

CLEMSON



IOWA





Table of Contents

<i>Web-based Integrated Drought Monitoring and Analysis with Copula Models</i>	4
Abstract	4
1. Introduction	4
1. Data and Research Area	6
2. Methodology	8
2.1. Drought Indices: SPI and SSFI	8
2.2. Bivariate dependence of SPI and SSI	9
3. Results and Discussion	10
3.1 Drought index development and temporal variations	10
3.2 Bivariate Copula selection.....	12
3.4 Conditional Probability of Occurrence	15
4 Conclusions	15
Acknowledgement	16
References	16
<i>Stream Temperature Variability and its Relationship to Large-Scale Climate Indices across the United States</i>	20
Abstract	20
1. Introduction	20
2. Methodology	21
2.1 Stream Water Temperature Data Stream Water Temperature (SWT) Data	21
2.2 Automated Data Retrieval Code.....	22
2.3 Data Processing and Quality Control.....	22
2.4 Auxiliary Data.....	24
2.5 Variability Analysis	24
3. Results and Discussion	26
3.1 SWT Variability Dashboard.....	26
3.2 SWT variability by HUC-02.....	28
3.3 SWT Trends by Aridity Index.....	29
3.4 SWT Trends by Stream Order.....	31
3.5 Link with climate oscillations	32
4 Conclusions	33
Acknowledgement	34
References	35
Appendix	37
<i>Clear Data, Clearer Waters: An Interactive Dashboard for Visualizing Stream Turbidity Trends</i>	39
Abstract	39



1. Introduction	40
2. Methodology	41
2.1 Study Area:	41
2.2 Data Retrieval:	41
2.3 Data cleaning and analysis:	42
3. Results and Discussion	44
3.1 Exploratory Data	44
3.2 XGBoost model	46
3.3 Challenges	48
4. Conclusions	48
References	48
Appendix	50



Web-based Integrated Drought Monitoring and Analysis with Copula Models

Suman Dhamala¹; Kunal Bhardwaj¹; Ethiopia Zeleke²; Moses Kiwanuka²

¹Glenn Department of Civil Engineering, Clemson University, Clemson, SC 29634

²Department of Earth and Environment, Florida International University, Miami, FL 33199

Advisors: Vidya Samadi, Carlos Erazo, Demir Ibrahim

Abstract

Our study, 'Web-based Integrated drought Monitoring and Analysis with Copula Models', addresses a pressing issue in the field of hydrology and climate change. Drought, a threatening hydro-climatic extreme event, is one of the costliest natural disasters globally, causing widespread famine, loss of life, and various socio-economic impacts. In the United States, drought effects range from individual hardships to regional economic slowdowns and environmental degradation, including increased wildfire hazards. Droughts are increasingly influenced by climate change, affecting their frequency and severity. Various studies have mainly focused on understanding drought vulnerability using indices like the Drought Vulnerability Index (DVI) and the Multivariate Drought Index (MDI). However, progress is still lacking when it comes to the simultaneous occurrence of different drought types, which has limited our understanding of drought propagation. This study introduces an online portal developed within the Hydrolang infrastructure, providing easy access to large-scale data services from the United States Geological Services (USGS). The portal enables bivariate drought analysis, offering tools for assessing various drought types. As a test case, we evaluate droughts over the Alabama Coosa-Tallapoosa (ACT) River Basin, covering the northeastern and east-central sections of Alabama, northwestern Georgia, and part of Tennessee. Given the projected increase in drought frequency and severity due to climate change, this study examined the simultaneous occurrence of different drought severities and intensities, focusing on precipitation and streamflow deficits modeled using the elliptical and Archimedean family of bivariate copulas. We expect that our findings will provide valuable insights for water resource managers and policymakers, particularly in drought-prone regions, to enhance water allocation, conservation, and infrastructure planning.

Keywords: Drought propagation, Precipitation, Streamflow, Hydrolang, Copulas

1. Introduction

Drought is an extreme hydro-climatic event marked by severe deficiencies in water availability (Konapala & Mishra, 2017). Droughts have caused huge losses globally, with severe impacts



resulting in increased plant mortality, reduced water availability, and loss of livelihood from the affected socio-economic conditions of the region. In the U.S., the drought impacts are from personal to regional levels. Small-scale farmers are left with no option but to kill their livestock due to a lack of enough water for drinking or feeds. At the same time, environmental degradation, economic slowdown, and multiple hazards like wildfires are felt at the regional level (Engström et al., 2020). Droughts commonly occur naturally. However, their frequency and severity are subject to change due to rampant climate change (Dai, 2011; Keellings & Engström, 2019; Madadgar & Moradkhani, 2013; Mukherjee et al., 2018). Due to the multifaceted nature of drought, various drought types are defined based on the monitoring method for any impacted sector. Drought events are mainly classified as meteorological (precipitation shortages), agricultural (decreased soil moisture), hydrological (decreased streamflow), and socioeconomic (failure to meet societal needs) (Engström et al., 2020). Based on the available conceptual models (Van Loon et al. 2016), the different types of droughts commonly occur in a particular sequence: a deficit in the precipitation leads to a meteorological drought, and its combination with high evapotranspiration leads to agricultural drought (Hagenlocher et al., 2020). Hydrological droughts occur because of temperature anomalies and precipitation shortfalls. Socioeconomic drought occurs because of less supply of particular economic goods due to other drought types, such as meteorological, agricultural, and hydrological droughts (Van Loon et al., 2016; Wang et al., 2016).

Numerous scholars and researchers have conducted studies on drought; Engström et al., 2020 studied Drought Vulnerability in the U.S. using three main factors: sensitivity, exposure, and adaptive capacity. In this study, the adaptive capacity was assessed through a set of socio economic and environmental indicators such as existing infrastructure, water management practices, and community preparedness that reflect a state's ability to cope with drought. A Drought Vulnerability Index (DVI) was used by combining biophysical and socio-economic data. However, the study acknowledges that the current indicators may not capture the complexity of adaptive capacity fully, suggesting that future research should focus on developing more comprehensive quantitative measures such as regional forecasting and more detailed studies for areas with high uncertainty DVI. Rajsekhar et al., 2015 studied drought causality, vulnerability, and hazard assessment for futuristic socioeconomic scenarios by developing an integrated framework that quantifies drought causality, hazard, and vulnerability by considering various drought forms (meteorological, agricultural, and hydrological) and their interactions with future socioeconomic scenarios. In this study, a Multivariate Drought Index (MDI) was used to quantify different physical forms of drought simultaneously, Causal Analysis to identify dominant drought triggers and develop causal maps that visualize information flow within the natural system, and Risk Assessment in order to incorporate vulnerability indicators alongside drought hazard assessments. However, the study identified the need for more socioeconomic indicators and their interactions and the need for long-term monitoring to track changes in drought patterns. Fangyue et al., 2021 studied how observed precipitation reveals longer and more variable drought. Mann-Kendall trend tests and Regional Kendall tests were utilized to assess trends in the mean state and variability of these precipitation metrics. There was an increased interannual variability in both precipitation and dry intervals.

(Van Loon et al., 2012) studied drought propagation in large-scale hydrological models, looking at



how these models reproduce the transition from meteorological drought (caused by insufficient precipitation) to hydrological drought (characterized by reduced water availability in rivers, lakes, and groundwater). Results from the models effectively simulated the relationship between meteorological and hydrological droughts, with minimal lag time between the onset of precipitation deficits and hydrological drought. However, there were accuracy issues regarding the simulation of drought characteristics in slowly responding systems, which led to an overestimation of short drought events. It also noted that future studies should consider the anthropogenic influence such as groundwater extraction and representation of storage processes in the modeling process.

Over the past decades, this disaster has affected the well-being of the people, environment, and economy. It is believed that its frequency and severity will significantly increase across the U.S. and the world at large due to climate change (IPCC 2012; Trenberth et al. 2013, UNCCD 2016). Although many scholars in the U.S. have conducted several vulnerability studies on more damaging hazards such as tornadoes (Boruff et al., 2003; Senkbeil et al., 2014; Strader et al., 2017), hurricanes (Cho & Chang, 2017; Pita et al., 2015; Song et al., 2020), and floods (Alipour et al., 2020; Cho & Chang, 2017; Cutter et al., 2013; Nasiri et al., 2016), less attention has been given to a greater understanding of the simultaneous occurrence of multiple drought types in the U.S. Therefore, this study seeks to understand the simultaneous occurrence of different drought types, explicitly looking at precipitation deficit translates into stream flow deficits and the conditional probability that a hydrological drought occurs given a meteorological drought for drought propagation evaluation. This study will enhance our understanding of drought dynamics based on watershed uniqueness, which is paramount for predicting and managing water scarcity and other related adverse impacts. It will act as a piece of valuable information for water resource managers and policymakers in making informed decisions for water allocation, conservation strategies, and infrastructure planning, particularly in regions prone to drought.

As such, the objective of this paper is to examine how precipitation droughts are transformed into streamflow/hydrological droughts at stream-gauging locations. We intend to answer two questions:

1. What is the probability of observing both hydrological and meteorological drought at once $P(SSFI \leq -1, SPI \leq -1)$?
2. What is the probability of observing hydrological drought given that meteorological drought has occurred $P(SSFI \leq -1 | SPI = -1)$?

Before performing any statistical analysis, we need to make sure that we meet the stationarity requirements. To this end, we will standardize the dataset to remove influences of autocorrelation and seasonality in our datasets.

