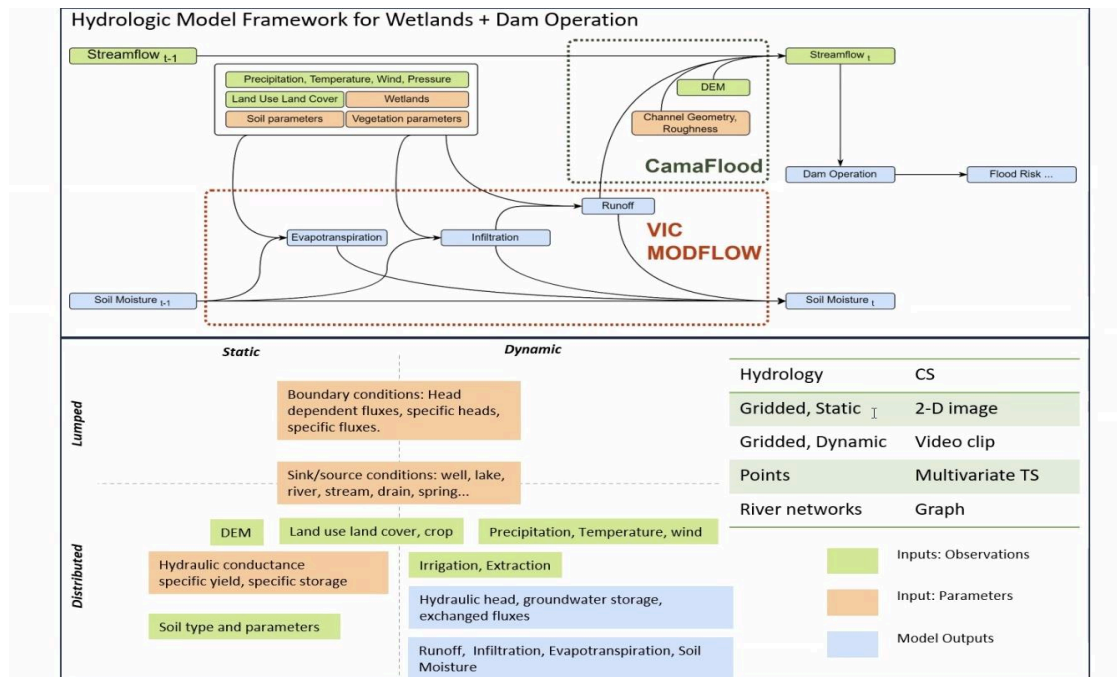# *Designing Nature to Enhance Resilience of Built Infrastructure*
(GR40695)
Quarterly Report - March 2024

Wetland Siting/Dam Operations--Brazos PMP

Activities

The Wetland Siting/Dam Operations team has been working to develop models (code), data, scientific papers and communications materials to build capacity in the Brazos River Basin to develop wetland construction and restoration projects. Our primary goal is to amplify the impact of and lengthen the lifespan of functionality of built infrastructure (USACE flood control reservoirs) and state storage reservoirs to provide water supply and flood protection in the face of climate change. To this end we are building surface water (VIC-4L) and groundwater (MODFLOW) models to understand total system storage and how it can be optimized with operations of reservoirs to deliver safe flood protection and enhance firm yield of conservation storage system wide. We work with the data science team to develop machine learning representations (emulation and imputation) of these complex physically based models. We also work with this team on optimization–both of spatial location of wetland sites and temporal release strategies that when combined with optimal siting of wetlands can enhance climate adaptation on the high- and low-end of water extremes in the river. We have built the full VIC-4L model and have the model setup for the MODFLOW model completed. We also have the data infrastructure to implement emulation (deep learning), spatial optimization (deep learning) and temporal optimization (non-linear integer programming). The illustration in this section represents the process flow of data for emulating VIC-4L using deep learning. We have submitted or published several papers on these methods. Currently we are gathering datasets to write two high-profile papers. The first paper will develop scientific indicators for evaluating the success and needs for water infrastructure–social and physical–for addressing climate adaptation. The second paper will develop a physics-informed prioritization for wetland siting and construction that allows climate adaptation engineers to deliver particular climate adaptation

