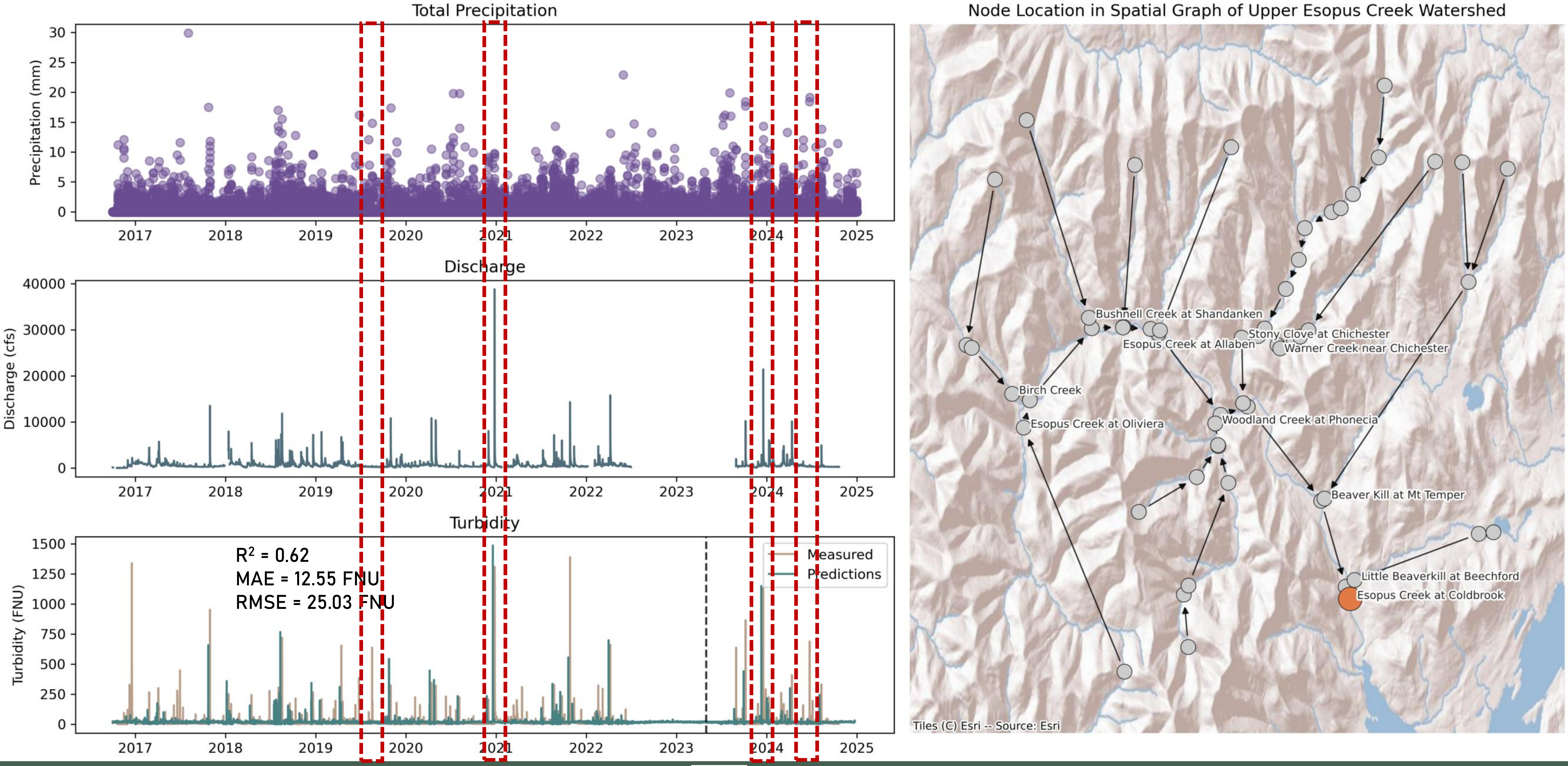
Median RMSE: 15.4 FNU

Median R^2: 0.29

- 3 best performing nodes are on the main branch of the Esopus
- 3 poorest performing nodes are towards the headwaters of the Stony Clove
- Generally, nodes with more upstream nodes (i.e., ancestors) perform better