1. Data and Research Area

1.1. Observed precipitation and streamflow

The precipitation and discharge data used in this study was sourced from the United States Geological Survey (USGS) National Water Information System (<http://waterdata.usgs.gov/nwis>)



for the period 1990 to 2023 at a daily time step. We used a JavaScript-based approach to visualize and interact with hydrological data using the HydroLang framework data retrieval module. The core functionality was implemented through the following essential functions:

1. Initialization - The *Hydrolang* object was created to manage hydrological data and map interactions. Global variables were set up to handle overlay variability and GeoJSON data.
2. Polygon layer loading - The *loadPolygon* function fetched and loaded the GeoJSON file containing polygon boundary data. This extracted geographic coordinates to initialize a map view and overlay the boundary
3. Data retrieval: the *retriveData* function accessed the USGS WaterOneFlow API to obtain site data within the specified bounds. The data was transformed into a suitable format for further processing.
4. Data filtering and mapping: the *renderLocations* function processed the retrieved data to filter stations located within the defined polygon. It classified them based on the presence of discharge and precipitation measurements. Stations were then displayed on the map with interactive pop-ups showing detailed information.
5. Data Visualization and Download: the *retrieveValues* function enabled retrieval and visualization of discharge and precipitation data for selected sites. This also included functionality to display in charts and tables and offered download options for the data in CSV format.

The daily data was then rescaled to a monthly period for further analysis. This was done in a python-based environment by appending the columns within the retrieved CSV and creating a data frame of stations and associated monthly discharge and precipitation.

1.2. Study area

The study area comprises the Alabama Coosa-Tallapoosa (ACT) River Basin, which covers the northeastern and east-central sections of Alabama, northwestern Georgia, and part of Tennessee (Fig. 1). The Basin has an approximate area of 59,100 km² with 14 US Hydrologic Subbasins (HUC08s). The Coosa and Tallapoosa rivers form the Alabama River near Montgomery. The terrain of the subregion is relatively flat, with few mountainous regions in the north. As per the National Elevation Dataset, the area elevation ranges from sea level - 1278 m (Gesch et al., 2002). The watershed receives an average yearly precipitation of 1379 mm, mainly from rainfall. The main land cover is the Forest, which results in higher evapotranspiration (762–1067) mm. Streamflow is highly regulated by reservoirs (15 large reservoirs) and (or) a system of dams (5 Federal dams) (Ruddy & Hitt, 1990).

In order to analyze drought propagation, the uppermost upstream part of the watershed was selected as it would reduce the need to evaluate additional variables and inflows that may originate from them. The study area shown in Figure 1 encompasses multiple discharge measuring stations and is suitable for the analysis. In addition, this sub-watershed experiences little to no spatial variability in observed precipitation, thus reducing the computation complexity of analysis.

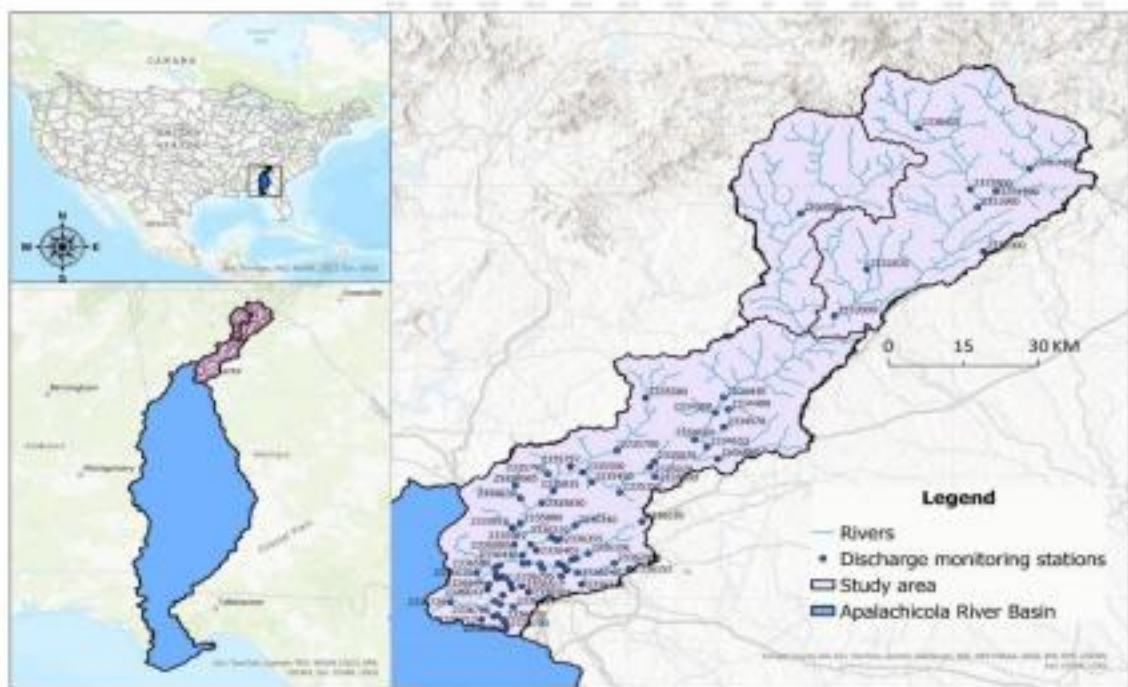


Figure 1: The study area

2. Methodology

2.1. Drought Indices: SPI and SSFI

In our study, we focus on analyzing precipitation to streamflow droughts within the ACT watershed. The Standardized Precipitation Index introduced by McKee et al, 1993, was designed to determine the precipitation deficit over a given period in relation to a historical mean. It shows the standardized deviation of the actual rainfall from the rainfall probability distribution function (Naresh Kumar et al., 2009). Over the recent years, it has gained much visibility as an indicator of potential droughts. The SPI is determined by fitting a probability distribution to the rainfall data and transforming it into a standardized normal distribution (Balew et al., 2021). Its primary purpose is to transform the mean and the standard deviation of the precipitation values to 0 and 1.0, respectively. Any skewness of the existing data can be readjusted to 0. Standardized Streamflow Index follows a similar methodology, where monthly streamflow is fitted to an appropriate distribution, and then the observed occurrence probabilities are transformed into z-scores.

Non-parametric drought indices offer a flexible alternative to traditional parametric indices by not assuming a specific statistical distribution for the data. These indices rely on empirical probabilities (from plotting position formulas) to evaluate the conditions. This approach is particularly advantageous in regions where climatic data do not fit common parametric distributions, allowing for more accurate and region-specific drought assessments. Additionally, non-parametric indices can adapt to varying climatic conditions, making them robust tools for global drought monitoring and management in the context of climate change.



In our study we developed a nonparametric Standardized Precipitation Index (SPI) and nonparametric Standardized Streamflow Index (SSFI) from the corresponding observed precipitation and streamflow for a watershed. Additionally, both SPI and SSI can be calculated at different accumulation scales, where the observational values are converted to an n-month running sum/average. Accumulation scales are used to make the drought index more relatable to ground-based phenomena, with more minor accumulation scales (< 3 months) important for agricultural drought and larger scales (> 12 months) for monitoring groundwater droughts. The following steps are used when calculating the non-parametric drought indices from the observations (X).

1. Apply the rolling window for accumulating the observations
2. Group data by climatology, i.e. months
3. Rank(r) the precipitation (or runoff) values $X_1, X_2, X_3, \dots, X_n$ in ascending order.
4. Calculate the empirical cumulative probability $P(X)$ based on the rank.

$$P(X) = r/n + 1$$

5. Transform $P(X)$ to the standard normal distribution to get the Standardized Index (SI):

$$SI = \phi^{-1}(P(X))$$

Where ϕ^{-1} is the inverse cumulative distribution (CDF) of standard normal distribution.

2.2. Bivariate dependence of SPI and SSI

Recognizing the interconnected hydrological processes governing streamflow/runoff generated from precipitation and modeling their dependence is necessary. While other forms of multidimensional analysis, like multivariate normal distribution, could be used, they assume that marginal follows specific distributions. To resolve this, the use of a copula is necessary, which can model the dependence between the variables even if they have varying marginal behavior (Sklar, 1998).

Copulas are generally classified into four classes: Archimedean, extreme value, elliptical, and other miscellaneous classes. In this study, bivariate classes of elliptical copula (Gaussian and Student's t) and eight from the Archimedean class (Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, and BB8) were considered. BB1, BB6, BB7, and BB8 are from the two-parameter families. The two-parameter families of copula can be instrumental in capturing more than one type of dependence, e.g., one parameter for the upper tail and lower tail dependence each, or one parameter for concordance while the other captures the lower tail dependence [Joe, 1997].

Since copula takes its input $[0, 1]$, therefore the calculated SPI and SSI are to be converted into a uniform distribution. This is done by calculating Empirical Cumulative Distribution (ECDF) by taking the whole time series.

Pseudo Maximum Likelihood Estimation is used to estimate the parameters of each family of copula, and the best copula is selected based on the least Akaike Information Criteria.