functions–groundwater storage, flood control, surface water storage and environmental flows. This team works fluidly with the social science team to measure collaboration success of our research team and the interaction of our research team with a broad base of stakeholders. Finally, Sabo writes monthly in Forbes and has a regular podcast on topics relevant to this research effort.
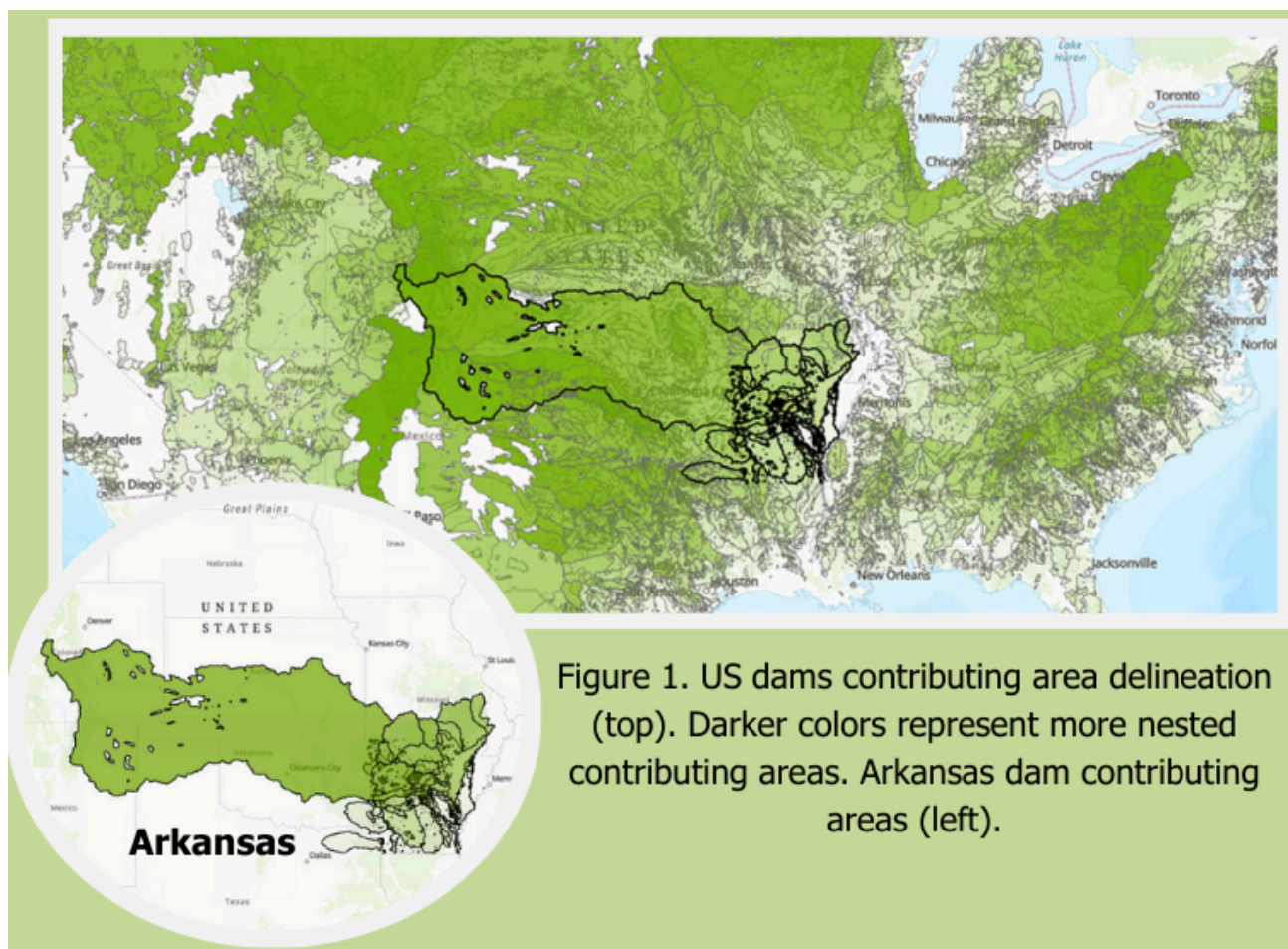
- In progress-Data collection for prioritizing sites to engineer wetland for flood mitigation, improving surface and groundwater security
- In progress-Basin scale rainfall-streamflow simulation model via spatial-temporal Graph Neural Networks (GNN)
- Completed simulating initial scenario for wetlands using physics-based computational model and its impact on flood peak.
- Constructed hybrid process-guided machine learning models that leverage strengths of hypothesis-driven physical model and data science
- Created and deployed survey to research team to measure collaboration success (internally)
- Developed a stakeholder map for potential interactions with a broad base of stakeholders including experts and non-experts and embracing DEI.

