Designing Nature to Enhance Resilience of Built Infrastructure (GR40695)

Quarterly Report - September 2024

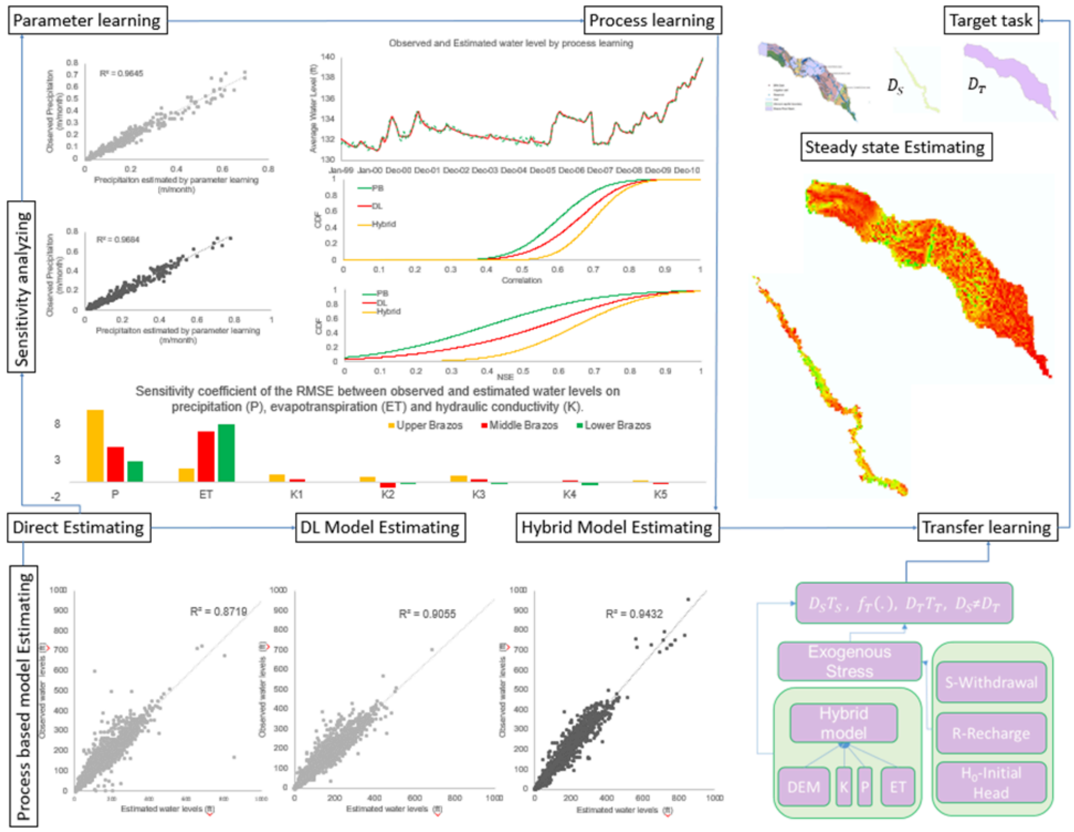
Wetland Siting/Dam Operations--Brazos PMP

The Wetland Siting/Dam Operations PMP has been actively building a modeling framework and ensemble to simulate strategic wetland siting in the Brazos River Basin. This ensemble consists of at four models: The 3-layer Variable Infiltration Capacity Model (VIC 3L) that simulates land surface processes and runoff generation, a 4-layer version of this model (VIC-4L) that also captures shallow groundwater dynamics through fusion with NASA's GRACE mission fused to the model, MODFLOW which simulates three dimensional groundwater fluxes, and CaMa Flood (CMF) which routes surface (and ground) water discharge through the river network. The goal is to fuse these models such that we get the best combination of model skill applied to estimating four key water fluxes relevant to water management: 1) Evapotranspiration (ET, "up"), 2) recharge from surface to groundwater (R, "down") and lateral flows in the river channel originating from surface and subsurface fluxes (Q, "horizontal"). None of these single models captures all four fluxes well. Existing models that attempt to capture all four are computationally expensive. We are working towards integrating these models with DataStorm, with the goal of creating an easy to use, but skilled representation of all four fluxes. The challenge is a space-time integration problem. Typically, routing models like CMF are implemented serially (after) land surface models like VIC and MODFLOW. This is problematic for wetland simulation in the river channel because inundated wetlands have different vertical fluxes (ET and R) than a non-flooded channel. Running VIC or MODFLOW with CMF with daily updates to water levels is computationally expensive. We will use DataStorm to connect and update input data (from routed floods) to land surface processes and create daily updated data in a more computationally efficient manner.