This way, we find linkages between different scales of drought index: 1 and 3 months. Hence a

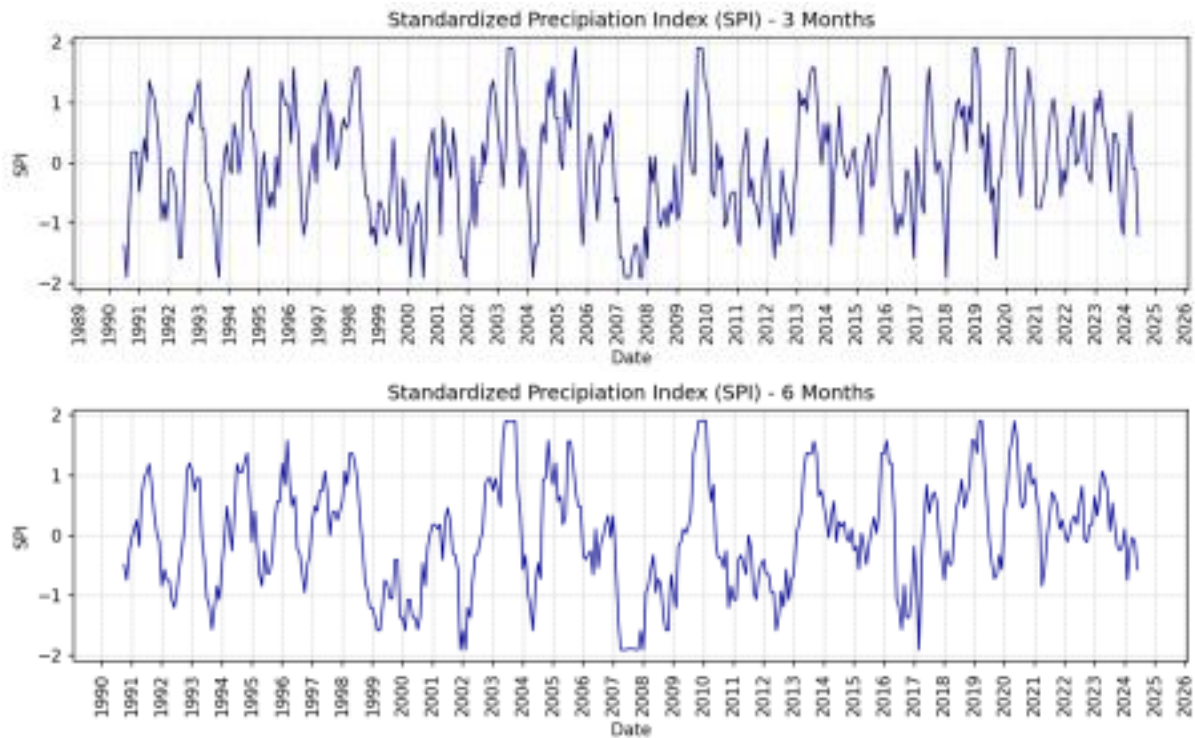


best fit copula for each station based on its own SSI and a common average SPI 1 and 3 is found. Then we check what is the probability of observing hydrological drought given that meteorological drought has occurred $P(SSFI \leq -1 | SPI = -1)$ and what is the probability of observing both hydrological and meteorological drought at once $P(SSFI \leq -1, SPI \leq -1)$?

3. Results and Discussion

3.1 Drought index development and temporal variations

The estimated standardized dataset includes normalized values for each site, with columns representing stations, making it easy to visualize and summarize. Figure 2 indicates the time series of the 3, 6, and 12-month SPI for an averaged observation over the study area. Figure 3 shows the 3, 6, and 12-month SRI estimated over all stations in the study area. Seasonal droughts are captured using the 3-month SPI and SRI, while intra-annual and annual events are highlighted in the 6- and 12-month indicators, respectively.



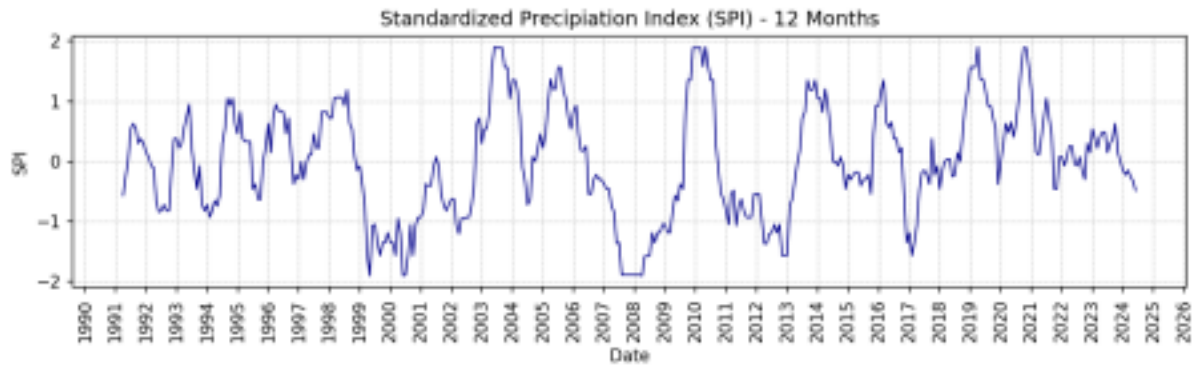


Figure 2: Standardized Precipitation Index (SPI) time series for various time scales averaged over the study area

Periods of significant precipitation deficit include 1999/2000, 2008, 2013, and 2017, which are captured using the 12-month SPI. These periods represent $SPI < -1$, which indicates severe meteorological droughts, according to McKee et al., 1993. The frequency distribution of SPI showed that 15% of the data lies between SPI of -1.0 and -0.5, 9% between -1.5 and -1.0, and 6% between -2.0 and -1.5. The majority lay within the mild (18%) to moderate (15%) drought range.

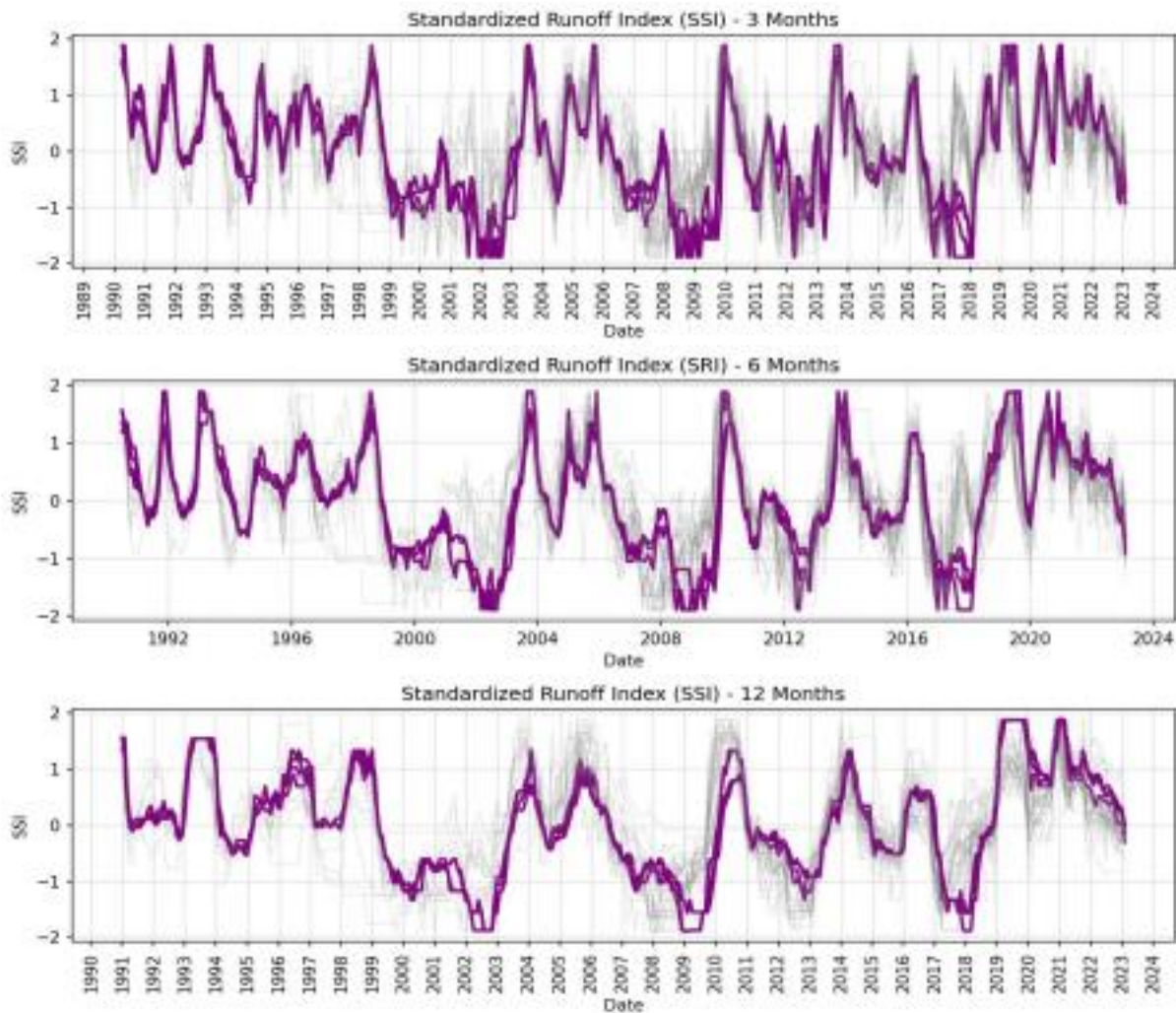




Figure 3: Standardized Streamflow Index (SSI) time series for various time scales in the study area (gray lines show all stations while the solid line shows selected stations along the main river)

SRI values, which represent standardized streamflow or discharge, show that periods 2002/03, 2008/09, 2012/13, and 2017/18 had significant deficits ($SRI < -1$), as represented in Figure 3. The frequency distribution of SSI also showed similar results for the SRI-12 time series. 15.5% of the data lies in the range -1.0 and -0.5, 9.3% between -1.5 and -1.0, and 6.2% between -2.0 and -1.5, revealing a good agreement between the standardized rainfall and streamflow distributions. As these discharge monitoring stations are positioned along various points in the primary and adjoining streams, they have a range of contributing areas that may be influenced by multiple agrometeorological as well as anthropogenic factors. This may reduce the linear correlation between precipitation and streamflow and requires an evaluation of each station for a nuanced assessment of correlation and synchronicity.