Outcomes
- VIC4L-CaMaFlood set-up
- MODFLOW setup
- Data- Simulated impact of wetlands on flood peak
- Published science paper- Shah R, Tsai Y, Stampoulis D, Damavandi H G and Sabo J 2023, Design principles for engineering wetlands to improve resilience of coupled built and natural water infrastructure. Environ. Res. Lett. 18 114045
- Huang, L. and J. Sabo (2023). "Physics Informed Machine Learning for modeling groundwater levels." American Geophysical Union (AGU) Fall Meeting.
- Over 12 column posts in Forbes: https://www.forbes.com/sites/johnsabo/?sh=3b4b7e379221 with one post receiving over 14K reads: https://www.forbes.com/sites/johnsabo/2023/08/22/a-scientists-take-on-oppenheimer-its-time-for-a-constructive-not-destructive--national-lab/?sh=3a7bfaa46e32
- Six episodes of Audacious Water (the podcast) in a season on nature, infrastructure and water in the Mississippi River Basin: https://www.audaciouswater.org/ including one episode interview with Todd Bridges https://www.audaciouswater.org/episodes/todd-bridges-us-natural-infrastructure-strategy

Team
- John Sabo, PhD
- Reepal Shah, PhD
- Li Huang, PhD
- Qi Deng, PhD

Ag/WQ PMP

Figure 1. US dams contributing area delineation (top). Darker colors represent more nested contributing areas. Arkansas dam contributing areas (left).

Activities

We have three primary research projects we are working on this PMP. First, we are in the mid-point of a data scoping project where we are quantifying the extent of watersheds in the U.S. that have dams, agriculture, and subsequent water quality concerns. We have completed the first half of this project (as described in outcomes) and are continuing on with the incorporation of water quality data into the scoping to complete the analysis. Second, we are working closely with the Data PMP group to develop and test a causal learning framework to inform water quality modeling in these complex watersheds. Our case study watershed is the Trinity River basin in Texas where there is both extensive agriculture and multiple large dams and reservoir systems. Third, we have identified that a key gap in doing both process-based or data-driven modeling of water quality in these complex systems is a lack of water quality observations (especially temporally) to train the models. Therefore we are working to develop a graph neural network approach to "fill in the gaps" of water quality observation predictions. We are testing this approach in the Maumee River watershed in the Midwestern U.S. because it is one of the few areas where there is significant spatial and temporal coverage of water quality.

Outcomes

Significant outcomes of this project include:

- The training of 2 postdoctoral and 1 doctoral scholars

- The development of a code to delineate contributing areas for all dams in the National Inventory of Dams (NID) using stream network data from the USGS
- A major finding that more than 41% of dams in the NID have a contributing area that is greater than 50% agricultural. This finding is important because it highlights the extent of "complex systems" within the U.S., furthering the need for coordinated modeling and management of these systems.
  - We currently have one paper in preparation for this outcome
- We have presented our current findings in six conference presentations:
  - Liu T, Deng Q, Ding K, Sabo JL, Liu H, Candan KS, Muenich RL. Applying Graph Neural Networks to Improve the Data Resolution of Stream Water Quality Monitoring Networks. American Geophysical Union Annual Meeting, December 12-15, 2023, San Francisco, CA.
  - Villarreal D, Liu T, Deng Q, Steissberg T, Sabo JL, Muenich RL. Do upland agricultural areas and downstream reservoirs counteract efforts to improve water resources? American Geophysical Union Annual Meeting, December 12-15, 2023, San Francisco, CA
  - Muenich RL. Changing the Paradigm of Water Quality Modeling in the Age of Big Data and Machine Learning. Arkansas Water Resources Conference, July 18-20, 2023. Fayetteville, AR
  - Steissberg TE, Muenich RL. Integrating water resources infrastructure with upland management to advance nature-based solutions for water quantity and quality. International Soil and Water Assessment Tool Conference, June 26-30, 2023, Aarhus, Denmark
  - Arslan B, Mandal P, Sheth P, Muenich RL, Candan KS. Streamflow Prediction using Spatiotemporal Deep-Learning-based Framework with Reinforcement Learning. HydroML Symposium, May 22-24, 2023, Berkeley, California.
  - Deng Q, Liu T, Sheth P, Muenich RL, Sabo J. Time-Space Transformer for Agriculture Evaluation Applied to Streamflow Simulation. HydroML Symposium, May 22-24, 2023, Berkeley, California
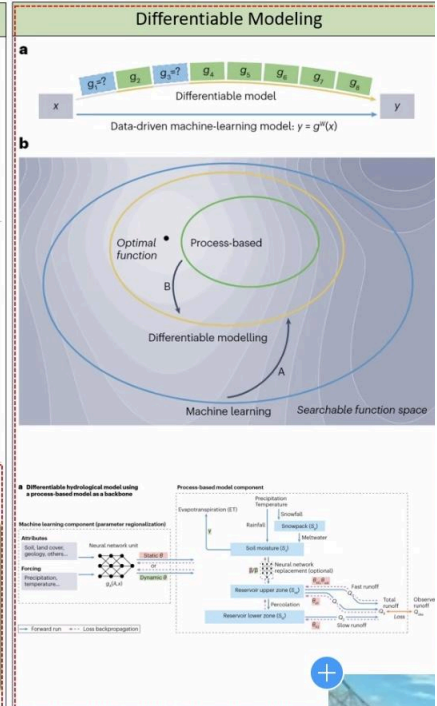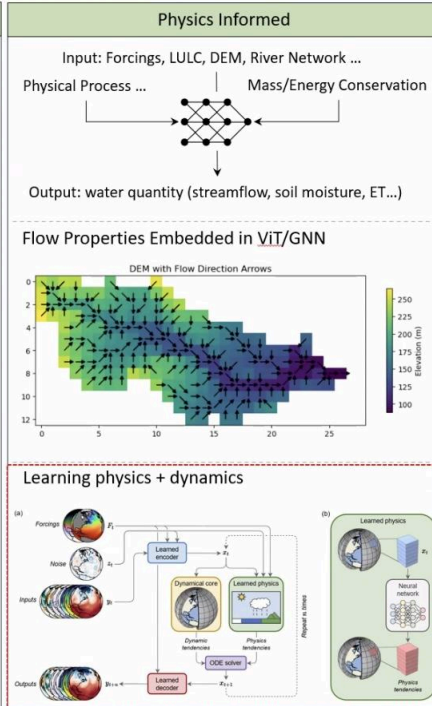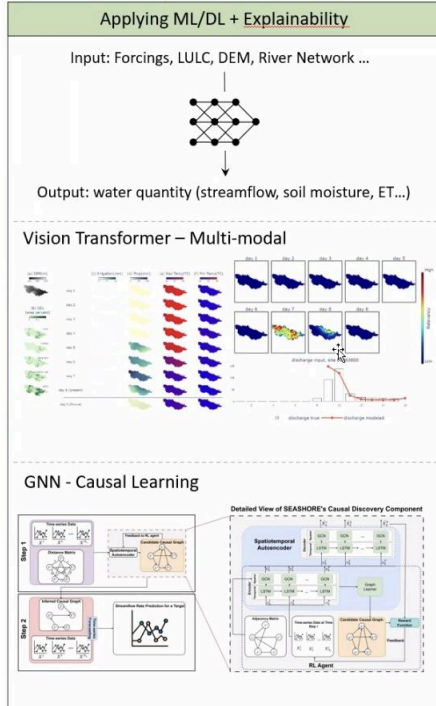
Team
- Rebecca Muenich, Ph.D.
- Selcuk Candan, Ph.D.
- Ting Liu, Ph.D.
- Danna Villarreal
- Qi Deng, Ph.D.
- Kaize Ding, PhD

Flood-MAR PMP

Activities

We are working to develop three CONUS-wide geospatial datasets that will be used to calculate the FloodMAR suitability index.

