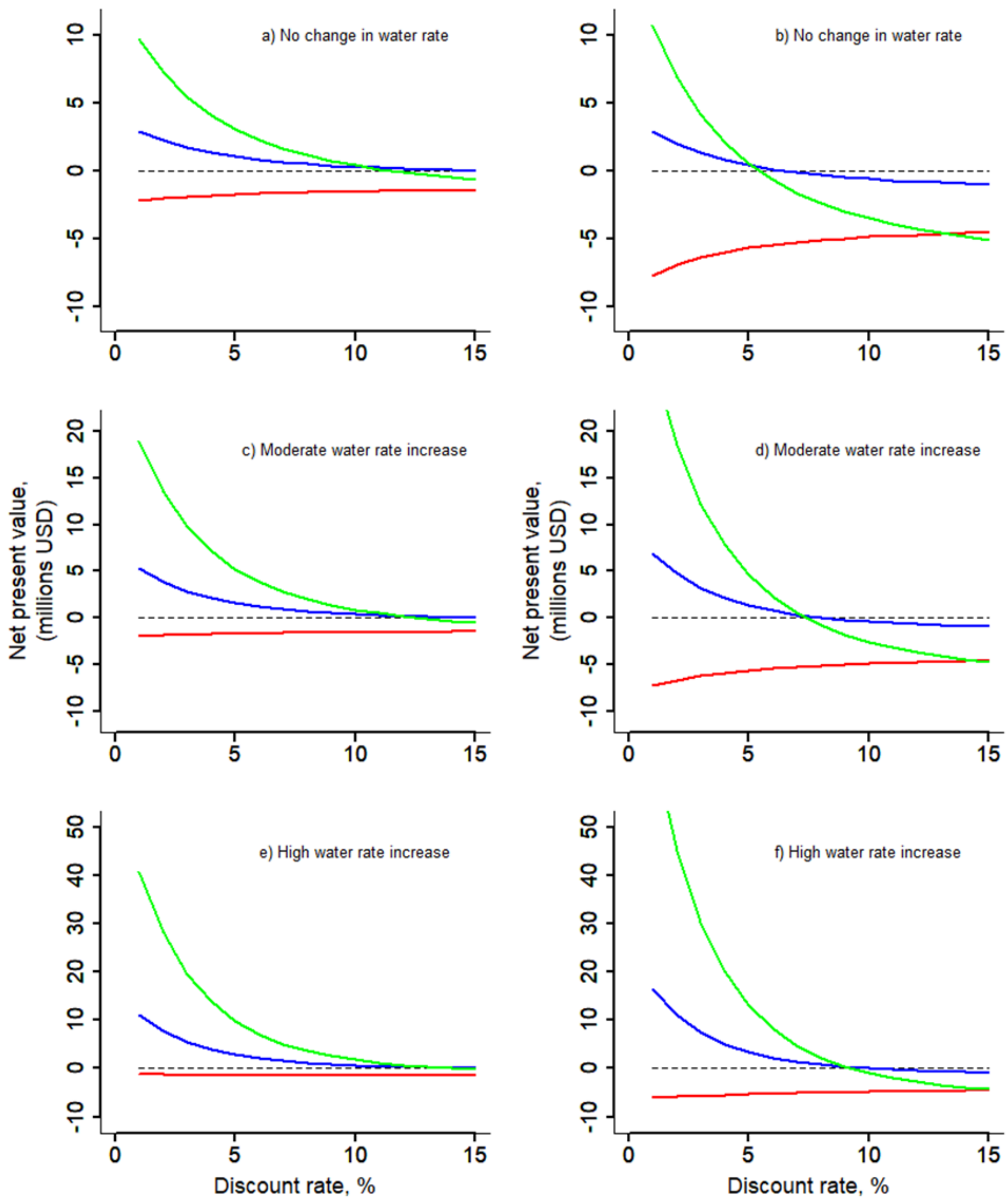
Designing Nature to Enhance Resilience of Built Infrastructure (GR40695)

Quarterly Report - June 2024

Wetland Siting/Dam Operations--Brazos PMP



Financial sensitivity analysis. Net present value (millions of USD) plotted as a function of the discount rate for no change in water price (top row), low increases in water price (middle row) and large increases in water price (bottom row) and for small projects (10 km², left) and large projects (30

km², right). Line colors represent wetland created in upstream (red), midstream (blue) and downstream (green) part of the Brazos basin.