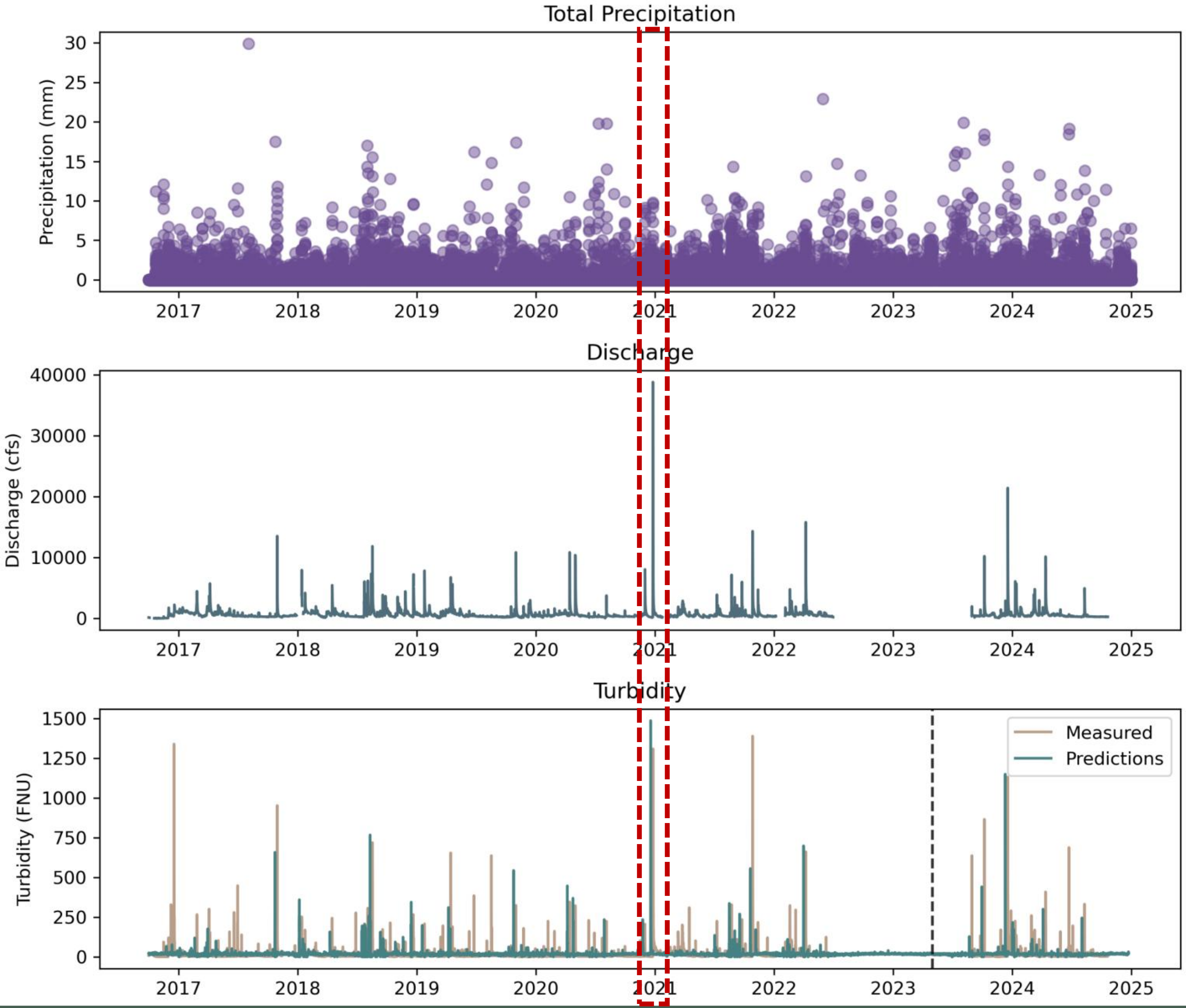
Esopus Creek at Coldbrook

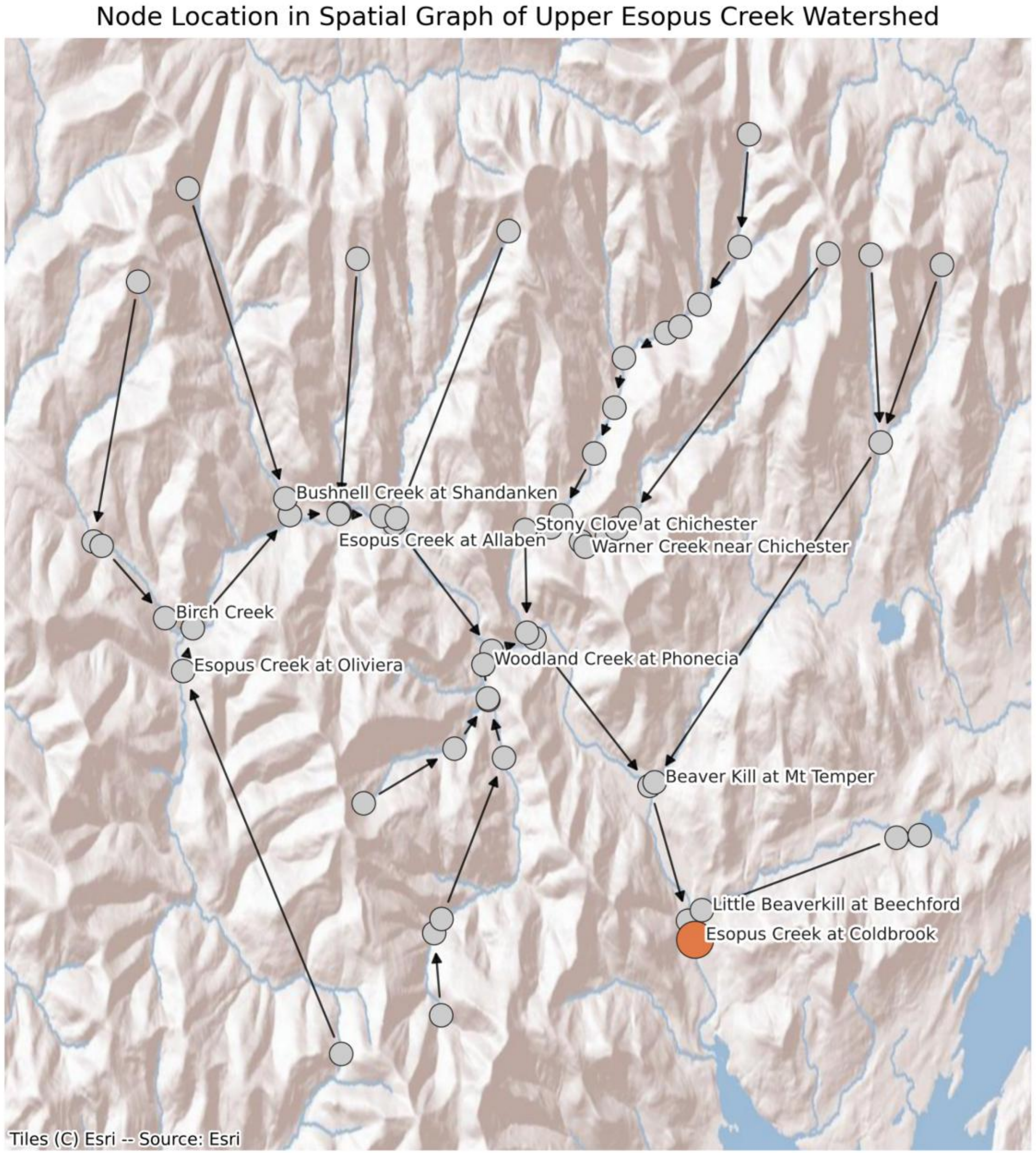


Preliminary Results

Source tracing of discrete event turbidity loads