- Surface water availability - We are developing a statistical method for predicting the volume and duration of high flow events based on basin characteristics and the historical reanalysis of NOAA's National Water Model. Using this method we can estimate the average annual volume of water available for FloodMAR in ungauged stream reaches.
- Aquifer Storage Potential - A machine learning algorithm model has been trained using observed groundwater levels at monitoring wells to simulate groundwater table response using climate variables, land use land cover change, physiographic properties (e.g., topography, soil texture), and with and without remotely sensed products of terrestrial water storage and fluxes. The machine learning model is being used to create monthly gridded groundwater table depth across CONUS including areas with no or scarce groundwater level observations. Areas with deep water tables have more capacity for subsurface storage, and areas in which the water table has experienced recent drawdown are particularly well-suited for MAR.
- Land Surface Suitability – A map of suitable land areas for FloodMAR is being delineated based on the slope, soil permeability, and land use.
- Working closely with the Social Science PMP team to study factors that affect the social suitability of FloodMAR, including leveraging perspectives and feedback from USACE and other floodplain managers

Outcomes

- The statistical method for predicting volume and duration of high flow events has been coded in Python, including a debiasing procedure. The method and its results will be published in a peer-reviewed paper.
- A Python script has been developed to select the NHD stream reach closest to the outlet of each HUC8 in the Watershed Boundary Dataset.
- The machine learning algorithm has been trained and deployed for the Western United States to estimate monthly groundwater depth. We are currently finetuning the model before deploying it for the entire CONUS. This work will be published in a peer-reviewed paper.
- The methodology for delineating suitable land areas has been determined based on a literature review and scripted in Python.
- We have had several meetings with the USACE Institute for Water Resources to gather feedback on the project plan and ensure it aligns with USACE policy goals as stated in WRDA 2022
- We have presented our current findings in the following conference presentation:
  - Tiwari, S., and G. Mascaro, G. Quantifying the high-flow volume availability for flood managed aquifer recharge across the Continental United States. AGU Fall Meeting 2023, December 2023.

Team
- Aubrey Harris, PE
- Aaron Byrd, PhD
- Daniel Siegel
- Glen Low
- Giuseppe Mascaro, PhD
- Suraj Tiwari
- Tianfang Xu, PhD
- Qinyuan Dai

Social Science PMP

Activities

The Social Science team has developed several activities across different PMPs
- Designed data collection methods:
  - Designed a survey to understand diverse perspectives to develop a feasibility-level tool for siting a flood management alternative: Flood-MAR (Managed Aquifer Recharge).
    - USACE sample (Water Managers)
    - External Sample (State, county, city government; irrigators; conservation groups)
  - Conducted 5 cognitive interviews with USACE stakeholders to ensure survey was appropriately capturing perceptions of, and concerns about, Flood-MAR
  - Designed data collection protocol for conducting participant observation activities to explore how the collaborators behave, cooperate, and exchange ideas for developing research tools.
  - Designed and applied online survey to compare how science is used in decision-making contexts.
  - Designed participatory modelling protocol for for conducting interviews and participate in workshops who work on wetland reservoir
systems and downstream water resources infrastructure
- Attended conferences
  - National Conference on Ecosystem Restoration (NCER), Assessing the Social Suitability of Managed Aquifer Recharge Sites.
- Academic manuscripts
  - MAR scoping review. The aim is to comprehensively analyze the scope of social dimensions associated with Managed Aquifer Recharge.