Activities

1. Building a plug-and-play tool to facilitate geospatial analysis for identifying potential sites for creating/restoring wetlands in the Brazos basin for multiple objectives: increasing floodplain storage (to reduce flood risk) and increasing groundwater recharge and enhancing biodiversity in the Brazos River Basin of Texas.

2. Developing framework to connect physical model with economic analysis, to estimate cash flow from wetland creation in different section of the basin.

3. We are building a multi-modal deep learning model using Vision Transformer architecture for rainfall-streamflow simulation against benchmark models including ConvLSTM and GNN, where unmatched spatial resolution inputs (as meteorological forcings, DEM, LULC) can be taken into account at once.

Outcomes

1. We have created maps of potential sites for wetlands for objectives increasing floodplain storage (at 1km x 1 km) and increasing ground water recharge (at 30m x 30 m) using physical variables (such as soil characteristics, land use land cover, aquifer characteristics, etc).

2. We found that net positive cash flow originating from avoided costs of water tariffs offset or exceeded capital costs of modestly sized wetlands built on low value land leading to water security solutions that generate revenue. Hence building back nature can be profitable.Our following work is in review:

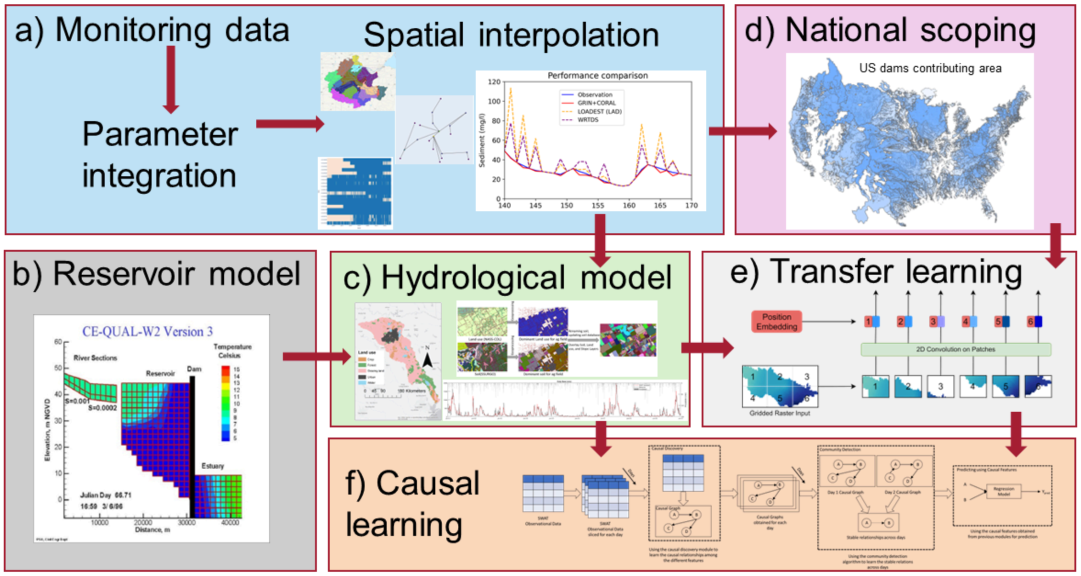
   Gao Hongkai, Shah Reepal, Yamzazki Dai, Finley Tim, Bohn Ted, Low Glen, and Sabo John, (2024), “Coordination of natural and built infrastructure to better manage extreme events in coastal water systems.”, Environmental Research Letters.

3. We have a version 1.1 model running with 4 KM spatial resolution and 1-day temporal resolution on HUC8 scale level with spatial temporal inputs including precipitation, min/max temperature, DEM, river flow direction graph.

Team

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

Ag/WQ PMP



Roadmap and latest visual results from AG/WQ PMP

Activities
This PMP aims to study the concurrent impact of agricultural field management and dam operations on river water quantity and quality. It will also explore incorporation opportunities between agricultural field management and dam operation improvement for water quantity and quality goals. To achieve the above goals, we are conducting national-level scoping and screening work on agricultural fields and dams to find watersheds with potential agricultural and dam management improvement needs. Meanwhile, we are conducting watershed-scale studies on the mechanisms of how agricultural management and dam operation impact water quantity and quality using both physical process-based and data science techniques.