Progress on VIC 3L: We have already coupled 3-layer VIC (Liang et al. 1994) with CaMa-Flood to represent variability in surface components of hydrologic cycle (evapotranspiration, infiltration, and stream water flow in the Brazos basin.

Progress on VIC 4L: Stampoulis et al. 2019 have developed 4-layer VIC to add representation of variability in subsurface water (groundwater) in the land surface model VIC-3L in the basin in California state. They also assimilated the VIC-4L framework using satellite-based groundwater estimated, to improve anthropogenic contributions (such as irrigation and recharge). Tulane ByWater has subcontracted (in-kind) the construction and calibration/validation of VIC-4L for North America to Dimitrios Stampoulis at the Southwest Research Institute (SWRI) and we are 50% complete with the build of the CONUS portion of this which includes Texas. The variability in ground water simulated by the VIC-4L is in natural condition. For the Brazos basin, we will assimilate groundwater from NASA's GRACE and GRACE-FO mission, to incorporate anthropogenic variability in groundwater. VIC-4L will be calibrated for the Brazos River Basin using gauge-based groundwater and streamflow data in the basin.

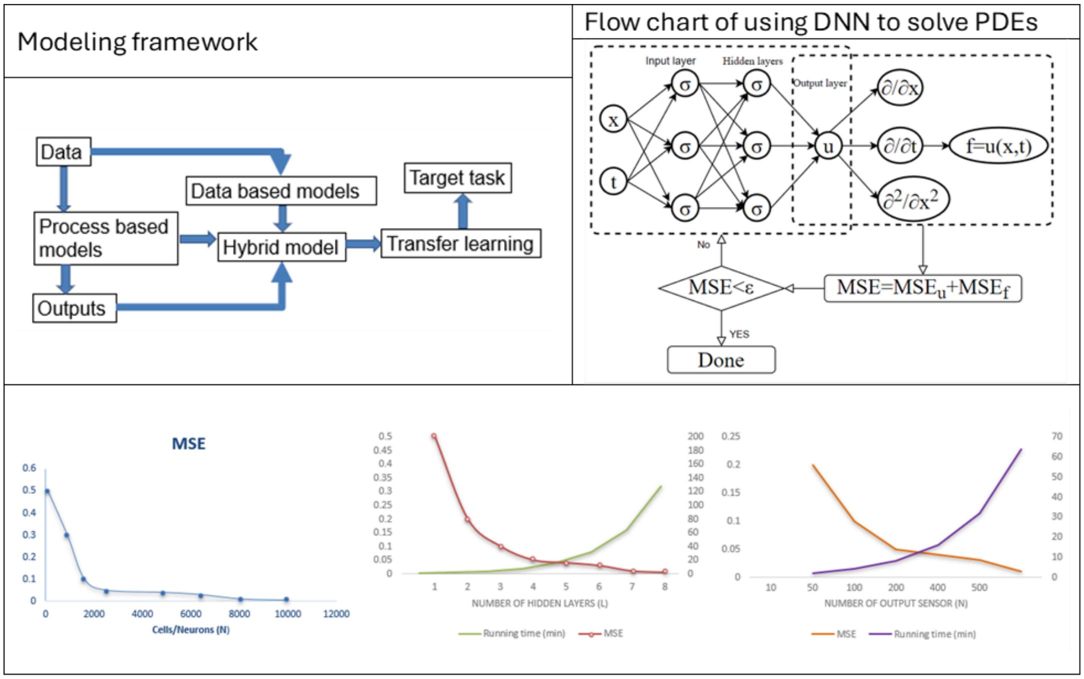
Progress on MODFLOW:



This PMP aims to study the impacts of wetland restoration and dam operations on water dynamics. It will explore impacts of proposed designed EWN strategies on overall water budget and different fluxes, including ET, groundwater recharge, river flux, etc. To achieve the above goals, we are constructing high resolution basin scale land surface models, groundwater flow models and river routing models to simulate the Brazos water system. Meanwhile, we are conducting basin-scale studies on the mechanisms of groundwater dynamics in response to inner and external forces using both physical process-based and machine learning models. (94-word summary)

The groundwater flow equation describes the movement of groundwater through a porous medium, it is a nonlinear partial differential equation (PDE). The nonlinear nature of the equation arises from the dependency of hydraulic conductivity (K) on the hydraulic head. Solving such nonlinear PDEs often requires numerical methods, such as finite difference, finite element, or finite volume methods, due to the complexity of the equations and the limited availability of analytical solutions in most cases. Recently, Machine Learning (ML) methods for solving partial differential equations (PDEs) have experienced a surge in popularity. Physics-Informed Neural Networks (PINNs) are innovative deep learning frameworks designed to address forward and inverse problems involving non-linear PDEs.

In this study, we explore the capability of PINNs in solving higher-dimensional flow equations, we analyze the approximation error for field cases in a statistical learning framework. Lastly, we compare our results obtained from PINNs with those from the Finite Difference Method (FDM) to highlight the advantages of PINNs in solving PDEs without the need for discretization. Our study suggests that the data driven model is effective when provided with systematic and consistent data over a continuous time period, the PIML models have higher accuracy, efficiency, and ability to deal with data. The Number of Cells/Neurons (N), Number of hidden layers (L) and Number of output sensor (n) have vital influence on the model performance. It remains challenging to balance the computational time, the complexity of models and the accuracy of model outputs.



Progress on DataStorrm: The ASU SCAI team has datasets from VIC-3L, CMF and MODFLOW in hand and is actively working on a DataStorm integration of these data workflows. A wealth of datasets generated by physical models will represent a wide variety of hydrologic processes. One of the grand challenges is to integrate datasets to get robust representation of natural/engineered process. Also, integration of spatio-temporal datasets of various granularity. Datastorm platform is explored to fully couple the hydrologic process in VIC3L and CaMa-Flood, and VIC4L and CaMa-Flood. Basically, this will improve our representation of hydrologic variables in wetlands, such as evapotranspiration (ET) and infiltration. Next, we will integrate datasets of ET, streamflow and groundwater coming from multiple modeling framework (VIC3L-CaMaFlood, VIC4L-CaMaFlood and MODFLOW). Modeled data are calibrated for VIC-3L but not yet for MODFLOW. Calibration of MODFLOW is happening in parallel, but the structure of the data from the calibrated model will not differ from those with which we are building the DataStorm infrastructure.

Activities

1. Building geospatial analysis tool for the Brazos River basin to identify potential wetland creation/restoration sites to increase floodplain storage and reduce flood risk in downstream.
2. Coupling hydrology (VIC) and hydrodynamic model (CaMa-Flood) for the basin.
3. Building and calibrating three representations of MODFLOW: physically-based (PB), deep learning emulation (DLE) and physics-informed machine learning implementations (PIML).
4. Editing science article on coupling hydro-economic modeling framework for wetland creation

5. Planning for initial stake-holder focus groups to co-produce analytical tools and scenarios for wetland planning in the Brazos River Basin

Outcomes

1. We have initial map showing sites for potential wetland sites developed with objective of increasing floodplain storage (at 1km x 1 km). Geospatial analysis was done on physical variables (such as soil characteristics, land use land cover, etc).
2. Set-up for hydrologic model ready to be studied for use of integration in Datastorm platform
3. Calibration results for MODFLOW indicate better performance and efficiency of the PIML MODFLOW implementation. Data ready for DataStorm integration.
4. In review science article:

   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.