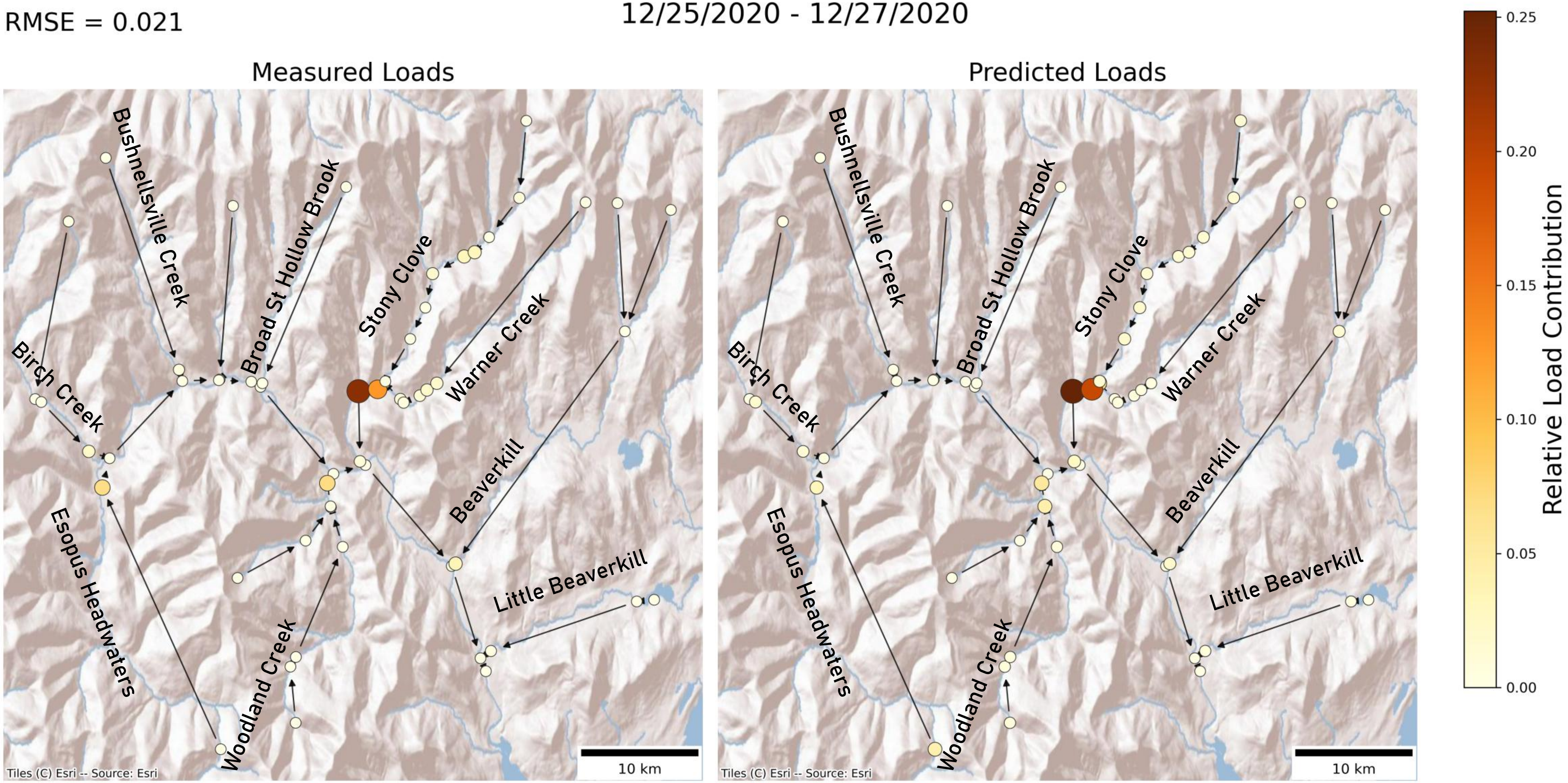
Esopus Creek at Coldbrook



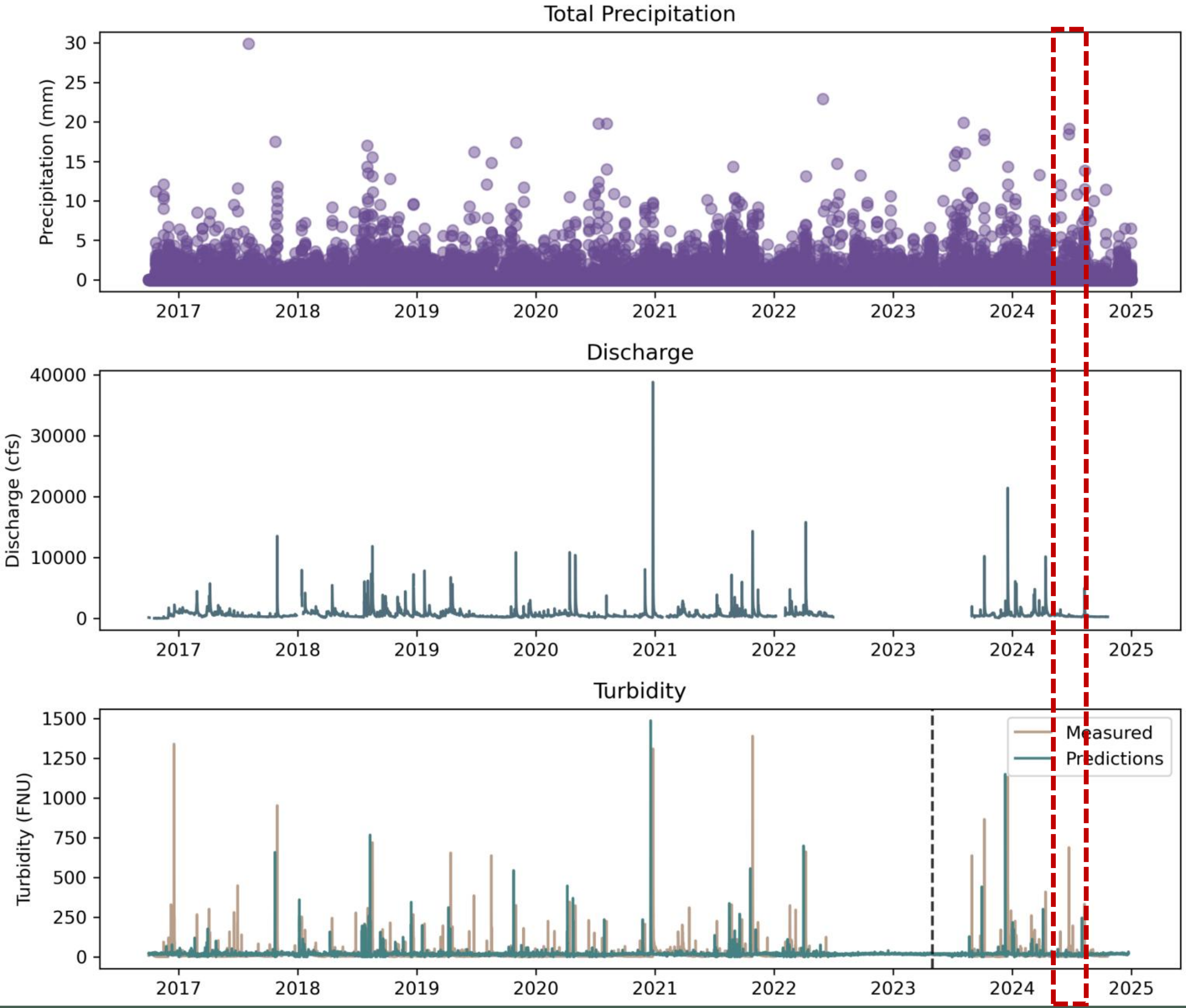


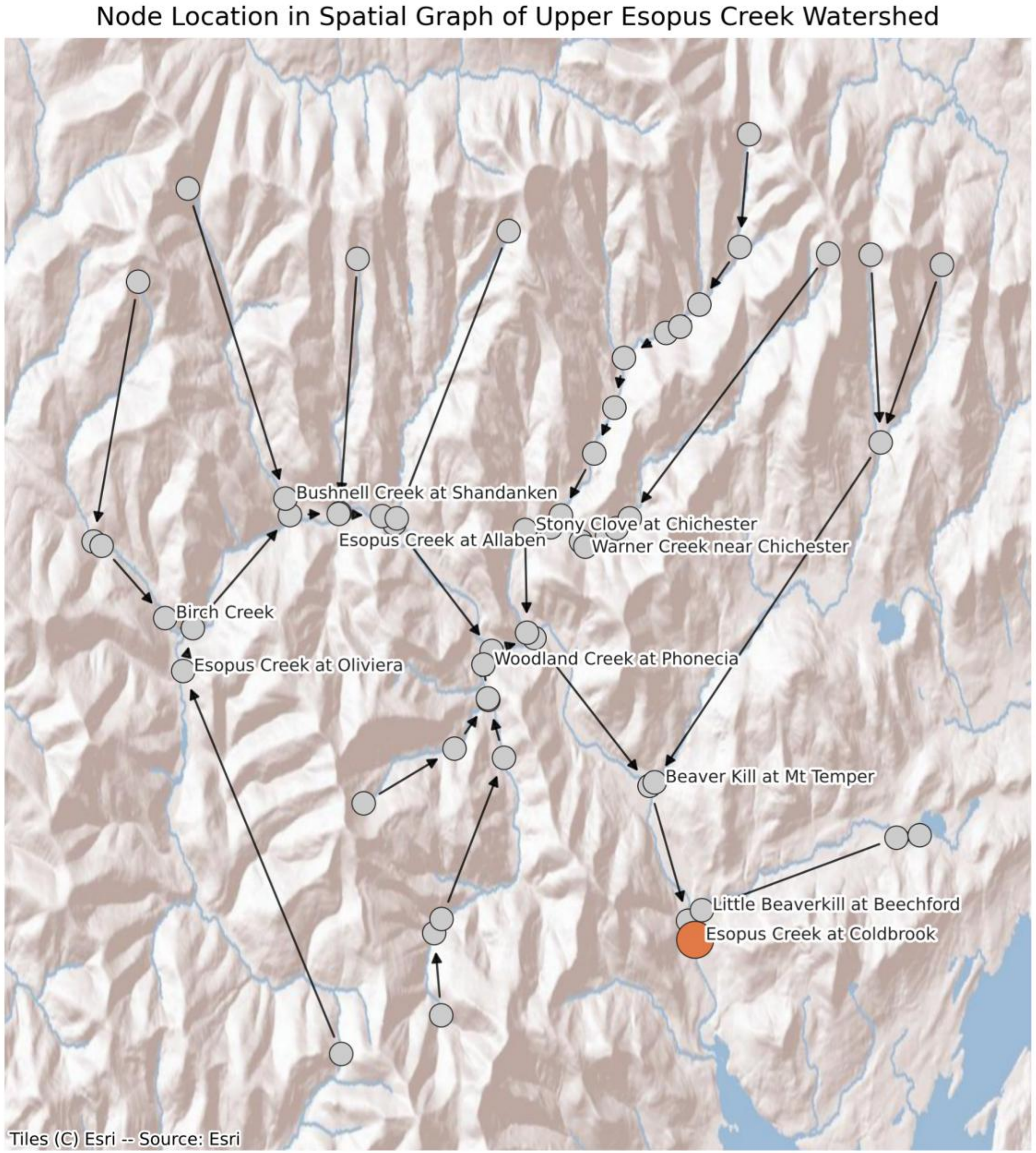
Proportion of Tributary Turbidity "Load" Contribution