Outcomes
- Data collection tools:
  - Survey to understand USACE perceptions about the feasibility of FloodMAR
  - Survey to understand non-USACE stakeholders perceptions of FloodMAR.
  - Online survey to explore how science is used in decision-making context
  - Stakeholder's network protocol
  - Interview protocols for stakeholder analysis
  - Workshops protocols for stakeholder analysis
- IRBs
  - ASU approval for collecting FloodMAR survey
  - ASU approval for conducting participant observation
  - ASU approval for collecting collaborators survey
  - ASU approval for conducting stakeholder analysis
- Miscellaneous
  - Created an onboarding packet

Team
- Amber Wutich, PhD
- Melissa Beresford, PhD
- Margaret du Bray, PhD
- Jelena Jankovic-Rankovic, PhD
- Laura Castro-Diaz, PhD
- Incoming Postdoc: Cara Jacob, PhD

Data Science PMP

Finished/Ongoing Machine Learning Models/Applications

Aa Application	Model	Input	Output	ForecastWindow	Scale	Region
Rainfall-Streamflow Forecast	VisionTransformer	ST-Forcing ST-LULC S-DEM MultiVar-Streamflow G-Rivernetwork	MultiVar-Streamflow	1D 7D	HUC8 1D 1KM	Trinity
Runoff-Streamflow Fo ▣ OPEN	GNN	ST-Runoff S-DEM SingleVar-Streamflow	SingleVar-Streamflow	1D	HUC4 1D 6KM	Brazos
Streamflow Routing Forecast	Transformer	MultiVar-Streamflow	MultiVar-Streamflow	72H	HUC4 1H	US
Water Yield Causal Learning	GNN	SWAT-Inputs SWAT-Simulations	WaterYield	1D	HUC4 1D	Trinity
MultiVar-Interpolation	GNN	ST-Forcing ST-LULC S-DEM MultiVar-Streamflow G-Rivernetwork MultiVar-WQ	MultiVar-WQ	1Y	HUC4 1D	Maumee
Rainfall-Streamflow Forecast	GNN	ST-Forcing ST-LULC S-DEM MultiVar-Streamflow G-Rivernetwork	MultiVar-Streamflow	1D 1Y	HUC4 1D 6KM	Brazos
Rainfall-streamflow Forecast	VisionTransformer	ST-Forcing ST-LULC S-DEM MultiVar-Streamflow G-Rivernetwork	MultiVar-Streamflow	1Y	HUC4 1D 4KM	
Differentialable-SWAT	SWAT	SWAT-Inputs	SWAT-Outputs	1Y	HUC4 1D	Trinity

Activities

During the past period, we have worked on multiple projects related to the N-EWN effort.

- Our first project involved predicting the streamflow in the Brazos River basin using a novel spatiotemporal Deep Learning-based framework. This work was published in the proceedings of HydroML 2023. We are currently collaborating with Dr. Reepal Shah and Dr. Qi Deng from the ByWater Institute at Tulane University on research focused on spatiotemporal causal learning for predicting streamflow. We leveraged physics-based simulations (CaMa Flood) to identify the contextually meaningful definition of what constitutes upstream for a chosen dam using a Long short-term memory (LSTM) network-based Autoencoder with Attention Mechanism. Our approach involves using the river flow graph as prior knowledge and creating a spatiotemporal graph neural network to forecast the streamflow rates at target sites. Our work has been recognized and will be presented at the inaugural N-EWN symposium scheduled for May this year.
- We developed the Spatio-temporal causal discovery algorithm aided by reinforcement learning for streamflow rate prediction. Our study introduces STREAMS, a novel framework combining Reinforcement Learning (RL) with an LSTM-GCN based autoencoder for spatiotemporal causal discovery in streamflow rate prediction. Recognizing the importance of accurate streamflow forecasting for reservoir management and environmental impact assessment, we tackle the complexity of modeling such systems. STREAMS is designed to identify crucial watershed outlets and their interactions, enhancing hydrological resource management through improved prediction accuracy. The framework's efficacy is validated through extensive experiments on the Brazos River Basin. These tests confirm STREAMS's capability in optimizing the search space for causal discovery and inferring spatiotemporal causal features, thereby offering a substantial advancement in streamflow modeling. Our results demonstrate the potential of leveraging advanced machine learning techniques for efficient and informed water resource management. We published this work in the CIKM'23 conference and presented it in October'23.
- The Wetland Identification project focuses on prioritizing potential wetlands using various datasets, including the National Land Cover Database (NLCD), the Soil Survey Geographic