On the national level:

- National data scoping - We have developed an automated tool to generate upstream watersheds of major artificial dams for the whole US, based on the latest USGS NHDPlus watershed and river network database.
- Transfer hydrological learning - We are testing convolutional machine learning methods such as spatial emulation to transfer hydrological simulation results across different watersheds.

On watershed-level:

- Trinity River SWAT model - We have built a near-fully calibrated physical-based Soil & Water Assessment Tool (SWAT) water quantity and quality hydrological model in Trinity River Basin, TX. We selected this basin as a test region for an agricultural dam management integration study as this watershed is dominated by agricultural land uses and contains 15 major reservoirs. The model is developed at a high resolution on agricultural lands and a lower resolution in urban areas to improve model fidelity and help identify priority areas for agricultural water quality management. The field-level resolution at agricultural land allows customized input on agricultural management activities.
- Reservoir-SWAT model integration - Through collaboration with N-EWN and LimnoTech, we integrate the numerical reservoir USACE CE-QUAL-W2 model into the SWAT model to improve the flexibility and precision of simulating dam controls and reservoir water dynamics.

- Spatial-temporal causal learning framework - Through cross-efforts with the Data Science PMP, we are building a temporal-spatial causal learning framework that allows us to study the causal relationships between land and soil, agricultural management, dam operation, and water quantity and quality parameters described and calibrated by the SWAT model.
- Water quality parameter integrator - Currently, USGS and US EPA integrated water quality stations have over 20 parameters that describe the concentration of nutrient pollutants in water due to the species and measurement differences. We are building integrating tools that target maximizing the use of these parameters and simplifying them into common model parameters.
- Water quality interpolation with GNN - Water quality measurements in most watersheds are sparsely distributed in temporal and spatial scales, which leads to incomplete records at any time or location point. We apply graphical neural network (GNN) methods to interpolate the water quality measurement in spatial-temporal dimensions to support national-level studies as well as the SWAT model calibration.

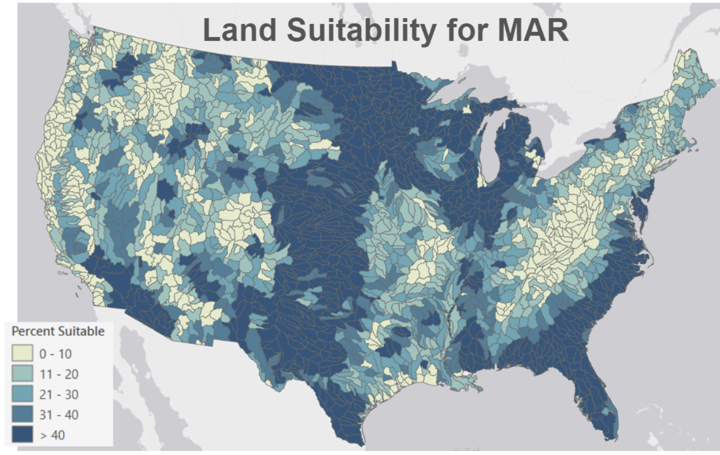
Outcomes

- National data scoping - We have finished a national dam upstream watershed map. Poster presented the latest progress in the N-EWN 2024 Symposium with a student poster award. Currently preparing result analysis for academic journal publication.
- Transfer hydrological learning - We have proven better accuracy using spatial emulation methods than traditional ensembled methods. Currently under calculation efficiency improvement.
- Trinity River SWAT model - Water quantity is fully calibrated, and water quality requires slight improvement waiting for the completion of the water quality parameter integrator before publication. Poster presented the latest progress in the N-EWN 2024 Symposium.
- Reservoir-SWAT model integration - Our collaborator has finished testing this integration in one of the 15 major reservoirs in the Trinity River basin.
- Spatial-temporal causal learning framework - We have proved the effectiveness of the methodology by qualitative tests. Currently developing a universal model with the Data Science PMP.
- Water quality parameter integrator - Database completed. Currently developing Python-based package-style application.
- Water quality interpolation with GNN - Proved a better performance on water quality interpolation than current popular tools. Currently improving the algorithm in preparation for an academic journal publication.

Team

- Rebecca Muenich, Ph.D.
- Selcuk Candan, Ph.D.
- Ting Liu, Ph.D.
- Danna Villarreal
- Qi Deng, Ph.D.
- Kaize Ding, Ph.D.
- Paras Sheth (Ph.D. Student)
- Pratanu Mangal (Ph.D. Student)
- Todd Steissberg, Ph.D.
- Billy Johnson, Ph.D.