12/25/2020 - 12/27/2020



Esopus Creek at Coldbrook

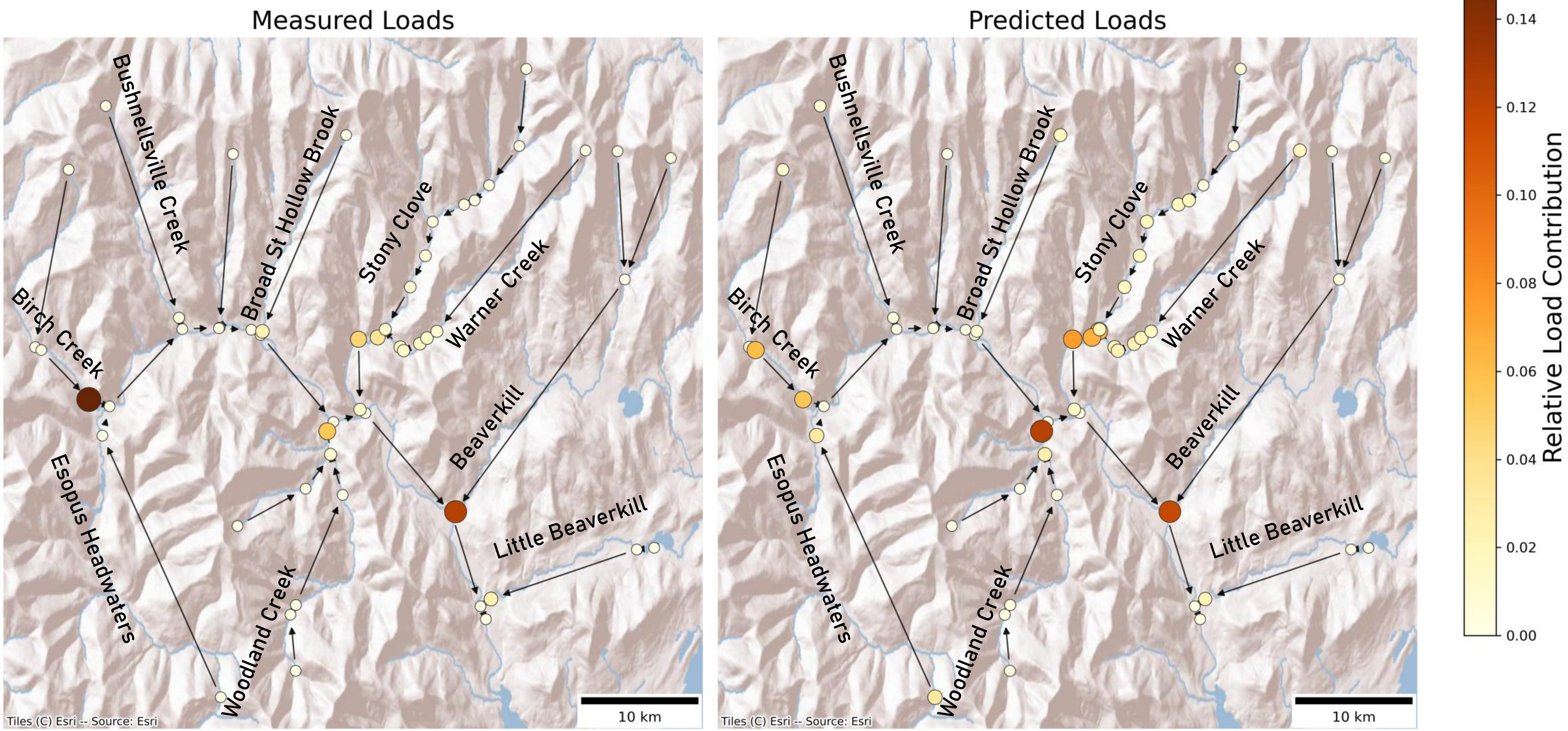




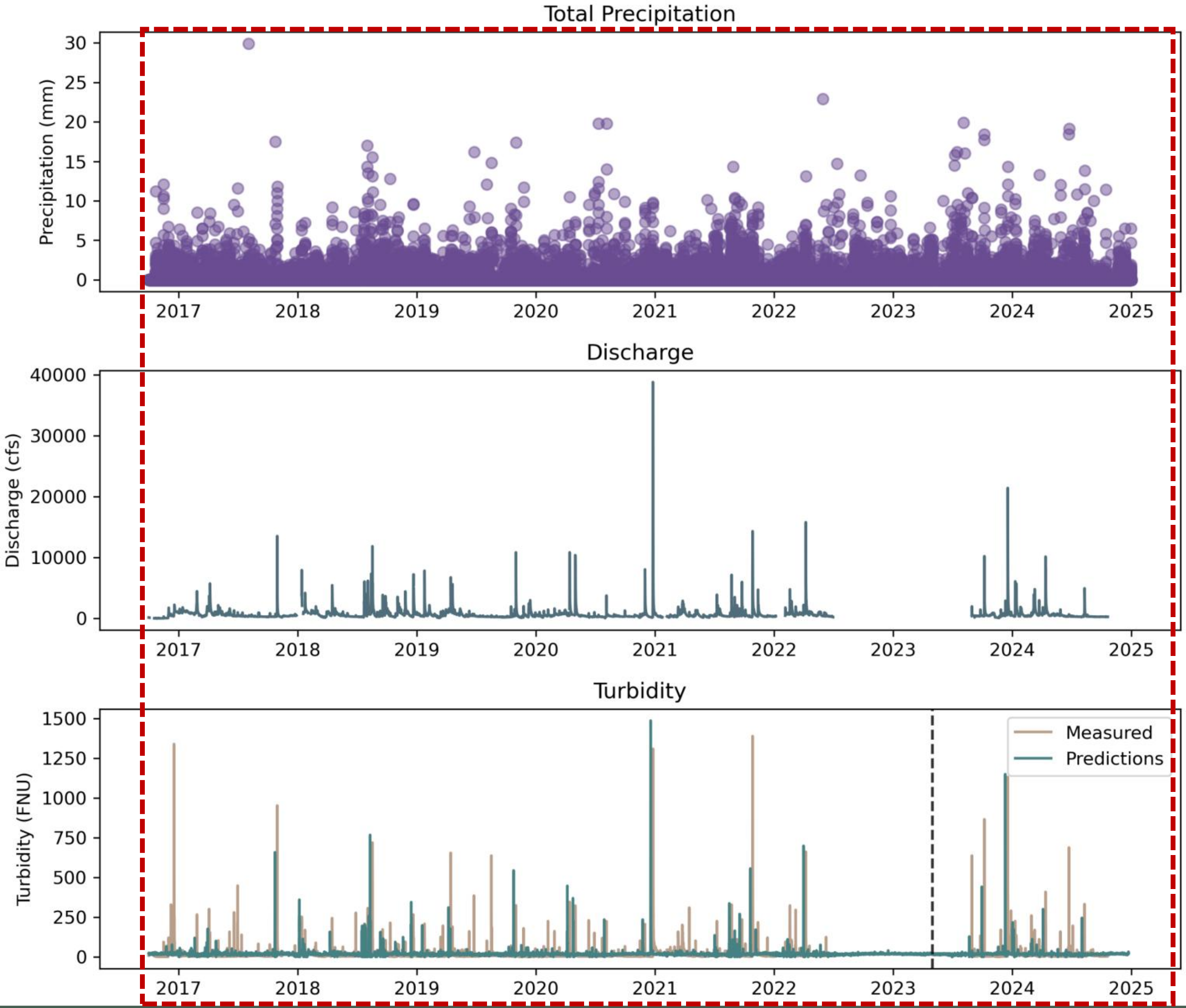
Proportion of Tributary Turbidity "Load" Contribution