Database (SSURGO), and the Height Above Nearest Drainage (HAND) layer. These resources provide comprehensive information on land cover, soil, and drainage, essential for identifying potential wetland areas, particularly within the Brazos River Basin across Texas and New Mexico. The project aims to train models for downstream tasks such as flood risk reduction and water storage enhancement, leveraging features like soil characteristics, flood/drainage frequency, and the HAND layer. Initially, the algorithms aim to improve flood risk predictions by leveraging characteristics from the soil dataset and employing feature processing techniques, such as neighbor feature aggregation and contrastive learning with aggregated features. The challenges encountered during feature processing, especially with high-dimensional data, underscore the importance of normalization, embedding, and addressing class imbalance. The iterative process of testing various models, adjusting loss weights, and experimenting with soil feature vectorization highlights the project's dynamic approach to refining its predictive accuracy and effectiveness in reducing flood risks. Furthermore, the model seeks to identify potential wetlands using multiple features using knowledge transfer (domain disentanglement), graph neural networks, and contrastive learning to address the challenge of the limited availability of wetland labels in specific areas. the Wetland Identification project represents a pioneering effort in environmental conservation and management, harnessing the power of cutting-edge computational techniques and diverse datasets to tackle the complexities of wetland identification and flood risk mitigation. By integrating the NLCD, SSURGO, and HAND layer data with sophisticated modeling approaches, the project not only aims to enhance our understanding of wetland dynamics but also to pave the way for innovative solutions in water management and ecological preservation. The challenges encountered in processing and analyzing high-dimensional data have prompted the development of robust methodologies, highlighting the project's commitment to precision and efficiency. As the project progresses, its outcomes are expected to contribute significantly to the conservation of wetland ecosystems, demonstrating the critical role of technology and data science in addressing environmental challenges. This endeavor not only exemplifies interdisciplinary collaboration but also sets a benchmark for future initiatives in the era of environmental science and conservation.

- We are exploring the use of the DataStorm project for simulating traditional hydrological models such as the Soil & Water Assessment Tool (SWAT). This can open new avenues in hydrological research by allowing us to simulate alternate timelines for more robust predictions. Furthermore, the parameter tuning process for hydrological models often involves manual calibration based on domain knowledge which may often be time-consuming and error-prone. It may be possible to use our proposed technique to tune existing hydrological models more efficiently.

- We are further investigating the use of causal information to improve the efficiency of skyline discovery for decision support queries. The skyline operator is used to find the Pareto-optimal frontier for a dataset. Hydrological data often contains underlying causal structures that we are attempting to exploit for more efficient skyline discovery. For instance, if we want to detect which crops to grow for the highest yield while using the lowest amount of fertilizers and pesticides, underlying causal variables such as the season or annual rainfall might allow us to make the discovery process more efficient.

- We are applying spatio-temporal graph learning to the problem of water stream data imputation. Specifically, we are tackling this problem from two aspects: (1) transfer the knowledge from one domain to another via spatio-temporal graph domain adaptation, (2) develop spatio-temporal graph diffusion model to improve the imputation capability of the learning backbone. We have experimented with series of baseline methods for each of the line of research and implemented our prototype models (e.g., https://github.com/keyangds/Domain-Adaptation), showing that domain adaptation and

diffusion model are able to improve the data imputation performance. We have successfully delivered one oral presentation based on our initial results on the AGU conference.