5. Human Subjects protocol approved through ASU and the USACE for our proposed stakeholder engagement activities.

Team

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

Ag/WQ PMP

Activities

1. Verifying and correcting dam and reservoir locations from the USACE National Inventory of Dam database.
   a. Have hired two undergraduate students to assist with this process
2. De-bugging USGS API for delineating contributing areas
3. Creating Soil and Water Assessment Tool (SWAT) model parameter diagram to help connect SWAT with data science research for the next steps.
4. Building USGS nutrient and sediment parameter integrator to standardize the nutrient measurements into uniform chemical forms and units that align with the SWAT model simulations.

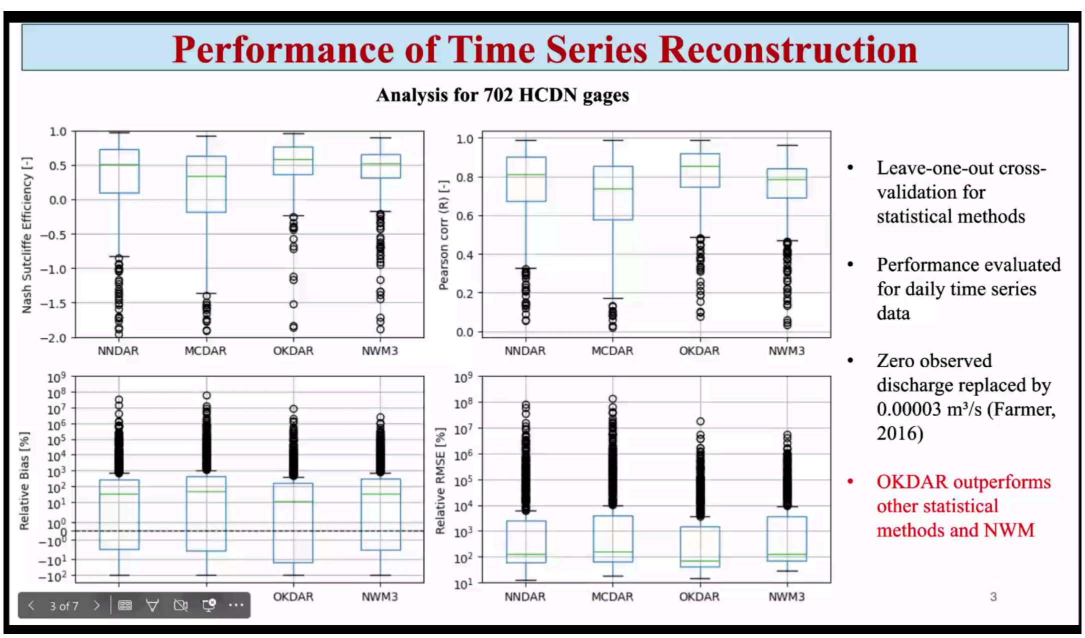
Outcomes

1. We have now recruited two undergraduate students to participate in the massive work of verifying and correcting dam and reservoir locations from the USACE National Inventory of Dam database.
2. Dr. Muenich has co-initiated an agricultural working group in the larger N-EWN.
3. We have distributed the first edition of the SWAT model parameter diagram to collaborators, which includes detailed indexing of all sub-watershed level water quantity parameters in the SWAT model.
5. We have completed the code and validation of the USGS nutrient parameter integrator, while still working on the sediment integrator.
6. We have submitted two abstracts which expecting to be presented in the AGU 24 fall meeting and expect both to become full paper drafts by the end of the year: “Comparison of agricultural activities and reservoir management on water quality at the watershed-scale” and “Do upland agricultural areas and downstream reservoirs counteract efforts to improve water resources?”.