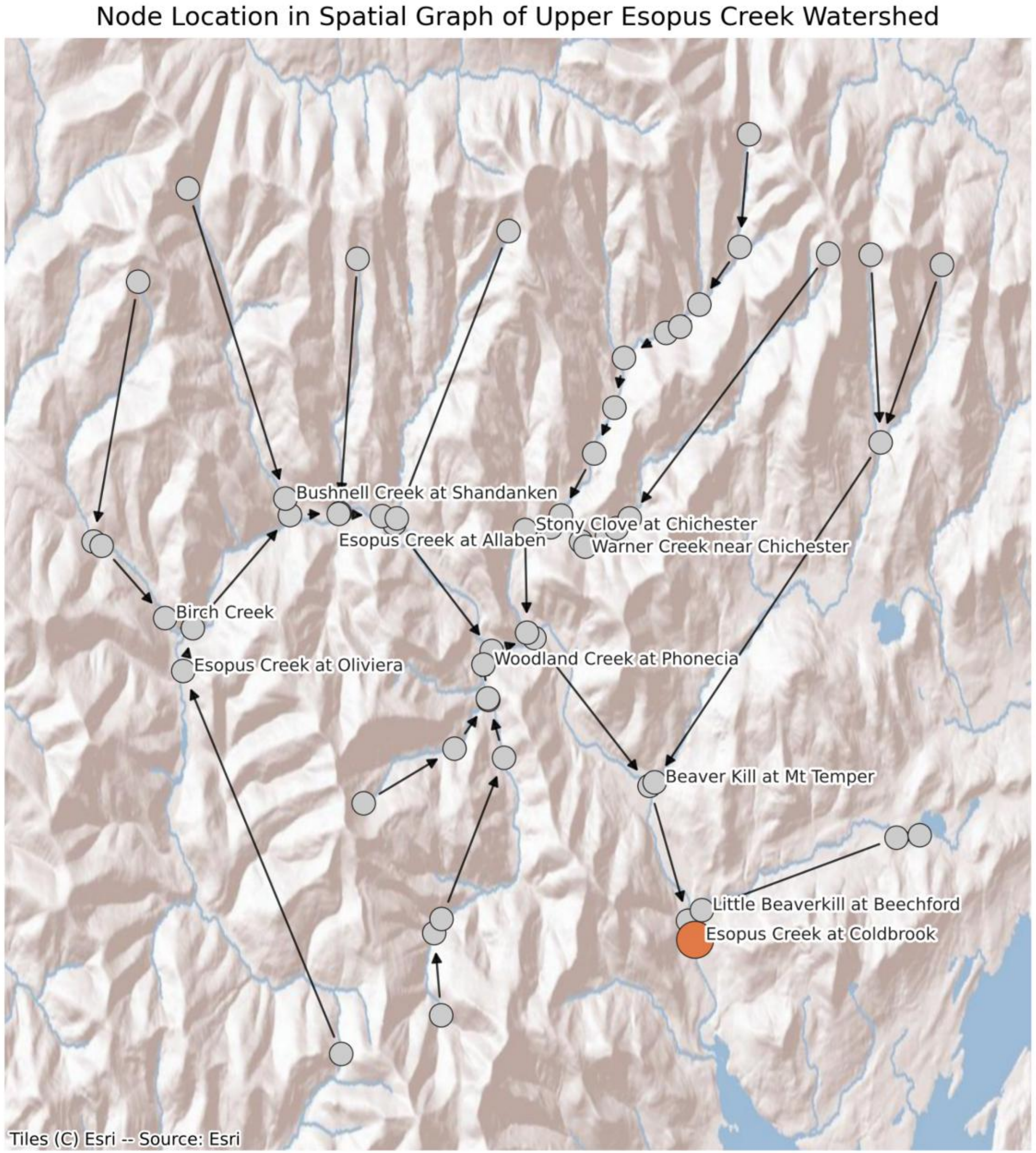
RMSE = 0.027

08/09/2024 - 08/11/2024



Esopus Creek at Coldbrook

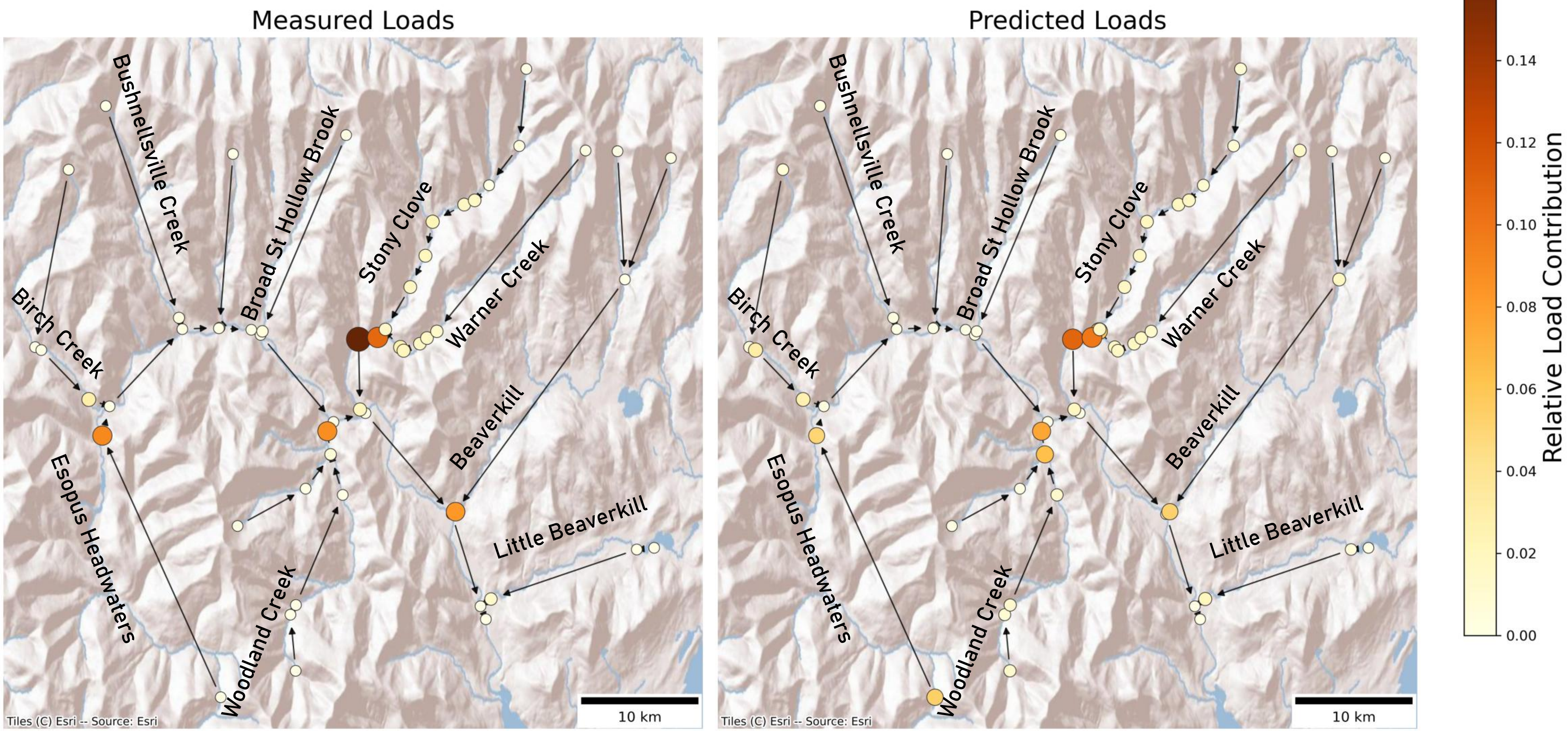




Proportion of Tributary Turbidity "Load" Contribution

RMSE = 0.019

10/01/2016 - 10/20/2024



Key takeaways

- Initial model performance is promising, with nodes closer to the Esopus generally performing better, and nodes closer to headwaters generally performing worse
- The model more accurately predicts Tn spikes associated with elevated discharge, and less accurate during low discharge turbidity export events that typically occur in summer
- Even with moderate predictive performance, the model is able to source trace Tn accurately
- Despite poor performance during low discharge Tn spikes, the model still produces accurate source tracing

Thank you! Questions?

PhD Candidate, Civil and Environmental Engineering, UVM
Gund Barrett Fellow | Department of Education GAANN Fellow
ghmyers@uvm.edu

References

1. Chen, S., Zwart, J. A., & Jia, X. (2022). Physics-Guided Graph Meta Learning for Predicting Water Temperature and Streamflow in Stream Networks. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2752–2761. https://doi.org/10.1145/3534678.3539115
2. Davies-Colley, R. J., & Smith, D. G. (2001). Turbidity, suspended sediment, and water clarity: A review. *JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION*, *37*(5), 1085–1101. https://doi.org/10.1111/j.1752-1688.2001.tb03624.x
3. Mukundan, R., Pierson, D. C., Schneiderman, E. M., O'Donnell, D. M., Pradhanang, S. M., Zion, M. S., & Matonse, A. H. (2013). Factors affecting storm event turbidity in a New York City water supply stream. *CATENA*, *107*, 80–88. https://doi.org/10.1016/j.catena.2013.02.002
4. Siemion, J., Bonville, D.B., McHale, M.R., and Antidormi, M.R., 2021, Turbidity–suspended-sediment concentration regression equations for monitoring stations in the upper Esopus Creek watershed, Ulster County, New York, 2016–19: U.S. Geological Survey Open-File Report 2021–1065, 27 p., https://doi.org/10.3133/ofr20211065.
5. Sun, A. Y., Jiang, P., Mudunuru, M. K., & Chen, X. (2021). Explore Spatio-Temporal Learning of Large Sample Hydrology Using Graph Neural Networks. *Water Resources Research*, *57*(12), e2021WR030394. https://doi.org/10.1029/2021WR030394
6. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020). T-GCN: A Temporal Graph ConvolutionalNetwork for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3848–3858. https://doi.org/10.1109/TITS.2019.2935152
7. Zhi, W., Appling, A. P., Golden, H. E., Podgorski, J., & Li, L. (2024). Deep learning for water quality. *Nature Water*, *2*(3), 228–241. https://doi.org/10.1038/s44221-024-00202-z
8. Zwart, J. A., Diaz, J., Hamshaw, S., Oliver, S., Ross, J. C., Sleckman, M., Appling, A. P., Corson-Dosch, H., Jia, X., Read, J., Sadler, J., Thompson, T., Watkins, D., & White, E. (2023). Evaluating deep learning architecture and data assimilation for improving water temperature forecasts at unmonitored locations. *Frontiers in Water*, *5*. https://doi.org/10.3389/frwa.2023.1184992

Model inputs and target variable

Meteorological: Air temperature, downward long-wave radiation flux, downward short-wave radiation flux, pressure, specific humidity, total precipitation, U and V component of wind

Hydrological: Discharge

Reach Attributes: % sand, % silt, % clay, # of dams, hydraulic conductivity, headwater node drainage area, length of reach, max elevation, drainage area, slope, stream order