3.2 Bivariate Copula selection

Various Elliptical and Archmedian families of copula were tested for fitting SPI and SSI time series. Two sets of results are presented here by varying the accumulation scale of the modeled Standardized Indices (SPI/SSI): a one-month and a three-month scale. The spatial distribution of results obtained includes the best-fit copula model, the joint probability of occurrence, and the conditional probability of occurrence.

Looking at the family of copula selected for both one month and three months, most of the families exhibit higher upper tail dependence apart from Frank and Gaussian, which don't exhibit upper or lower tail dependence and are more suitable for capturing moderate dependencies without significant tail dependence. In contrast, most BB families, gumbel, joe, and student's t exhibit either both tail dependence or higher upper tail dependence. High upper tail dependence structures mean higher correlation at upper values of variables and are typically seen in fast-responding catchments, where higher slopes and the absence of abstractions.

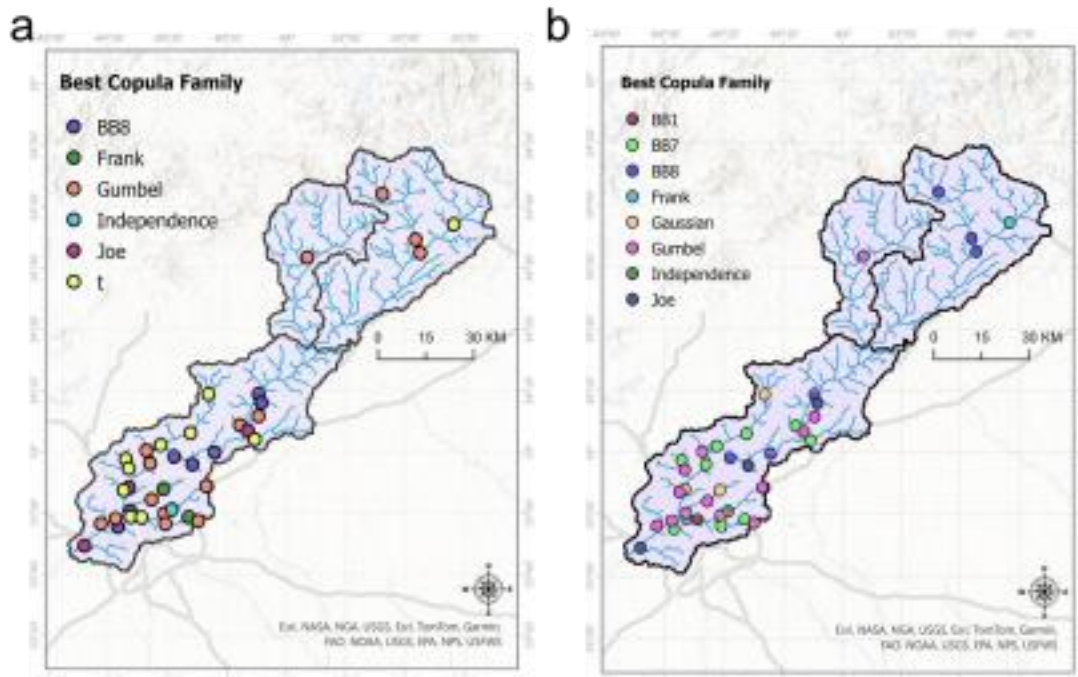
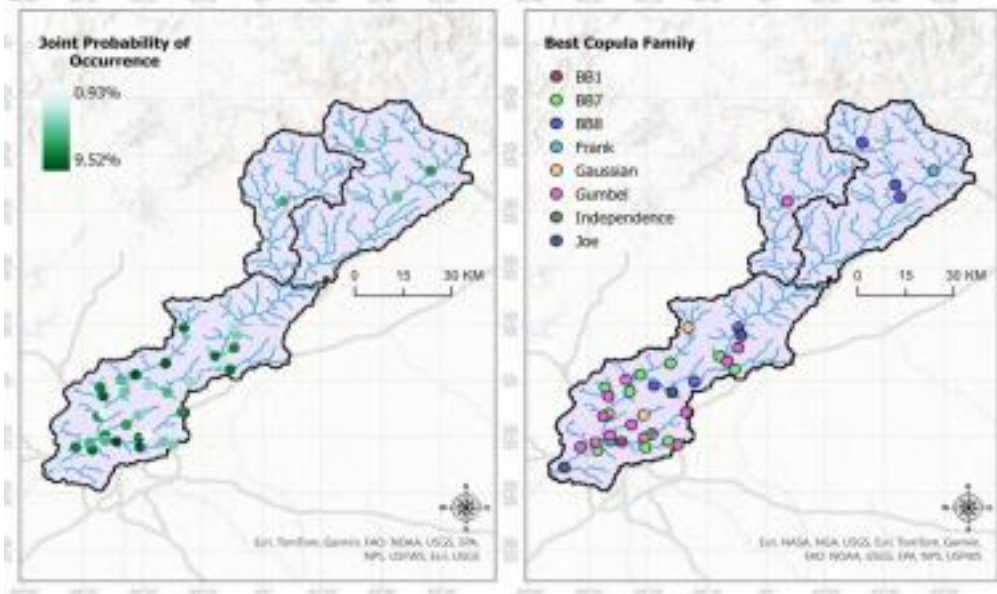


Figure 4: Best family of Copula identified for catchments for a) SPI1-SSI1 and b) SPI3-SSI3 **4.3 Joint probability of occurrence**

For both 1-month and 3-month scenarios, the joint probability of occurrence of meteorological drought and hydrological drought is similar statistically but low compared to conditional probability. Furthermore, it increases as we move downstream, which makes sense because streamflow downstream is related to precipitation due to the higher catchment area.



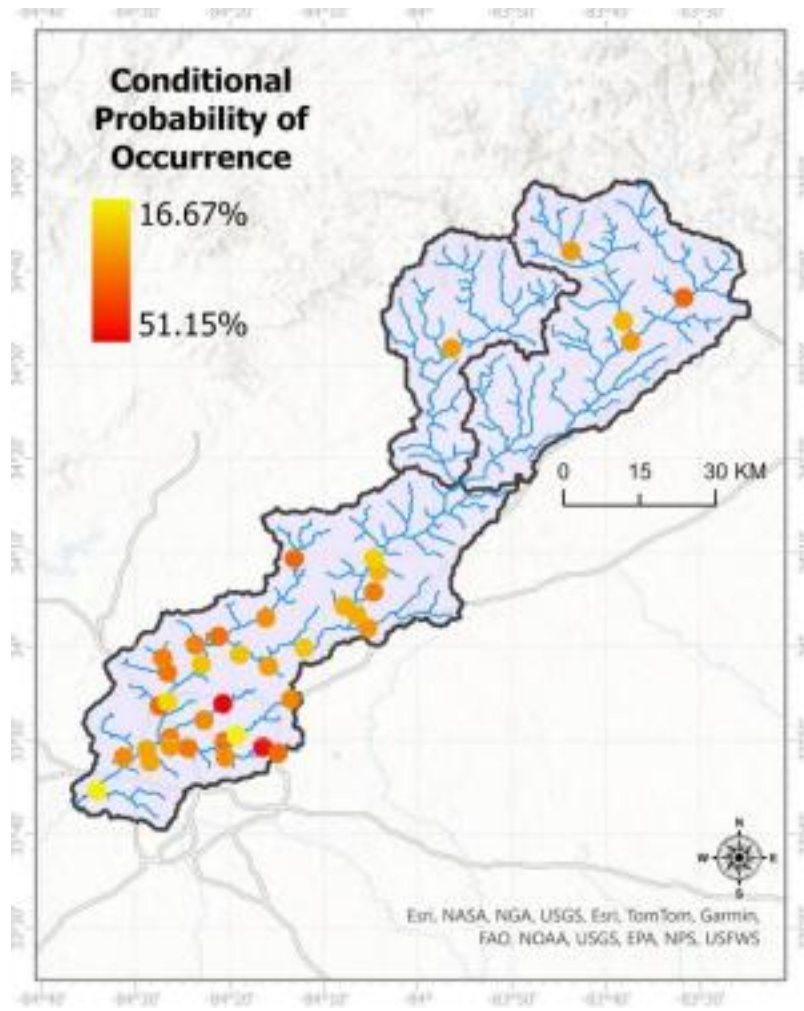


Figure 5: Joint probability of occurrence of $SPI \leq -1$ & $SSI \leq -1$ identified for catchments for a) SPI1-SSI1 and b) SPI3-SSI3