Team

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

Flood-MAR PMP



Comparison of four different statistical methods for interpolating streamflow in ungauged streams.

Activities

We are working to develop three CONUS-wide geospatial datasets that will be used to calculate the FloodMAR suitability index and a web-based GIS application that incorporates these data into a FloodMAR suitability screening tool. One of these datasets (land surface suitability for FloodMAR) is complete.

- Surface water availability - We are developing a statistical method for predicting the volume and duration of high flow events in ungauged watersheds.
- Aquifer Storage Potential - A machine learning algorithm model is being 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 will be used to create monthly gridded groundwater table depths across CONUS including areas with no or scarce groundwater level observations.
- Screening Tool - We are 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. These factors will inform the design of the screening tool.

Outcomes

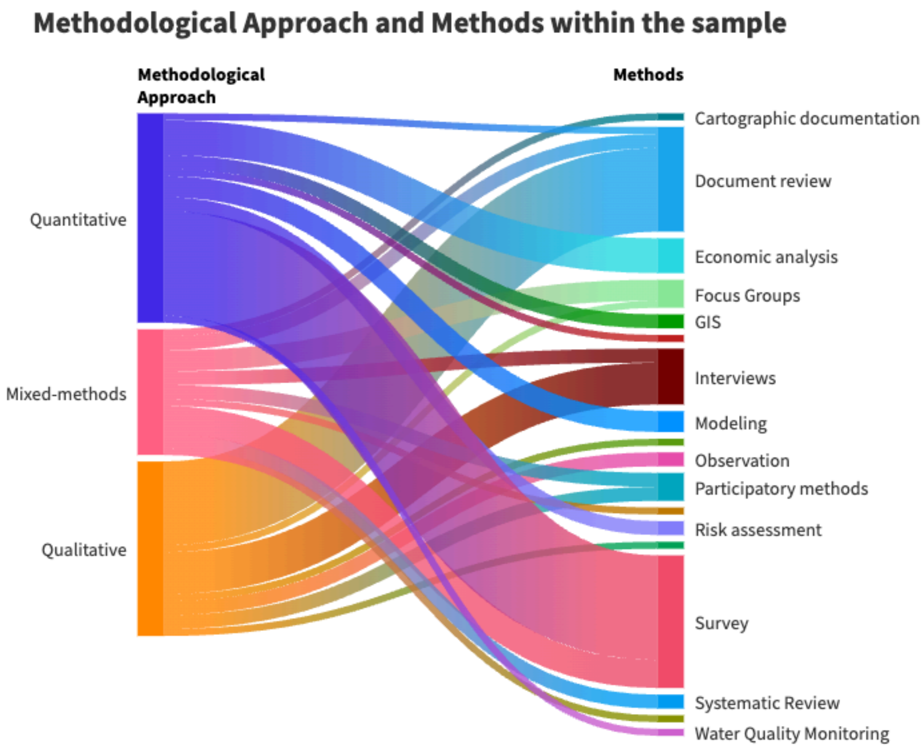
- Multiple sources of flowrate data were evaluated, including the National Water Model, Nearest-Neighbor Drainage Area Ratio (NNDAR), Map-Correlation Drainage Area Ratio (MCDAR), and Ordinary Kriging of the logarithms of discharge per unit area (OKDAR). We've determined that OKDAR out performs the other statistical methods and the National Water Model, when compared to USGS stream gage data.
- A methodology has been developed for determining total flow in a HUC8 based on the interpolated flow of all component HUC12s. This requires determining which of the HUC12s are most downstream, as well as which ones are ocean/lake draining or endorheic.
- A new machine learning model for interpolating groundwater levels has been developed that uses USGS monitoring wells directly for training, instead of aggregating them into a raster grid first. This is much more compute-intensive, and requires the use of ASU's supercomputer, but is performing much more accurate interpolations.
- An approach was developed for incorporating GRACE gravitational anomaly data into the groundwater interpolation model.
- Submitted two abstracts for presentations at AGU:
    - "Managed Aquifer Recharge: Physical and Social Considerations for Siting"
    - "Varying impacts of climate variability and human activities on groundwater response in the western United States"