- In the pursuit of advancing hydrological modeling, the challenge of data scarcity across diverse regions stands as a significant barrier. This issue is particularly acute in the field of hydrology, where the availability of comprehensive datasets is not uniform across all geographical areas. Addressing this challenge, our research introduces a groundbreaking causality-aware disentanglement model, designed to transcend the limitations posed by data scarcity through innovative machine learning techniques. At the core of our model lies the ability to separate input data into two distinct categories: domain-invariant (causal) factors and domain-dependent factors. This disentanglement process is pivotal, as it allows our model to isolate and leverage causal features that hold true across different regions, thereby enabling the transfer of knowledge from data-rich areas to those where data is sparse. Such an approach not only enhances the model's predictive accuracy but also its applicability in a wide range of hydrological scenarios. The practical implications of our causality-aware model are vast and varied. For instance, in flood forecasting, the model could utilize data from regions with extensive historical records to predict flood risks in areas where such data is limited. Similarly, in predicting drought patterns, the model's ability to focus on causal relationships could improve predictions in regions that have historically been under-monitored. This model's versatility was further demonstrated through its successful application in hate speech detection, as published in WSDM'24 and presented in March'24. This accomplishment underscores the model's potential beyond its initial hydrology-focused applications, suggesting its broader utility in addressing challenges where generalization across diverse domains is crucial.
- Additionally, we are creating a general framework for representing spatiotemporal causal knowledge to build better causal discovery algorithms. The intricate spatiotemporal dynamics of complex systems are not well captured by traditional approaches, such as causal graphs and structural equations. Thus, it is imperative to have a robust causal framework that is rich enough to capture the complex causal relationships present in hydrological data. A vision of our proposed framework is currently under review at the Association for Computing Machinery (ACM) Transaction on Spatial Algorithms and Systems journal. We note that modeling causal relationships over patio-temporal data is a challenging issue. The most commonly used model, directed acyclic graphs (DAGs), are generally based on Bayesian causality. However, there are several issues with DAG representations and these shortcomings present a need for an alternative model. We have been investigating different representations of causality for better analysis and modeling of causal systems. We have therefore been further investigating Petri Net based representations — more specifically, we have been developing algorithms for mining petri nets from multivariate temporal data and providing an alternative representation of a causal system.

Outcomes
- Paras Sheth, Ruocheng Guo, Lu Cheng, Huan Liu, Kasim Selçuk Candan. Causal Disentanglement for Implicit Recommendations with Network Information. ACM Trans. Knowl. Discov. Data 17(7): 94:1-94:18 (2023)
- Paras Sheth, Ahmadreza Mosallanezhad, Kaize Ding, Reepal Shah, John Sabo, Huan Liu, K. Selçuk Candan. STREAMS: Towards Spatio-Temporal Causal Discovery with Reinforcement Learning for Streamflow Rate Prediction. CIKM 2023: 4815-4821
- Ting Liu, Qi Deng, Kaize Ding, John L Sabo, Huan Liu, Kasim Selcuk Candan, and Rebecca Logsdon Muenich. Applying Graph Neural Networks to Improve the Data Resolution of Stream Water Quality Monitoring Networks. AGU 2023.
- Domain adaptation models, https://github.com/keyangds/Domain-Adaptation

- Bilgehan Arslan et al. Streamflow Prediction using SpatioTemporal Deep-Learning based Framework with Reinforcement Learning. HydroML poster, 2023.
- F. Azad et al. A Vision for Spatio-Causal Situation Awareness, Forecasting, and Planning. ACM Transaction on Spatial Algorithms and Systems, Under Review, 2023.
- Shu Wan, Reepal Shah, Qi Deng, John Sabo, Huan Liu, K. Selcuk Candan, Spatiotemporal Causal Learning for Streamflow Forecasting, accepted for presentation at the N-EWN Symposium, 2024.
- P. Mandal, Y. Choi, R. Shah, J. Sabo, H. Liu. K.S. Candan. Identifying Potential Wetlands via Causality-based Data amputation and Knowledge Transfer. , accepted for presentation at the N-EWN Symposium, 2024.

Team (not all funded through this contract)
- K. Selcuk Candan (Prof.)
- Huan Liu (Prof.)
- Kaize Ding (Asst. Prof.)
- Yoonhyuk Choi (PostDoctoral Researcher)
- Bilgehan Arslan (PostDoctoral Researcher)
- Paras Sheth (PhD Student)
- Pratanu Mangal (PhD Student)
- Shu Wan (PhD Student)
- Ahmet Kapkic (PhD Student)
- ..and members from other PMPs