3.4 Conditional Probability of Occurrence

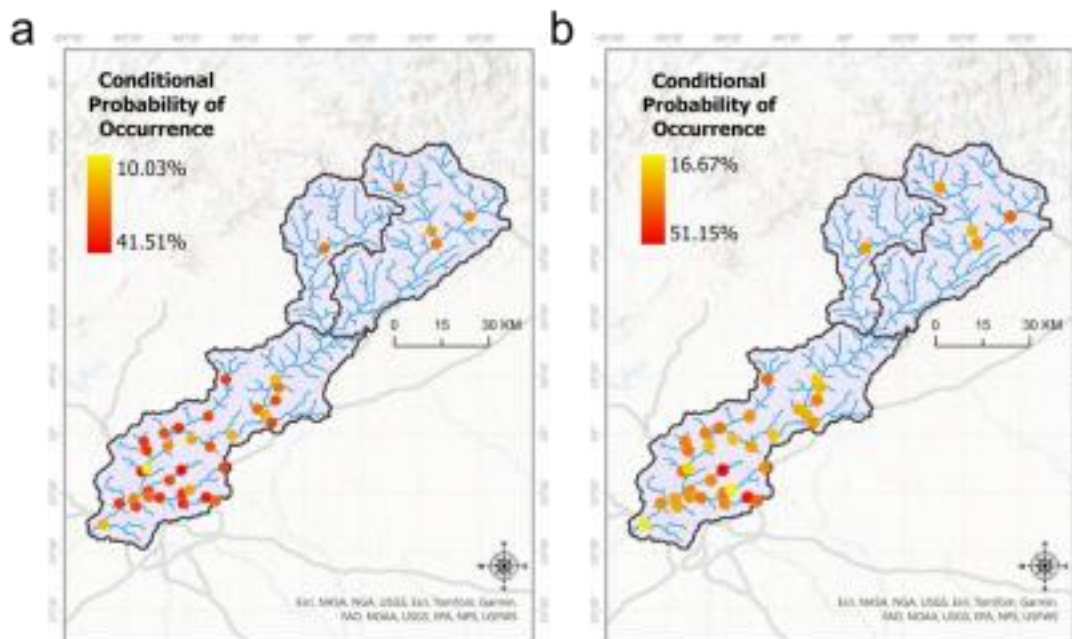


Figure 6: Conditional probability of occurrence of $(SSI \leq -1 \mid SSI = -1)$ for catchments for a) SPI1-SS11 and b) SPI3-SS13

While the joint probability of occurrence was found particularly low, the conditional probability signifying the occurrence of a moderate to severe streamflow drought, provided precipitation drought has already occurred, shows higher values in both 1 and 3-month accumulation scales of the indices. Similar to joint probabilities, increasing conditional probabilities are observed in the lower sections of ACT watershed.

4 Conclusions

This study provides an analysis of drought propagation using copula models. It focuses on the simultaneous occurrence of meteorological and hydrological droughts as well as the conditional probability that a hydrological drought occurs given a meteorological drought. By leveraging the HydroLang infrastructure, we developed a web-based tool that extracts precipitation and discharge data from large-scale services such as USGS that were used to estimate drought indices.

Our findings highlight the importance of understanding how precipitation deficits (meteorological droughts) translate into streamflow deficits (hydrological droughts), as past analyses have indicated the lack of clear-cut thresholds determining when meteorological droughts transition into hydrological droughts. Distinct periods of drought events with notable overlap between severe meteorological and hydrological droughts were observed. The use of bivariate copulas allowed us to model the dependence between precipitation and streamflow deficits and captured the nuanced interaction between these variables over 1- and 3-month time scales. Most downstream stations showed a greater probability of a hydrologic drought occurrence as a result of meteorological drought occurrence, particularly in the 3-month analysis, indicating the influence of seasonally -



aggregated rainfall and streamflow. This high probability suggests that when one anomaly is observed, the likelihood of occurrence of the other anomaly is also elevated, which can enhance our ability to predict drought conditions more reliably. The incorporation of additional catchment and climate controls such as elevation, evapotranspiration, and landcover can enhance these outputs by explaining the presence of low-value conditional probabilities among high-value ones.

Future studies should consider applying the copula model to a global scale to assess its effectiveness in various climatic and hydrological contexts. Additionally, incorporating soil moisture data into the framework could provide a more holistic view of drought impact and improve predictive capabilities.

Acknowledgement

We extend our sincere thanks to the WaterSoftHack initiative and its collaborative efforts with Clemson University, the University of Iowa, and the Consortium of Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI). Special thanks to Dr. Vidya Samadi for her leadership and guidance throughout the project. We are also indebted to our co-principal investigators, team members, and fellows who contributed their expertise and invaluable insights to this research. Additionally, we appreciate the contributions of the Glenn Department of Civil Engineering at Clemson University and the Department of Earth and Environment at Florida International University for their support and collaboration.

References

List of References

- ALIPOUR, A., AHMADALIPOUR, A., ABBASZADEH, P. & MORADKHANI, H. 2020. Leveraging machine learning for predicting flash flood damage in the Southeast US. *Environmental Research Letters*, 15, 024011.
- BALEW, A., LEGESE, B. & SEMAW, F. 2021. A Review of Some Indices Used for Drought Monitoring.
- BORUFF, B. J., EASOZ, J. A., JONES, S. D., LANDRY, H. R., MITCHEM, J. D. & CUTTER, S. L. 2003. Tornado hazards in the United States. *Climate Research*, 24, 103-117.
- CHO, S. Y. & CHANG, H. 2017. Recent research approaches to urban flood vulnerability, 2006–2016. *Natural Hazards*, 88, 633-649.
- CUTTER, S. L., EMRICH, C. T., MORATH, D. & DUNNING, C. 2013. Integrating social vulnerability into federal flood risk management planning. *Journal of Flood Risk Management*, 6, 332-344.
- DAI, A. 2011. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change*, 2, 45-65.



- DING, Y., HAYES, M. J. & WIDHALM, M. 2011. Measuring economic impacts of drought: a review and discussion. *Disaster Prevention and Management: An International Journal*, 20, 434-446.
- ENGSTRÖM, J., JAFARZADEGAN, K. & MORADKHANI, H. 2020. Drought vulnerability in the United States: An integrated assessment. *Water*, 12, 2033.
- FARAHMAND, A., & AGHAKOUCHAK, A. (2015). A generalized framework for deriving nonparametric standardized drought indicators. *Advances in Water Resources*, 76, 140-145.
- GESCH, D., OIMOEN, M., GREENLEE, S., NELSON, C., STEUCK, M. & TYLER, D. 2002. The national elevation dataset. *Photogrammetric engineering and remote sensing*, 68, 5-32.
- HAGENLOCHER, M., MEZA, I., ANDERSON, C. C., MIN, A., RENAUD, F. G., WALZ, Y., SIEBERT, S. & SEBESVARI, Z. 2019. Drought vulnerability and risk assessments: state of the art, persistent gaps, and research agenda. *Environmental Research Letters*, 14, 083002.
- IPCC 2012 Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change ed CB Field et al. (Cambridge: Cambridge University Press)
- KEELLINGS, D. & ENGSTRÖM, J. 2019. The future of drought in the southeastern US: projections from downscaled CMIP5 models. *Water*, 11, 259.
- Khedun, C. P., Mishra, A. K., Singh, V. P., & Giardino, J. R. (2014). A copula-based precipitation forecasting model: Investigating the interdecadal modulation of ENSO's impacts on monthly precipitation. *Water Resources Research*, 50(1), 580-600.
- KONAPALA, G. & MISHRA, A. 2017. Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. *Journal of Hydrology*, 555, 600-620.
- MADADGAR, S. & MORADKHANI, H. 2013. Drought analysis under climate change using copula. *Journal of hydrologic engineering*, 18, 746-759.
- MCKEE, T. B., DOESKEN, N. J. & KLEIST, J. The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, 1993. Boston, 179-183.
- MUKHERJEE, S., MISHRA, A. & TRENBERTH, K. E. 2018. Climate change and drought: a perspective on drought indices. *Current climate change reports*, 4, 145-163.



- NARESH KUMAR, M., MURTHY, C., SESA SAI, M. & ROY, P. 2009. On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 16, 381-389.
- NASIRI, H., MOHD YUSOF, M. J. & MOHAMMAD ALI, T. A. 2016. An overview to flood vulnerability assessment methods. *Sustainable Water Resources Management*, 2, 331-336.
- PITA, G., PINELLI, J.-P., GURLEY, K. & MITRANI-REISER, J. 2015. State of the art of hurricane vulnerability estimation methods: A review. *Natural Hazards Review*, 16, 04014022.
- RAJSEKHAR, D., SINGH, V. P. & MISHRA, A. K. 2015. Integrated drought causality, hazard, and vulnerability assessment for future socioeconomic scenarios: An information theory perspective. *Journal of Geophysical Research: Atmospheres*, 120, 6346-6378.
- RUDDY, B. C. & HITT, K. J. 1990. Summary of selected characteristics of large reservoirs in the United States and Puerto Rico, 1988. US Geological Survey.
- SALIMI, H., ASADI, E. & DARBANDI, S. 2021. Meteorological and hydrological drought monitoring using several drought indices. *Applied Water Science*, 11, 11.
- SENKBEIL, J. C., SCOTT, D. A., GUINAZU-WALKER, P. & ROCKMAN, M. S. 2014. Ethnic and racial differences in tornado hazard perception, preparedness, and shelter lead time in Tuscaloosa. *The Professional Geographer*, 66, 610-620.
- SONG, J. Y., ALIPOUR, A., MOFTAKHARI, H. R. & MORADKHANI, H. 2020. Toward a more effective hurricane hazard communication. *Environmental Research Letters*, 15, 064012.
- STRADER, S. M., ASHLEY, W. S., PINGEL, T. J. & KRMENEC, A. J. 2017. Projected 21st century changes in tornado exposure, risk, and disaster potential. *Climatic Change*, 141, 301-313.
- TRENBERTH, K. E., DAI, A., VAN DER SCHRIER, G., JONES, P. D., BARICHIVICH, J., BRIFFA, K. R. & SHEFFIELD, J. 2014. Global warming and changes in drought. *Nature Climate Change*, 4, 17-22.
- UNCCD2016 The ripple effect: A Fresh Approach to Reducing Drought Impacts and Building Resilience (https://unccd.int/sites/default/files/documents/27072016_The_ripple_effect_ENG.pdf)
- VAN LOON, A., VAN HUIJGEVOORT, M. & VAN LANEN, H. 2012. Evaluation of drought propagation in an ensemble mean of large-scale hydrological models. *Hydrology and Earth System Sciences*, 16, 4057-4078.
- VAN LOON, A. F., GLEESON, T., CLARK, J., VAN DIJK, A. I., STAHL, K.,