Team

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

- Qinyuan Dai

Participatory Research PMP



Our review of research on Managed Aquifer Recharge assessed and classified the methodological approaches used to study MAR social dynamics: quantitative approach (41%), followed by qualitative (23%) and mixed methods (12%). Most of the studies used surveys (21%), document reviews (17%), interviews (10%), and focus groups (5%). Our findings can inform the future growth of the NEWN and its integration of interdisciplinary teams and approaches.

Activities

The goals of the social science PMP are to study human dimensions of engineering with nature, integrate and inform other PMPs of key human and social dynamics, and capacitate the NEWN to work with human stakeholders to advance engineering with nature. To this end, the social science PMP has taken the lead in developing a high-level, free-of-cost curriculum to train NEWN partners in participatory methods. Leveraging a $150,000 investment from the Arizona Water Innovation Initiative, which is funded by the Arizona Governor's Office, the social science PMP has partnered with ASU's Learning Enterprise to create an interactive continuing education course.

The social science PMP continues to be integrated into several other PMPs, including FloodMAR, Brazos, and Education. We have made significant progress in our partnership with the FloodMAR PMP in the last quarter. Our process included an 18-month effort, now complete, to secure IRB and other federal approvals required to launch human subjects research in partnership with the USACE. Once permissions were finalized, we were able to launch FloodMAR surveys to internal and external stakeholders who have expertise about MAR. To support this broader effort, we are also completing two review papers about the roles of social dynamics and participatory methods in MAR.

Highlights of the social science team several activities across PMPs:

- Meetings with Learning Enterprise to build the education platform
- Deployed the Perceptions of Flood-MAR survey to internal USACE Stakeholders (emails sent to 3 internal listservs, with more pending)
- Deployed the Perceptions of Flood-MAR survey to external stakeholders (283 emails sent to relevant stakeholders)
- Two review papers in progress
    - The first version of the MAR scoping review is complete
        - Invitation from journal editors secured & confirmed
        - The manuscript has been edited by lead coauthors.
        - Next steps to circulate to full team
    - Outline of the review Participatory methods & MAR
        - Preliminary outline – complete
        - Detailed outline – in progress
        - Circulate detailed outline to social science team for feedback – anticipated Sept 30

Outcomes

In this quarter, the social science PMP has made significant advances in studying human dimensions of engineering with nature, integrating and informing other PMPs of key human and social dynamics, and capacitating the NEWN to work with human stakeholders to advance engineering with nature. A major accomplishment is completing three modules for the high-level, free-of-cost curriculum to train NEWN partners in participatory methods in partnership with the Arizona Water Innovation Initiative and ASU's Learning Enterprise. These modules teach anyone who wishes to learn how to do: (1) citizen science, (2) participatory modeling, and (3) art-based methods.

Other major outcomes are related to the social science PMP's integration with the FloodMAR PMP. One major accomplishment was finalizing the IRB and other federal approvals required to launch human subjects research in partnership with the USACE. Another key outcome was the deployment and completion of > 100 FloodMAR survey interviews with internal and external expert stakeholders. A final key outcome is significant progress toward completion of two systematic review papers about the roles of social dynamics and participatory methods in MAR.