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 in ungauged watersheds. Multiple sources of flowrate data are being evaluated, including the National Water Model, Nearest-Neighbor Drainage Area Ratio (NNDAR), Map-Correlation Drainage Area Ratio (MCDAR), and Ordinary Kriging of the logarithms of discharge per unit area (OKDAR).
- 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.
- 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 machine learning algorithm for groundwater level interpolation has been trained and deployed for the entire CONUS. Once it has been validated and refined, this work will be published in a peer-reviewed paper.
- Poster on groundwater level interpolation work won an award at the Biennial Symposium on Managed Aquifer Recharge in Tuscon, AZ.
- Suitable land areas for MAR have been delineated across the entire CONUS, and the land suitability metric calculated for each region (see figure above). This data and methodology will be published in a peer-reviewed paper.
- 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:
    - Siegel, D. A CONUS-wide evaluation of the best locations for Managed Aquifer Recharge using floodwater (FloodMAR). Biennial Symposium on Managed Aquifer Recharge, April 2024.
    - Dai, Q. Estimating Groundwater Depth and Storage Responses to Climate Variability and Human Activities in Western U.S. Biennial Symposium on Managed Aquifer Recharge, April 2024.
    - Siegel, D. A CONUS-wide evaluation of the best locations for Managed Aquifer Recharge using floodwater (FloodMAR). AWRA 2024 Geospatial Water Technology Conference, April 2024.
    - Siegel, D., and Harris, A. A CONUS-wide evaluation of the best locations for Managed Aquifer Recharge using floodwater (FloodMAR). N-EWN Parner Symposium, May 2024.

Team

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

Participatory Research PMP

Activities

The Social Science team has developed several activities across different PMPs

Conferences

- Assessing the Social Suitability of Managed Aquifer Recharge. National Conference on Ecosystem Restoration. Albuquerque, New Mexico. April 14-19, 2024
- Assessing the Social Suitability of Managed Aquifer Recharge. N-EWN Partners Symposium. Saint Augustine, Florida. May 22-24, 2024

Academic manuscripts

- First draft MAR scoping review. The aim is to comprehensively analyze the scope of social dimensions associated with Managed Aquifer Recharge
- Manuscript in preparation: Participatory methods & MAR review.
- Outline Climate adaptation paper (with Brazos team)

Participatory methods class:

- Planned 6 modules for the methods class:
    - Community-Based Participatory Research
    - Participatory Mapping
    - Participatory Modeling

    - Citizen Science
    - Photovoice
    - Art-Based Methods
- Met with Arizona Water Innovation Initiative and Learning Enterprise (ASU continuing education platform) representatives to discuss co-branding with EWN and plan course structure
- Learning Enterprise planning and negotiation in process
- Met with Courtney Chambers at EWN to discuss co-branding and deliverable requirements

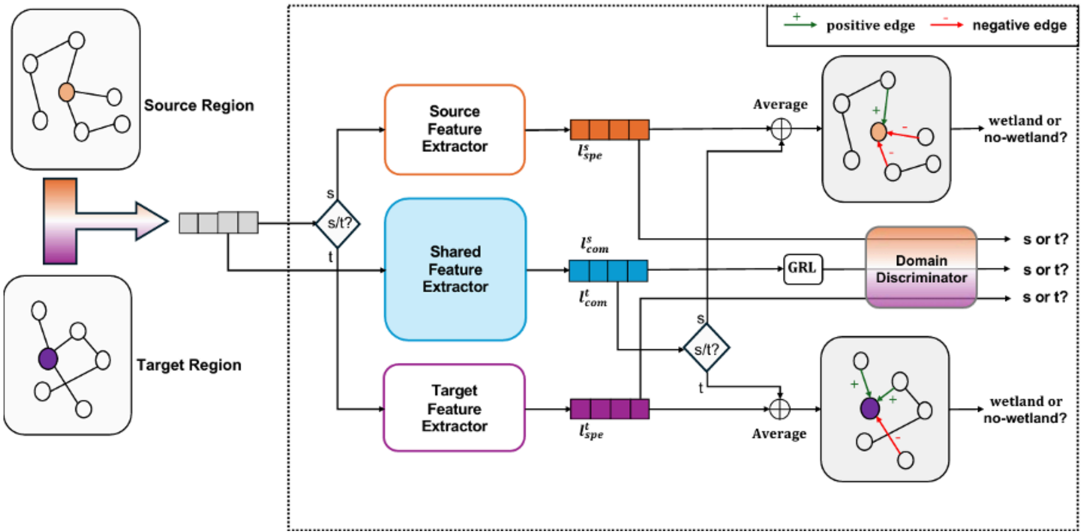
MAR IRB
- Supporting process for IRB and OMB approval for launching the MAR survey