HANNAFORD, J., DI BALDASSARRE, G., TEULING, A. J., TALLAKSEN, L. M. & UIJLENHOET, R. 2016. Drought in the Anthropocene. *Nature Geoscience*, 9, 89-91.

VICENTE-SERRANO, S. M., LOPEZ-MORENO, J. I., BEGUERIA, S., LORENZO-LACRUZ, J., AZORIN-MOLINA, C., & MORAN-TEJEDA, E. (2012). Accurate computation of a streamflow drought index. *Journal of Hydrologic Engineering*, 17(2), 318-332.

WANG, W., ERTSEN, M. W., SVOBODA, M. D. & HAFEEZ, M. 2016. Propagation of drought: from meteorological drought to agricultural and hydrological drought.



Stream Temperature Variability and its Relationship to Large-Scale Climate Indices across the United States

Claudia Corona¹, Mohamed Abdelkader², Steve Yoon³, Cielo Sharkus⁴, Ochithya Fernando⁵
¹Colorado School of Mines, ²Stevens Institute of Technology, ³Southern Methodist University,
⁴University of Massachusetts, ⁵University of South Florida
Advisor: Vidya Samadi, Clemson University

Abstract

Stream water temperatures (SWT) are crucial indicators of aquatic ecosystem health, which can be strongly influenced by climate patterns, disturbance events and anthropogenic factors. Understanding the dynamics of SWT variability across different geographical and climatic zones is essential for effective water resource management and ecological conservation. The purpose of this study is to analyze recent trends in SWT field data across the United States (U.S.) and assess the impacts of large-scale climate variability on SWT observations over time. Specifically, the study investigates how SWT correlates with major climate indices, including the North Atlantic Oscillation (NAO), Atlantic Oscillation (AO), El Niño Southern Oscillation (ENSO), and Pacific North American pattern (PNA). Examining field data from the U.S. Geological Survey's National Water Information System Database, we ultimately identified 883 stream gaging stations with 10 or more years of daily SWT observations for further analysis.

We first conduct a spatial-temporal examination of SWT variability as a function of aridity index, stream order and land cover/land use. We use a Mann-Kendall nonparametric trend test to evaluate the existence of increasing or decreasing trends. Analysis of SWT trends finds that 70% of USGS monitored stream gage stations across the U.S. have experienced a statistically significant increase (at the 5% level) in daily SWT. Of the hydrologic unit codes (HUC) considered, the South Atlantic-Gulf region experienced the most prominent increasing SWT trends, while the Great Basin region experienced the least. In terms of aridity index, nearly 75% of stations considered to be in "humid" areas have experienced increasing trends in SWT, followed by semiarid station areas, followed by subhumid stations and arid stations. Finally, we consider how global climatic oscillations (NAO, AMO, ENSO, PNA) can influence local SWT conditions. By understanding the linkage between global climate patterns and regional water temperatures, policymakers and environmental managers can better expect changes and develop strategies to mitigate adverse effects on freshwater habitats and biodiversity.

Keywords: stream temperature, thermal trend analysis, climate change, global warming

1. Introduction

Stream water temperature (SWT) is an important driver of ecosystem health, and societal function. As part of the hydrologic cycle, SWT regulates dissolved oxygen concentrations, biochemical oxygen demand rates, and chemical toxicities (Patra et al., 2015), which ultimately have a strong influence on the health, survival, and distribution of freshwater fish, amphibians and other riparian species (Ward, 1998; Rogers et al., 2020). SWT is also a key indicator of cumulative anthropogenic impacts on freshwater systems (Risley et al., 2010). For example, SWT observations and trend analysis can show the extent of impacts of anthropogenic stream flow regulation such as the impacts of dam structures, channelization or stream re-routing



(Risley et al., 2010; Fuller et al., 2023). SWT analysis can also be a harbinger of awareness of the impacts of disturbance (i.e., riparian alteration, forest management, wildfire) and large-scale climate change impacts (Barbarossa et al., 2021; Johnson and Jones, 2000).

Over the last century, technological advances have improved our ability to measure SWT in a dependable manner across the United States (Caissie et al., 1998; Benyahya et al., 2007). In recent decades, the increase for SWT data available has also made data analysis of stream temperature possible and strengthened our scientific ability to look for trends (Maheu et al., 2016; Siegel et al., 2023). As a result, the scientific community has increasingly moved towards data accessibility, an example being the United States Geological Survey (USGS) National Water Dashboard, an interactive web tool that any user can use to find extensive site-specific hydrological data from thousands of stream gages across the United States (Miller et al., 2022). Currently, the USGS National Water Dashboard (here, ‘USGS Water Dashboard’) visualizes changes in certain hydrological parameters expertly, such as showing color variations for increases or decreases in stream discharge (water flow rate) over time. Addressing such variations in real time with color and comparative information on trends can be a useful tool for stakeholders needing a rapid, in-context understanding of real-time trends prior to making critical decisions such as dam water releases, agricultural effluent or industrial wastewater flow rate allowances. Unfortunately, the USGS Water Dashboard does not extend such visualizations to other hydrologic parameters of concern, particularly stream temperature, the focus of our project.

Unfortunately, the USGS Water Dashboard does not extend such visualizations to other important hydrologic parameters, in particular stream temperature, which is an important knowledge gap that needs addressing to enhance stakeholders’ ability to access and data and understand observational variability. Our project objective is thus two-fold: 1) to create an interactive web dashboard of all stream gaging stations with 10+ years of stream temperature observations for the United States, and 2) to analyze the 10+ years of stream temperature observations for overall trends in SWT variability over time, and determine possible influences on SWT variability by (a) hydrologic unit code (HUC), (b) aridity, and (c) stream order. The significance of this work is in its never-before-seen ability to organize national stream temperature data in a user-friendly manner to further support a comprehensive view of SWT variability across the United States, while also using the data to discover long-term trends that were previously not known in a larger spatial-temporal context.

2. Methodology

2.1 Stream Water Temperature Data Stream Water Temperature (SWT) Data

The United States Geological Survey (USGS) operates a comprehensive network of stream water temperature monitoring stations across the nation. These stations provide critical data for understanding temporal and spatial variations in stream water temperatures, which are essential for water resource management and ecological studies. The locations of these monitoring stations were sourced from the USGS National Water Information System Database (USGS, 2017), which is currently accessible through the USGS National Water Dashboard (<https://dashboard.waterdata.usgs.gov>).

While USGS provides several tools for retrieving streamflow data, such as water level and discharge, through its APIs, there was no direct facility available to download large-scale water



temperature datasets. To address this gap, we developed a Python-based automated code that facilitates the retrieval of SWT data across all USGS gauging stations for extended periods. This automation aimed to support large-scale analyses without the limitations imposed by manual data retrieval processes.

2.2 Automated Data Retrieval Code

The core of our methodology involves an automated data retrieval script written in Python. The script interacts directly with the USGS web services to fetch daily water temperature data for a specified range of dates. An overview of the script's functionality is summarized in Table 1.

Table 1. Key functionalities and the SWT data retrieval code

Functionality	Description
Initialization	The script begins by reading a list of station numbers from a CSV file. These station numbers are formatted to match USGS requirements (e.g., ensuring eight-digit uniformity).
Data Request	For each station, the script constructs a URL to query the USGS API, specifying the water temperature parameter code ('00010'). It requests data between October 1, 1980, and September 30, 2023, or based on the user selected time period.
Data Retrieval and Storage	Upon receiving a successful response, the script parses the JSON data and stores it in a directory named 'Data', creating a separate JSON file for each station. If the data retrieval fails (e.g., due to server issues or an invalid station number), the script logs an error message.
Automation and Execution	The entire process is automated allowing for periodic or on demand execution, such that data is up-to-date for ongoing research needs.

To foster collaboration and ensure transparency in our research, the station codes used in this study, along with the automated data retrieval script, will be made available as a HydroShare resource. This initiative aims to provide the scientific community with the tools necessary to replicate our study, validate our findings, and conduct further research on stream water temperatures.

2.3 Data Processing and Quality Control

The robustness of our findings hinges on the reliability and consistency of the SWT data obtained from the USGS stations. To ensure the integrity of our analysis, a comprehensive data preprocessing and curation step was implemented. This step involved the detailed examination of each station's data for completeness and accuracy.