- ASU LE Continuing Education Participatory Methods course, in partnership with Arizona Water Innovation Initiative
    - 3 of 6 modules completed:
    - Citizen Science module
    - Art-based Methods module
- Flood-MAR internal USACE responses: n = 43
- Flood-MAR external USACE responses: n = 71
- Participatory Modeling moduleAcademic manuscripts
    - Citation (in prep): Kelly, E. M. Beresford, L. Castro Diaz, M. du Bray, A. Wutich, C. Jacob., O. Medina-Ramírez, J. Jancovich Rankovich, inter alia. Participatory Methods

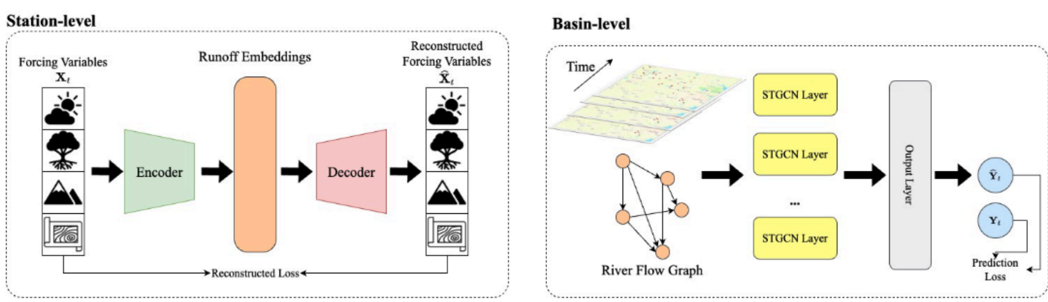
for Assessing Managed Aquifer Recharge (MAR) Projects: A Review. In prep for *Weather, Climate, and Society OR Water and Climate Change*

- Citation (in prep): Castro Diaz, L., M. du Bray, A. Wutich, M. Beresford, C. Jacob., O. Medina-Ramírez, J. Jancovich Rankovich, inter alia. Social Dimensions of Managed Aquifer Recharge (MAR) Projects: A Scoping Review. Invited to *WIRES Water*.

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



(a) **Station-level model:** The station-level model is a VAE designed to learn an embedding of the forcing variables. Conceptually, it mimics the learning process of traditional process-based models like VIC. The learned runoff embedding does not represent actual runoff; instead, it serves as a latent representation that is passed to the basin-level model in the next stage for further processing.

(b) **Basin-level model:** The runoff embeddings generated by the station-level model are passed as node embeddings in the basin-level Spatio-Temporal Graph Convolutional Network (STGCN). The river flow graph is constructed using DEM and HUC data, and serves as the spatial causal adjacency matrix to guide causal message passing within the model.

Activities

Wetland prioritization. We focused on integrating data-driven and causal-based learning with LLM-based soil documentation analysis to support wetland prioritization. We have initially carried out SVM and decision-tree based analysis of solid characteristics for wetlands and non-wetlands, for different states, based on the wetland occurrence frequency. Secondly, we have carried out causal machine learning and LLM based analysis — for this, our primary soil features included drainage class, ponding frequency, hydrologic group, slope gradient, water table depth, and available water storage at multiple depths. We first applied LLM-based soil documentation analysis to construct a causal graph. Although the graph was reasonable, it did not fully capture the true relationships between certain features. For the feature-based analysis, in Arizona, our analysis with the GES algorithm showed reversed causality between some features, such as water storage influencing slope gradient, but it successfully identified domain-specific causalities like the independence of ponding frequency. Combining LLM and feature-based analysis in Arizona allowed us to generate improved causal graphs, but we still encountered errors, such as the relationship between slope gradient and available water storage, which required correction. Moving on to California, we observed stronger causal relationships between features. However, most features converged toward the drainage class. Although the LLM and feature-based analysis in California required some adjustments, the initial results appear to be promising, which can be applied prior to the knowledge transfer algorithm to solve the data sparsity problem.