Outcomes
- Work Plan for Participatory Methods Course
    - Curriculum planned
    - Co-branding in progress

- Meta-Analysis workshop
    - Protocol for conducting reviews
    - Paper 1. Draft done. MAR scoping review- Social Dimensions associated with MAR
    - Paper 2. In preparation, Participatory methods & MAR
- IRB
    - Modification of FloodMAR survey approved by ASU
    - Internal USACE survey approved

Team
- Amber Wutich, PhD
- Melissa Beresford, PhD
- Margaret du Bray, PhD
- Jelena Jankovic-Rankovic, PhD
- Laura Castro-Diaz, PhD
- Cara Jacob, PhD
- Oswaldo Medina, PhD

Data Science PMP



Accuracy gains due to knowledge transfer: gains are higher when the knowledge is transferred from denser regions (LA and FL) to sparser regions (TX and AZ)

		Target						
		TX:0.1%	AZ:0.6%	OR:1.5%	WA:2.2%	LA:8.3%	FL:12%	AVG
Source	TX:0.1%	0.0%	1.1%	0.7%	0.3%	0.4%	0.5%	0.6%
	AZ:0.6%	0.6%	0.0%	0.8%	0.0%	0.3%	0.3%	0.4%
	OR:1.5%	2.1%	2.1%	0.0%	0.6%	0.3%	0.6%	1.1%
	WA:2.2%	1.9%	2.6%	1.6%	0.0%	0.4%	0.5%	1.4%
	LA:8.3%	5.6%	3.1%	0.9%	1.2%	0.0%	0.7%	2.3%
	FL:12%	3.7%	5.2%	1.2%	0.9%	0.6%	0.0%	2.3%
	AVG	2.8%	2.8%	1.0%	0.6%	0.4%	0.5%	

Activities

- **Wetland Prioritization -** Working with the Wetlands and Dams PMP, this effort 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, which is essential for identifying potential wetland areas in the United States. To begin with, we claim that wetlands are generally sparse, covering only 6% of the entire land use types on average. Particularly, some areas with desert climates, such as Arizona, were found to have a wetland ratio below 1%. This is well-known to be a data sparsity issue. To solve this problem under such conditions, we suggest two strategies: (1) non-local knowledge transfer from the dense regions (with plentiful wetlands) to the sparse ones, and (2) adaptive propagation between local neighborhoods. For the first method, we should consider the aspect that each region may have different geographic characteristics. Thus, we identify the unique characteristics of regions for wetlands through domain disentanglement, which separates transferrable (domain-shareable) knowledge from the others (e.g., Arizona has dry soil while Louisiana has damp soil). This can be achieved through a domain discriminator, which disentangles domain-specific and domain-shareable features using the gradient reversal function. For the second approach, adaptive propagation, we adjust the weight of message-passing in a way that differentiates between node pairs that have positive and negative impacts. Given that wetlands are generally sparse, the graph will generally have low homophily, and in this case, it was confirmed that the smoothing effect can be effectively reduced through adaptive propagation. To show the effectiveness of our method, we conduct extensive experiments using Accuracy (%) and Recall. In detail, we assume two different domains, source and target, and select three regions of dry, moderate, and humid for each domain. From the results, we notice that the proposed scheme achieves state-of-the-art performance among the recently proposed techniques. Furthermore, the performance gain was significant when knowledge was transferred from dense (e.g., Louisiana or Florida) to sparse domains (e.g., Texas or Arizona), which confirms our assumption. Lastly, we aim to integrate multiple downstream tasks such as flood risk reduction and water storage enhancement into the wetland prioritization. We believe that this can improve the quality of the feature extraction process, leading to an overall performance improvement. 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.