Target: Turbidity

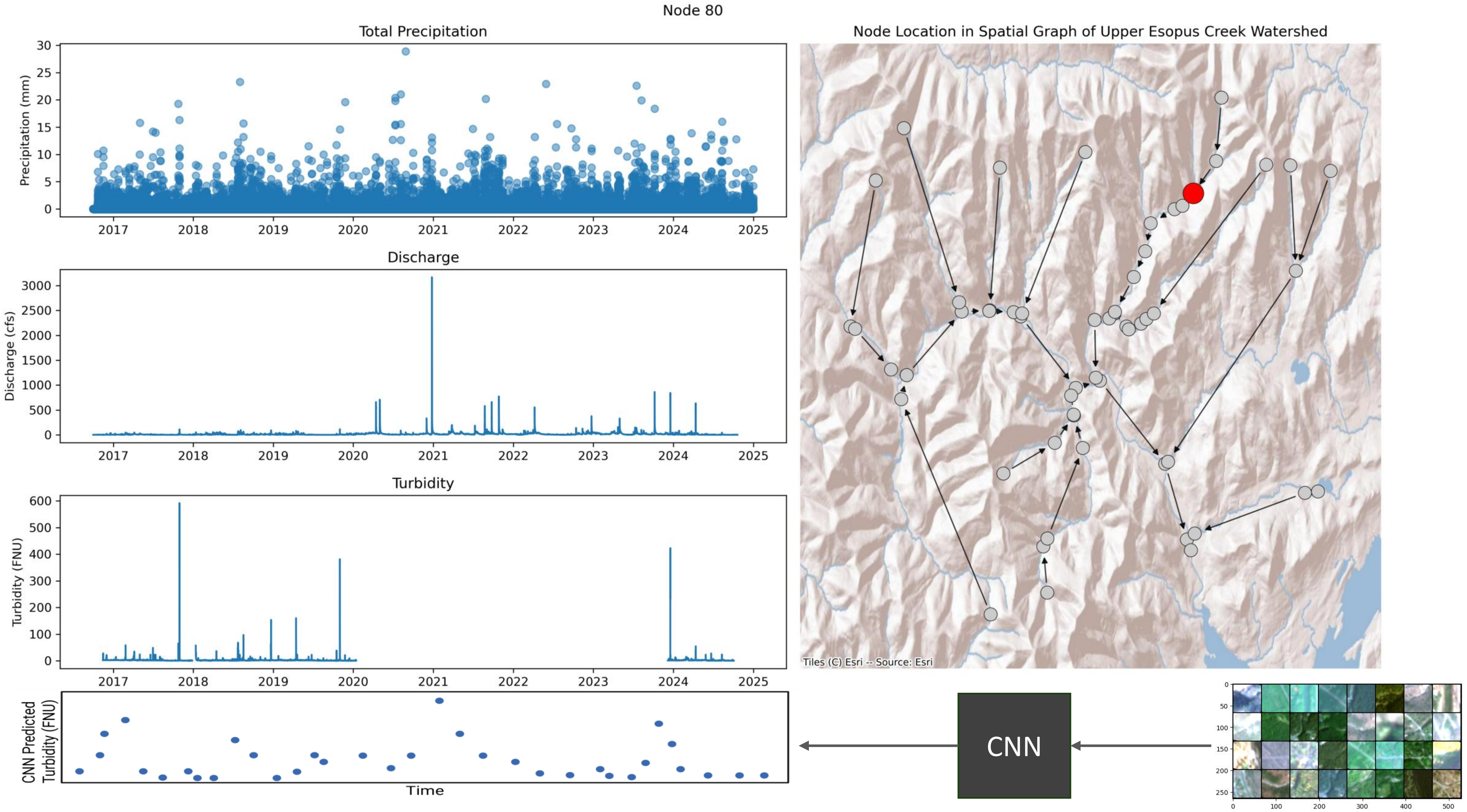
Next Steps

- Include additional training data (CNN predicted turbidity, change in temperature, accumulated precipitation, construction dates of STRPs) to improve model predictions

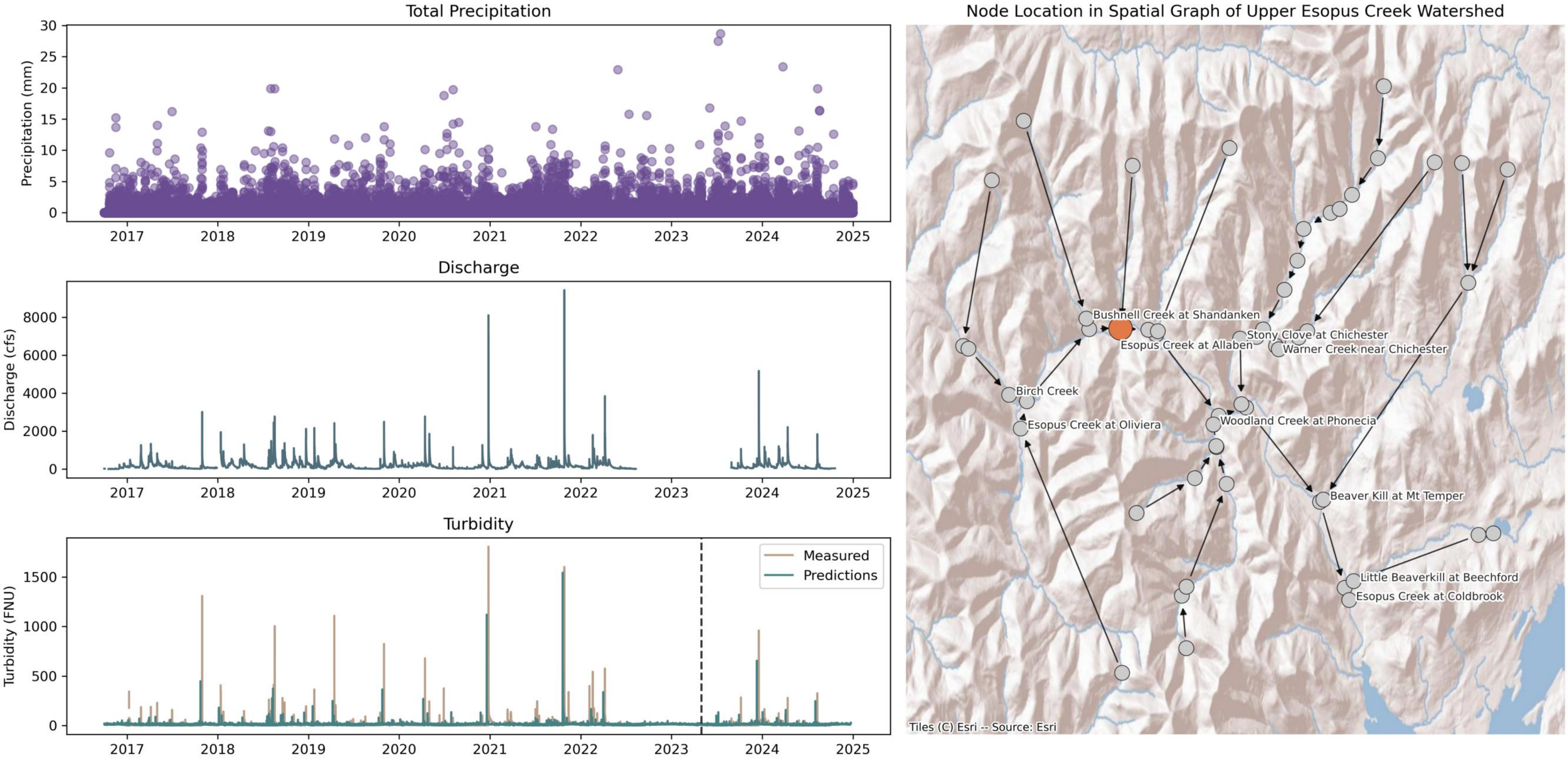
- Compare source tracing in the Stony Clove against sediment fingerprinting study from 2017 - 2020

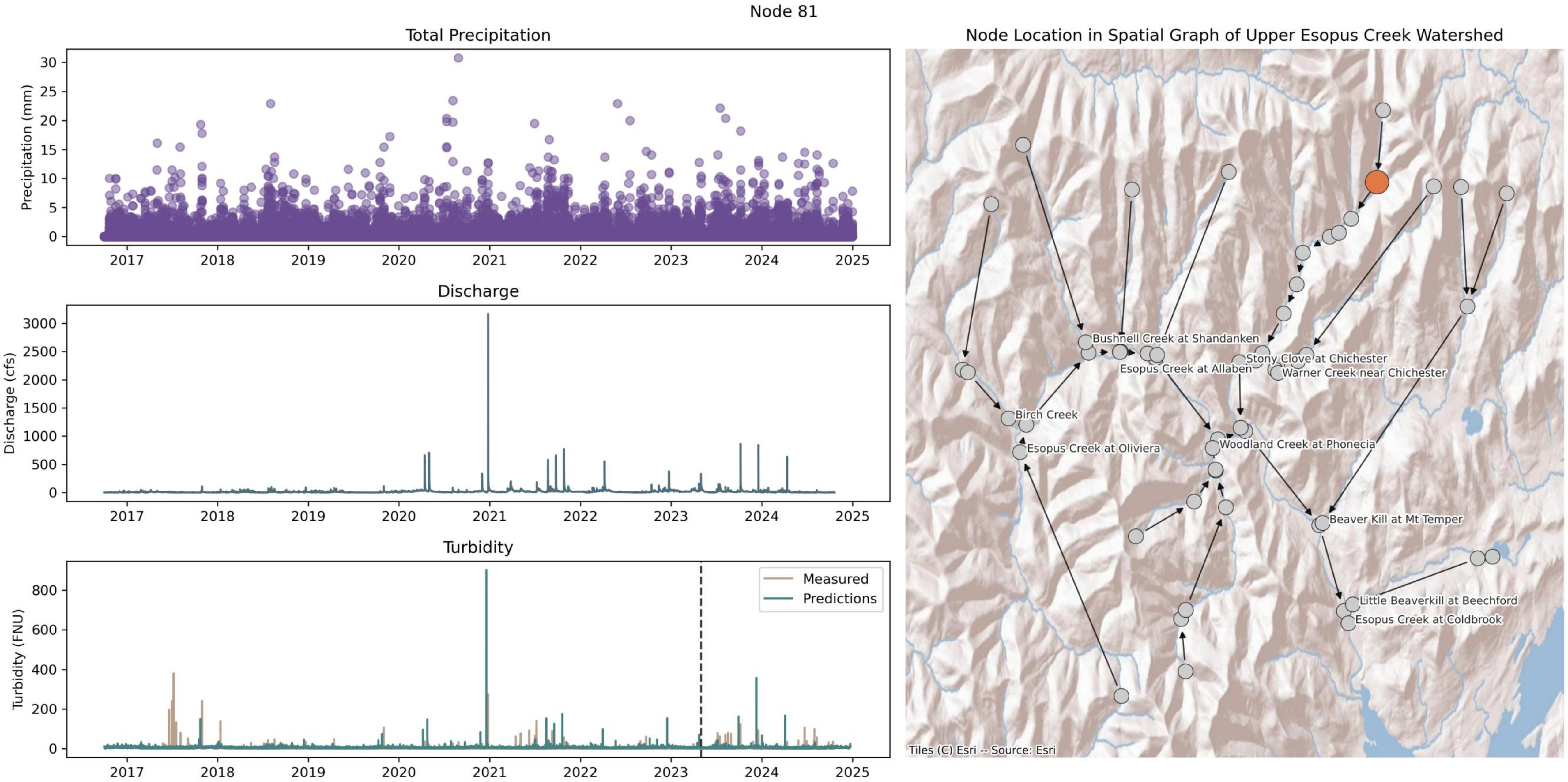
- Test additional GNN model architectures, hyperparameters, etc.