Streamflow prediction. Streamflow plays an essential role in the sustainable planning and management of national water resources. Traditional hydrologic modeling approaches simulate streamflow by establishing connections across multiple physical processes, such as rainfall and runoff. These data, inherently connected both spatially and temporally, possess intrinsic causal relations that can be leveraged for robust and accurate forecasting. Spatio-temporal graph neural networks (STGNNs) promise advances in streamflow management. However, learning causal relationships directly from observational data is theoretically and computationally challenging.We are developing a river flow graph as prior knowledge to facilitate the learning of the causal structure and then use the learned causal graph to predict streamflow at targeted sites. The proposed model, Causal Streamflow Forecasting (CSF) is tested in the Brazos River basin in Texas. Initial results demonstrate that our method outperforms regular spatio-temporal graph neural networks and achieves higher computational efficiency compared to traditional simulation methods. By effectively integrating river flow graphs with STGNNs, this research offers a novel approach to streamflow prediction, showcasing the potential of combining advanced neural network techniques with domain-specific knowledge for enhanced performance in hydrologic modeling

Data imputation. Missing hydrological data significantly impacts the accuracy and reliability of time series analysis and subsequent applications in assessing water quality and quantity. To deal with this challenge, we have developed a cross-domain time series imputation strategy that leverages data from a data-rich region to help with the imputation task in a target region with high missing rates. High missing rates make it difficult to accurately capture the underlying patterns in the target domain, while data imbalance between domains complicates effective knowledge transfer, limiting the adaptability in achieving optimal performance. Evaluation has been carried through a dataset that records daily river flow and sediment concentration from 20 stations in the United States, collected from United States Geological Survey and Water Quality Portal, consisting of two domains: Discharge (D) domain (no missing data) and Pooled (P) domain, spanning from March 1, 2017, to September 30, 2022.

Model Integration. Hydrological systems are characterized by complexity and dynamicity. Data-driven simulations represent a promising approach in understanding and predicting complex dynamic processes in the presence of shifting demands of hydrological systems. We have therefore started investigation model integration techniques to support integration of independently developed models, such as VIC, CaMa-Flood, and SWAT, under a unified execution and simulation framework to support effective whole-system modeling and simulation. To achieve this goal, we are currently investigating extension of the DataStorm continuous, coupled modeling and simulation engine, which provides an actor-driven model- and data-flow capability for plug-and-play model integration, to accommodate hydrology (VIC) and hydrodynamic model (CaMa-Flood) as actors to be integrated into a single workflow/model for a given basin.

Causal benchmarking. While witnessing the exceptional success of machine learning (ML) technologies in hydrological applications, we notice a critical shortcoming of ML in this context: correlation is a poor substitute for causation. The conventional way to discover causal relationships is to use randomized controlled experiments (RCT); in hydrological context, however, these are impractical or sometimes unethical. Causal learning from observational data offers a promising alternative. While being relatively recent, causal learning aims to go far beyond conventional machine learning, yet several major challenges remain. CausalBench is a transparent, fair, and easy-to-use evaluation platform, aiming to (a) enable the advancement of research in causal learning by facilitating scientific collaboration in novel algorithms, datasets, and metrics and (b) promote scientific objectivity, reproducibility, fairness, and awareness of bias in causal learning research. It provides services for benchmarking data, algorithms, models, and metrics. We are currently integrating hydrological data into CauslaBench for assessing causal learning models within the context of hydrological questions.

Outcomes

- In review article: Domain-Adaptive Conditional Diffusion Models for Cross-Domain Time Series Imputation. Submitted to AAAI24
- In preparation: Shu Wan, Reepal Shah, Qi Deng, John Sabo, Huan Liu, K. Selcuk Candan. Spatio-temporal Causal Learning for Causally-Guided Streamflow Forecasting
- Accepted for publication: Ahmet Kapkiç, Pratanu Mandal, Shu Wan, Paras Sheth, Abhinav Gorantla, Yoonhyuk Choi, Huan Liu, K. Selçuk Candan. Introducing CausalBench: A Flexible Benchmark Framework for Causal Analysis and Machine Learning. CIKM 2024.

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