- **Streamflows -** We are continuing our collaboration with Dr. Reepal Shah and Dr. Qi Deng from the ByWater Institute at Tulane University on streamflow prediction using spatiotemporal causal learning. We have developed a new model called Causal Streamflow Forecasting (CSF), a two-stage, physically-aware hierarchical network. In the first stage, a station-level model predicts runoff using forcing variables such as weather, soil, and elevation, with a Variational AutoEncoder (VAE) as its backbone. The second stage employs a basin-level model that uses Spatio-Temporal Graph Convolutional Networks (STGCN) and the river flow graph as the causal graph to learn streamflow rates at each station. During inference, only causally influential stations are utilized to make predictions at the target site. Our experiments on the Brazos basin demonstrate that CSF significantly outperforms other baseline models. Additionally, we are investigating the causal relationships between forcing variables by incorporating a causal discovery model at the first stage. Initial experiments indicate a positive impact on forecasting, though further analysis is needed to draw definitive conclusions.
- **Domain disentanglement** - In the field of hydrological modeling data scarcity is a prominent problem. Moreover, there exists data which is either generated through simulation or is noisy w.r.t the sensors available. To address this problem our effort introduces a weakly supervised causal disentanglement framework that aims to learn generalizable causal representations in the presence of noisy data. At the core of our model lies the ability to leverage denoising techniques such as high confidence sampling and confidence-based re-weighting to utilize weakly supervised data to learn two kinds of features (domain invariant – features that are highly reflective of the problem and domain dependent features – that are more focused about the domain e.g. location specific features). Such an approach allows to leverage the use of this model in scenarios where availability of data is limited, and data sources may be noisy. By leveraging causality, the model can learn vital relationships essential for the downstream task e.g. flood forecasting (where there are features that irrespective of the location affect the flood

levels), and secondly by leveraging denoising mechanisms our approach allows to train the model even with simulated data or data with missing values.

- **Causally Informed Pareto-Optimal Policy Recommendations -** Many decisions, such as wetland prioritization, involve multiple, often conflicting objectives (flood protection versus water quality). We are continuing our investigation regarding the use of causal learning to improve the efficiency of skyline discovery (i.e. pareto-optimal recommendations) for decision support queries. To improve efficiency, our approach works by splitting the data into groups based on control variates which are not part of the query. Our goal is to eliminate negative correlations between the preference variates while preserving the positive correlations. We are now developing an algorithm to discover the best possible set of control variates, if it exists.
- **Learn causal relationship from the calibrated physical-process-based hydrological model -** We have confirmed that the input and output data structure from the Soil & Water Assessment Tool (SWAT) is suitable for serving as the input source for the spatial-temporal causal learning framework – DataStorm. 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. Currently, we are building a table index that contains essential information (such as uncertainty, temporal and spatial variation, physical links between parameters, etc.) that stores all SWAT input and output parameters for the ease of data scientists' use.

- **New ASU-NEWN web site -**
    - https://www.asu-newn.org/

Outcomes

- Yoonhyuk Choi, Reepal Shah, John Sabo, Huan Liu and Selçuk Candan. Prioritizing Potential Wetland Areas via Region-to-Region Knowledge Transfer and Adaptive Propagation, Under Review at CIKM '24
- P. Mandal, Y. Choi, R. Shah, J. Sabo, H. Liu. K.S. Candan. Identifying Potential Wetlands via Causality-based Data Imputation and Knowledge Transfer. presented at the N-EWN Symposium, 2024.
- Shu Wan, Reepal Shah, Qi Deng, John Sabo, Huan Liu, and K. Selçuk Candan. Spatiotemporal Causal Learning for Streamflow Forecasting, Presented as poster at N-EWN Partner Symposium '24
- Fahim Tasneema Azad, K. Selcuk Candan, Ahmet Kapkic, Mao-Lin Li, Huan Liu, Pratanu Mandal, Paras Sheth, Bilgehan Arslan, Gerardo Chowell-Puente, John Sabo, Rebecca Muenich, Javier Redondo Anton, and Maria Luisa Sapino. (Vision Paper) A Vision for Spatio-Causal Situation Awareness, Forecasting, and Planning. ACM Trans. Spatial Algorithms Syst. Just Accepted (June 2024). https://doi.org/10.1145/3672556

- Ahmet Kapkiç, Pratanu Mandal, Shu Wan, Paras Sheth, Abhinav Gorantla, Yoonhyuk Choi, K. Selçuk Candan and Huan Liu. CausalBench: A Transparent Benchmark Framework for Causal Analysis and Machine Learning, Under Review at CIKM '24

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