The automated data retrieval code, described previously, serves as the foundation for this process. Following the initial data retrieval, a data investigation process was employed to parse and assess the quality of the water temperature records from each station. This process performs several tasks.

The preprocessing of SWT data involves a series of steps that transform raw measurements into a format suitable for detailed analysis. Initially, the raw JSON data from each USGS station file is meticulously converted into a structured panda DataFrame. During this conversion process, date entries are parsed into a datetime format to facilitate chronological analysis, while water



temperature values are converted to numeric types, with non-numeric entries being handled gracefully to ensure data integrity.

Following the conversion, the completeness of the data is assessed by calculating the total number of days covered by the records and comparing this to the actual count of data entries logged. This comparison is crucial as it highlights any significant gaps in the data that could potentially skew the analysis of trends and variability. To further ensure the robustness of the dataset, a statistical summary that includes the minimum, maximum, median, and mean temperatures is computed. These statistics provide an initial quality check and a snapshot of the range of conditions experienced at each station, facilitating preliminary interpretations and comparisons. The final stage of preprocessing involves applying strict data retention criteria, ensuring that only stations with at least ten years of daily records are included in the analysis. This criterion is essential for the reliable detection of long-term trends and variability within the dataset. The output from this preprocessing includes a comprehensive summary of each station's data quality metrics. These metrics are instrumental in determining the suitability of the data for inclusion in our trend analysis, ensuring that the dataset used for further study represents a reliable and scientifically robust basis for evaluating the impacts of various environmental factors on stream water temperatures.

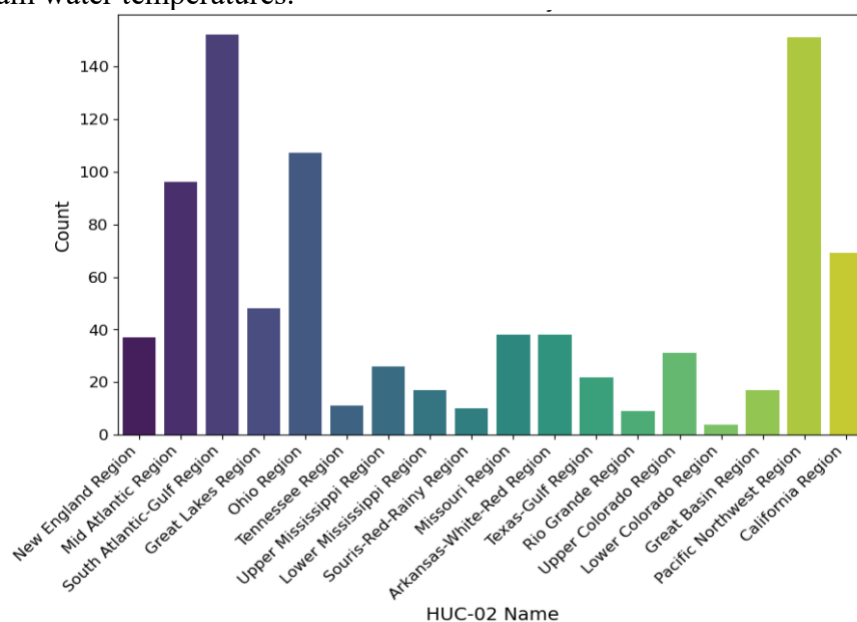


Figure 1. Distribution of USGS SWT monitoring stations across major hydrological units from the East Coast of the United States to the West coast.

Of the 2607 active USGS stations initially considered, 883 stations met the criteria discussed above for long-term data completeness and reliability. This selection was based on the requirement that each station have at least ten years of daily records, ensuring robust trend analysis and change point detection capabilities. To illustrate the geographical distribution and hydrological context of the selected stations, Figure 1 provides the count of the 883 stations, categorized by their respective Hydrological Unit Codes (HUCs) organized from East Coast to West Coast. We note that the South Atlantic-Gulf region, Ohio region, and Mid-Atlantic regions on the East Coast have more stream gaging stations than the West Coast regions, i.e. California and Pacific Northwest.



2.4 Auxiliary Data

To comprehensively assess the impacts of climatic, hydrological, and anthropogenic factors on stream water temperature variability, we obtained several datasets reflecting these elements. These datasets are essential for understanding the multifaceted influences on streamflow and water temperature regimes. The following summary (Table 2) provides an overview of the data used in our analysis.

Table 2. Summary of data type obtained, source, description, and time frame considered.

Data type	Source	Description/Link	Time Frame
Global climate oscillations	NOAA Physical Sciences Laboratory	NAO, AO, ENSO, and PNA time series. https://psl.noaa.gov/data/climateindices/list/	1980-2023
LC/LU data	USGS National Land Cover Database	https://www.mrlc.gov/data	2021
Aridity index	Abdelkader et al. (2023)	Aridity index was calculated as a function of the mean annual precipitation and mean annual potential evapotranspiration.	1979-2020
Stream order	National Hydrography Dataset Plus	https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus	-
Hydrologic Unit Codes (HUC-02)	USGS Watershed Boundary Dataset (WBD)	https://www.usgs.gov/national-hydrography/watershed-boundary-dataset	-

2.5 Variability Analysis

In this study, the assessment of variability in SWT data series is primarily performed through trend analysis, using the **Mann-Kendall (MK) test** to detect significant monotonic trends (Kendall, 1975). The MK test is a robust non-parametric statistical method widely used for analyzing time series data where the sequence of values shows either an increasing, decreasing, or no trend over time. This method does not require the data to conform to any distribution and is insensitive to abrupt breaks due to inhomogeneous time series data. Therefore, it is especially suitable for environmental studies where data may be non-normal. (Helsel and Hirsch, 2002). The test evaluates the strength and direction of a trend by comparing the relative magnitudes of sample data points. The null hypothesis of the MK test states that there is no trend, which is rejected if the test statistics exceeded the significance level value, where:

$$H_0: \text{Probability } [Y_j > Y_i] = 0.5, \text{ where time, } T_j > T_i$$

$$H_1: \text{Probability } [Y_j > Y_i] \neq 0.5 \text{ (2-sided test)}$$

The null (H_0) hypothesis expresses the existence of no trend while the (H_1) alternative hypothesis highlights a significant rising or declining trend in the data. Based on a 5% significance level, where the p-value < 0.05 , then the alternative hypothesis (H_1) is accepted, thereby signifying the presence of a trend in the data. If the p-value > 0.05 , then the null hypothesis (H_0) is accepted, thereby suggesting the absence of a trend in the data.



The test has been applied in hydrologic studies to analyze temporal changes in environmental records, such as examining changes in daily maximum precipitation (Westra et al., 2013). A strength of the Mann-Kendall test for trend is that it is not biased by any transformations (i.e., logarithmic, exponential) applied to the Y-variable, making it suitable for studies of many similar datasets, such as 10+ year records of stream water temperature (Helsel and Hirsch, 2002), as is the case for this study.

To complement the trend significance testing with the Mann-Kendall test, **Sen's Slope Estimator** is employed to quantify the magnitude of the observed trend. The Sen's slope test is a robust, non-parametric test used to quantify the magnitude of the observed trends (Gocic and Trajkovic, 2013). The Sen's Slope test is a non-parametric method used to estimate the true slope (change per unit time) of a dataset, providing a robust measure against outliers and non-normal data distributions. For a set of pairs (i, X_i) , where X_i is change over time, Sen's slope can be defined as:

$$d_k = \frac{X_j - X_i}{j - i} \text{ for } 1 \leq i < j \leq n$$

Where the d is the slope, X denotes the variable, n is the number of data, and i, j are indices. From this equation, Sen's slope is calculated as the median from all the computed slopes, where $b = \text{median } d_k$. The intercepts, a_t , are computed for timestep, t , using the following equation:

$$a_t = X_t - b * t$$

And the corresponding intercept is as well the median of all intercepts. For our analysis of SWT, the statistical tests were applied to over ten years of daily SWT data collected from the USGS gauging stations. This approach allowed us to systematically identify and quantify trends across various geographical and climatic zones within the United States. By detecting and measuring the rates of change in SWT, we can infer potential impacts linked to climatic oscillations, land use changes, and other anthropogenic activities, thus providing valuable insights for environmental management and policymaking.

The **Pettitt test** (Pettitt, 1979) is a nonparametric test that identifies a point at which the values in a series with continuous data change, called a change-point detection. Such an approach can be applied to hydrological or climate series datasets, where the null hypothesis (H_0) states that the T (time) variables follow one or more distributions showing no change, i.e., have the same location parameter compared to the alternative hypothesis (H_1) which states that a change point exists. The Pettitt test is defined as:

$$K_T = \max |U_{t,T}|,$$

where

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T \text{sgn}(X_i - X_j)$$

The change-point of the data series is located at K_T , provided a significant statistic. The significance probability of K_T is determined for the p-value ≤ 0.05 , where:



$$p \approx 2 \exp\left(\frac{-6K_T^2}{T^3 + T^2}\right)$$

The method is advantageous in identifying shifts in the median without presupposing a specific distribution for the data. The test operates by comparing data ranks across potential change points and calculating a statistic U_i , which is assessed against critical values from the asymptotic distribution. If the computed U_i exceeded the critical threshold, a change point is considered statistically significant. This method is particularly valuable for detecting abrupt changes in hydrological time series data.