- Include STRP data to capture effect of STRPs and test various STRP scenarios for Tn reduction potential


Node 80
Total Precipitation
Precipitation (mm)
Discharge
Discharge (cfs)
Turbidity
Turbidity (FNU)
CNN Predicted Turbidity (FNU)
Time
Node Location in Spatial Graph of Upper Esopus Creek Watershed
Tiles (C) Esri -- Source: Esri
CNN

Node 53

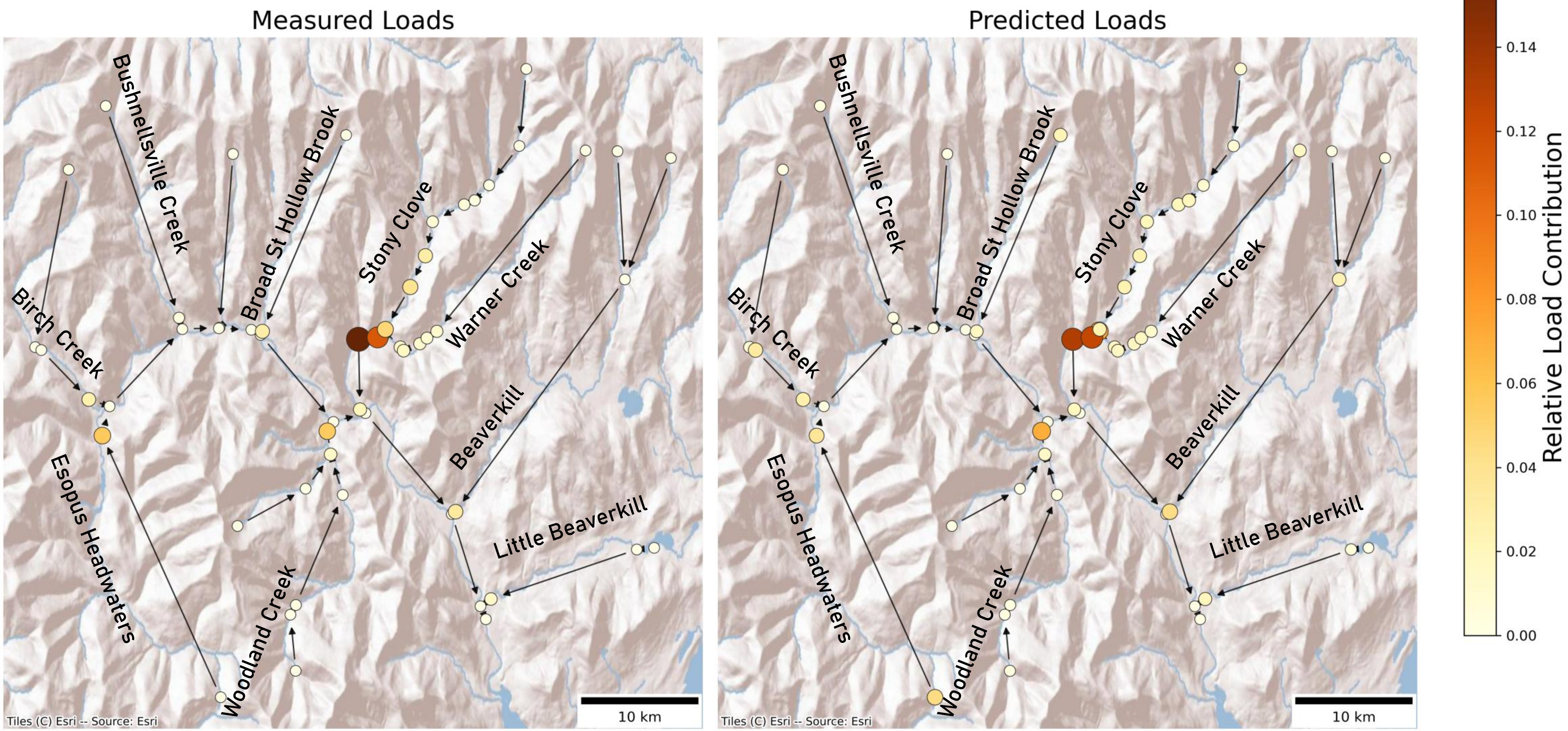

Total Precipitation
Precipitation (mm)
Discharge
Discharge (cfs)
Turbidity
Turbidity (FNU)
Measured
Predictions
Node Location in Spatial Graph of Upper Esopus Creek Watershed
Bushnell Creek at Shandanken
Esopus Creek at Allaben
Stony Clove at Chichester
Warner Creek near Chichester
Birch Creek
Esopus Creek at Oliviera
Woodland Creek at Phonecia
Beaver Kill at Mt Temper
Little Beaverkill at Beechford
Esopus Creek at Coldbrook
Tiles (C) Esri -- Source: Esri


Node 81
Total Precipitation
Precipitation (mm)
Discharge
Discharge (cfs)
Turbidity
Turbidity (FNU)
Measured
Predictions
Node Location in Spatial Graph of Upper Esopus Creek Watershed
Bushnell Creek at Shandanken
Stony Clove at Chichester
Esopus Creek at Allaben
Warner Creek near Chichester
Birch Creek
Esopus Creek at Oliviera
Woodland Creek at Phonecia
Beaver Kill at Mt Temper
Little Beaverkill at Beechford
Esopus Creek at Coldbrook
Tiles (C) Esri -- Source: Esri

Proportion of Tributary Turbidity "Load" Contribution

RMSE = 0.012

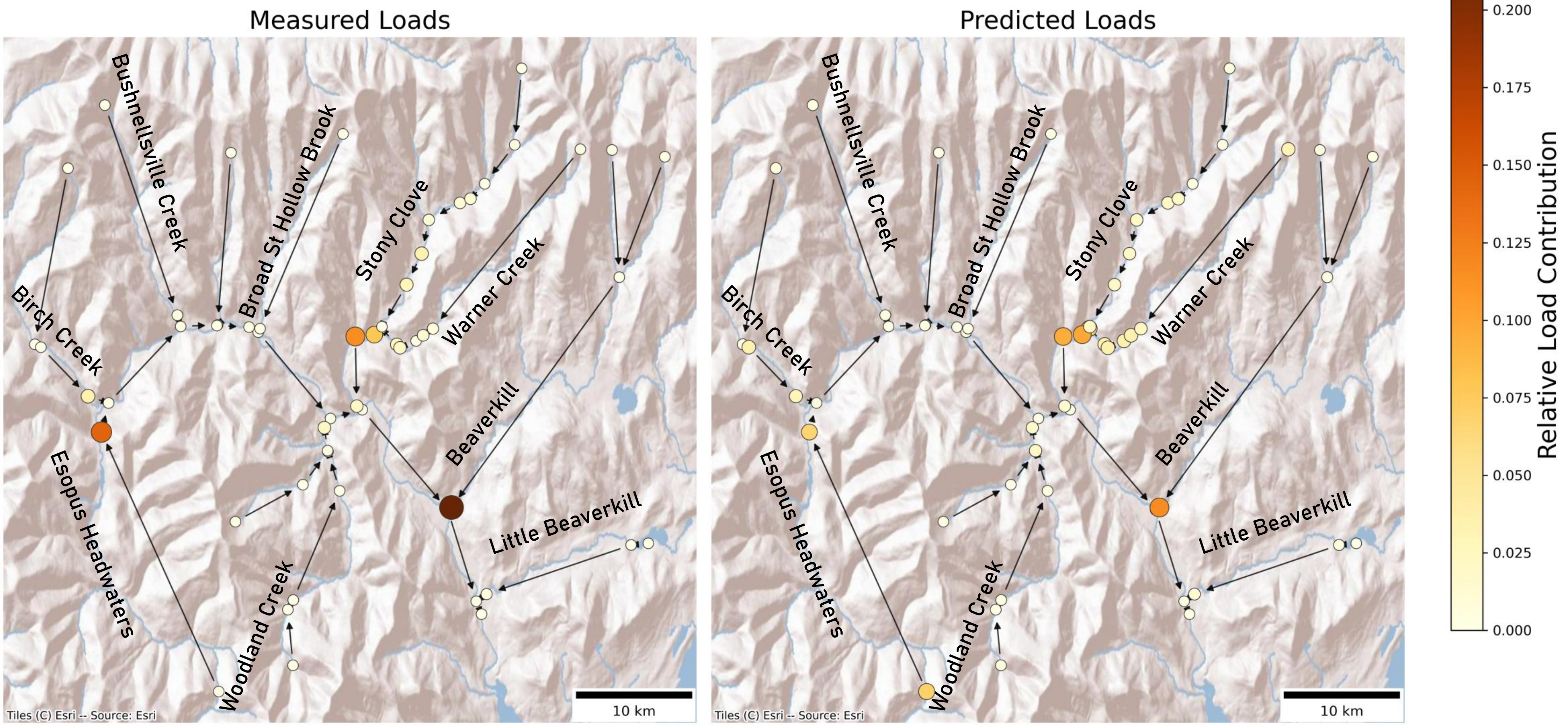
12/18/2023 - 12/19/2023



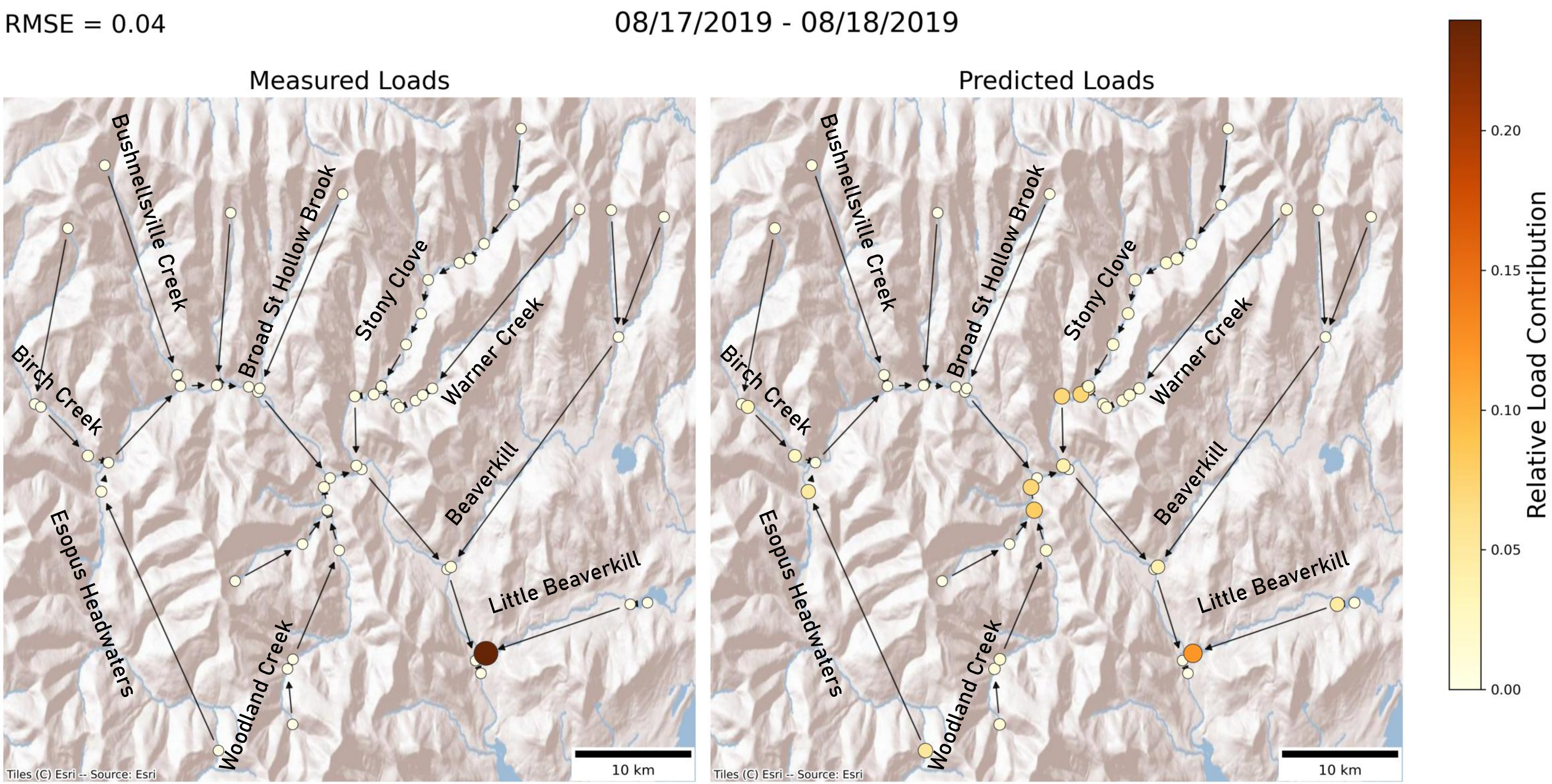
Proportion of Tributary Turbidity "Load" Contribution

08/17/2018 - 08/19/2018

RMSE = 0.031



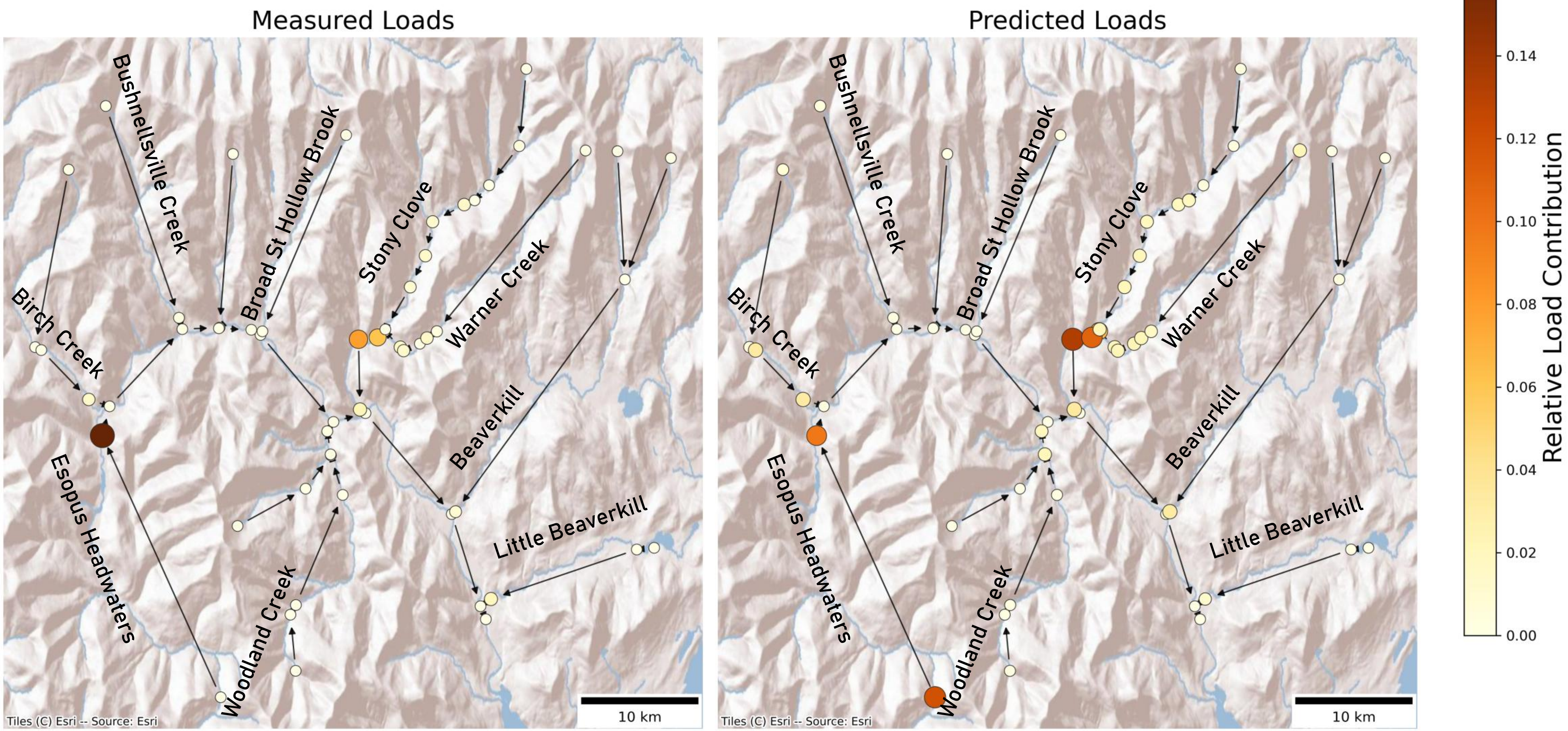
Proportion of Tributary Turbidity "Load" Contribution



Proportion of Tributary Turbidity "Load" Contribution

06/30/2017 - 07/03/2017

RMSE = 0.025



Model architecture and hyperparameters

- Temporal Graph Convolutional Network (TGCN) (Zhao et al., 2020)
- 80:20 temporal train: test split
- 32 hidden nodes in the TGCN, 32 hidden nodes in the regression head
- Learning rate: 0.001
- Batch size: 32
- Lookback period: 1 week
- Loss function: MSE
- Optimizer: Adam

Data Preprocessing

- Data were normalized from 0 – 1, then a small constant of 1e-5 was added to the entire dataset
- After the constant was added, missing data were filled with 0s
- The loss function masked out any missing turbidity values so they were not used to calculate loss and update model weights