3. Results and Discussion

3.1 SWT Variability Dashboard

In this study, we developed the SWT Variability Dashboard, a comprehensive web-based platform designed to advance the understanding of stream water temperature changes across the United States. This platform is structured around three key portals, each tailored to meet specific analytical and data access needs (Figure 2).

The first portal facilitates a dynamic exploration of recent trends in SWT. It features an interactive map displaying sites with increasing, decreasing, and stable SWT trends. This portal integrates additional geographic layers such as aridity zones and locations of significant dams—those exceeding 100 feet in height—providing a multi-dimensional perspective on the environmental and anthropogenic factors influencing SWT.

The second portal is dedicated to the retrieval of stream temperature data, enabling users to directly access and download data. This functionality supports a broad spectrum of users, from researchers conducting complex analyses to practitioners requiring immediate data for environmental management applications.

The third portal offers detailed information about active USGS SWT stations, allowing users to access basic statistical data for each station and related hydraulic features, such as nearby dams. This information is crucial for understanding the local conditions that might influence water temperatures at each monitoring site.

Significantly, the second and third portals are integrated with HydroShare, hosted as resources that direct users to CUAHSI's JupyterHub. This integration facilitates further data manipulation and analysis within a collaborative environment, enhancing the utility of the dashboard for comprehensive scientific research.



Figure 2. Interface of the SWT variability across the United States web portal

Figure 2 provides a visual representation of the dashboard's user interface, highlighting its functionality and user-friendly design. Overall, the SWT Variability Dashboard represents a significant advancement in the tools available for researchers, policymakers, and environmental managers, offering an integrated platform to analyze, visualize, and respond to changes in stream water temperatures effectively. Users can access the platform using the following URL: <https://softhack-mabdelka1-c776225f2bb346fd6c8bc81dd0be3fd85f94edb48ba3.gitlab.io/>



3.2 SWT variability by HUC-02

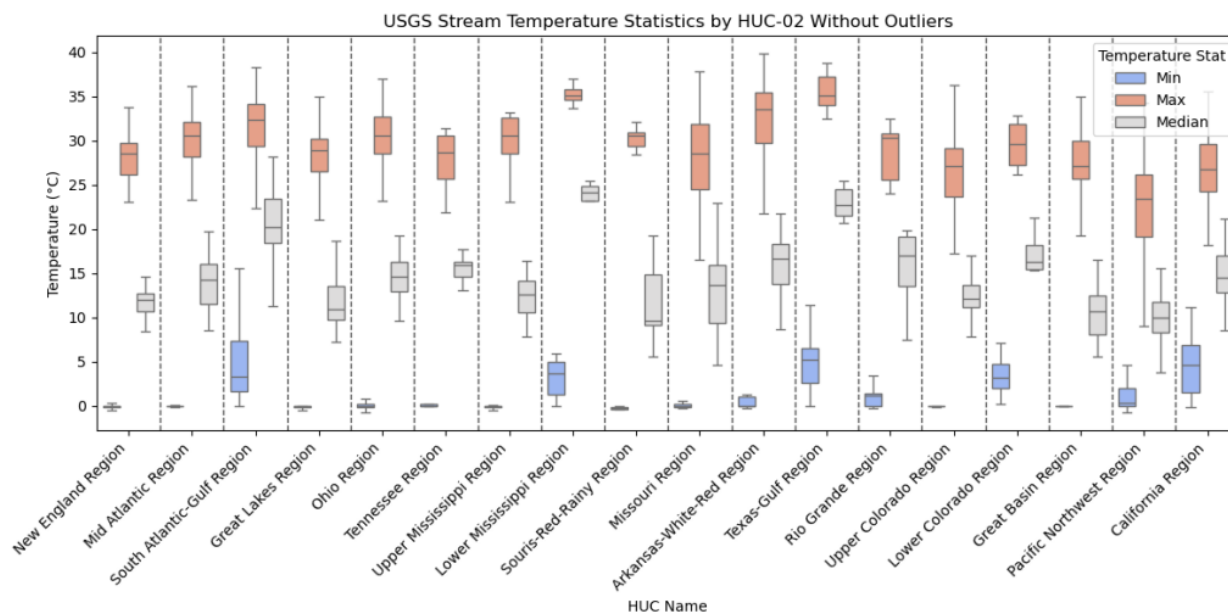


Figure 3. Box plot showing stream temperature observations minimum (blue), maximum (orange) and median (gray) values in order of HUC-02 region, organized from East Coast to West Coast.

Figure 4 organizes stream water temperature minimum, maximum and median statistics in box plot form by HUC-02 region from East Coast (New England region) to West Coast (California region). Overall, there does not appear to be an increasing or decreasing trend for stream gaging stations, by HUC-02 region. This is not surprising as HUC-02 regions encompass varying spatial scales that can add natural nonlinearity in the form elevation variability, climate variability, topographic variability, and aridity variability to data analysis. In contrast to the minimum/maximum/median box plots of stream water temperature, figure 3 shows a proportional analysis of trend types by HUC-02 region. Of interest are the trends showing increasing stream water temperatures across stream gaging stations, which appear to dominate across all HUC-02 regions. Specifically, the percentage of stream gaging stations depicting an increasing trend in SWT range from the Great Basin region showing ~33% of stations with increasing SWT over the 10+ year dataset, to the South-Atlantic Gulf region, which shows over ~90% of all stations experiencing an increasing trend in SWT. In contrast, decreasing trends in SWT range from ~18% (California and Missouri regions) to 0% (Rio Grande and Texas-Gulf regions). The percentage of stations with no discernible trend in SWT ranged from ~10% (Ohio region) to nearly ~60% (Great Basin region).

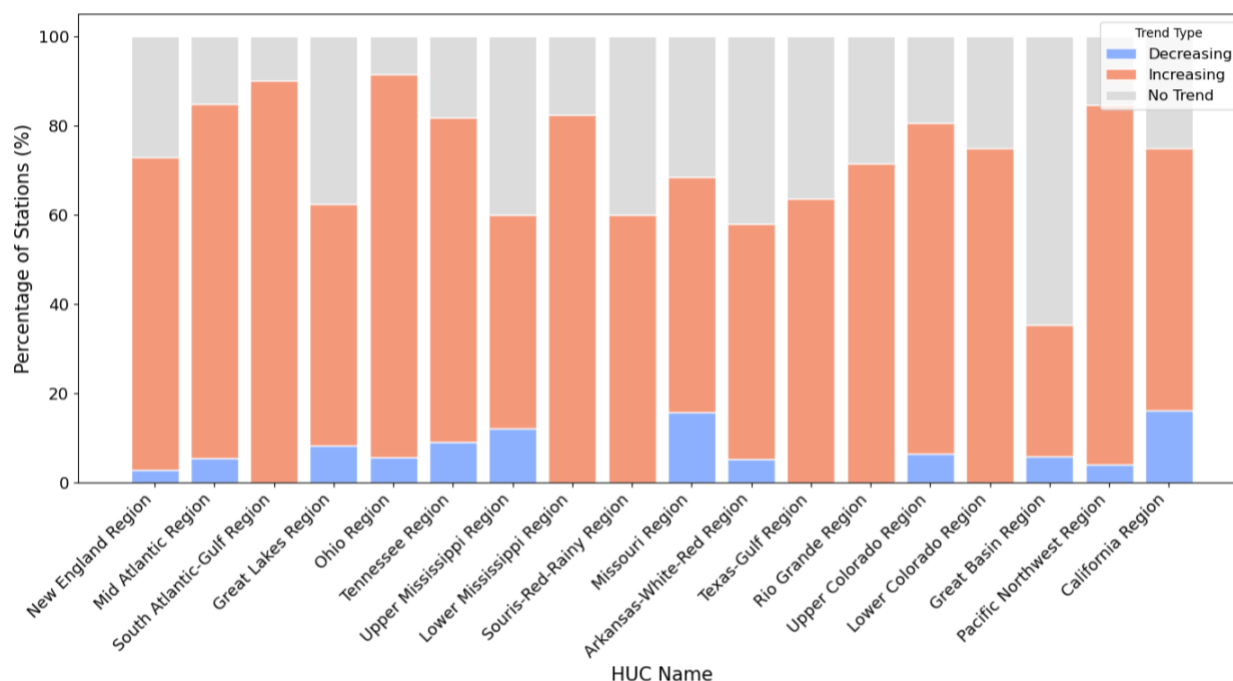


Figure 4. Bar plot shows proportional analysis of trend types, i.e., decreasing (blue), no trend (gray) and increasing (orange) by HUC-02 region, organized from East Coast to West Coast.

3.3 SWT Trends by Aridity Index

Figure 5 organizes stream water temperature minimum, maximum and median statistics in box plot form by aridity index. The aridity index represents the average amount of water available in the soil and is physically defined as the ratio between mean annual precipitation of an area and the mean annual evapotranspiration (Arora, 2002). Arid soils are thus considered soils with the least amount of water available in the soil (on average), followed by semiarid soils, subhumid soils, and finally humid soils. Figure 4 shows the stations considered semiarid and humid as having the greatest variability in terms of maximum stream water temperature. Of the trend types shown, the decreasing trend is least variable across the four aridity zones.

The stream gage stations categorized as humid show the greatest variability for minimum, median, and maximum SWT, which is physically sensible because humid areas tend to be less water limited, allowing liquid water to reach the highest and lowest range of temperatures. Interestingly, stream gage stations classified as semiarid have the second highest variability. We would initially expect that subhumid stations would have the second highest variability due to being less water limited than semiarid regions. Finally, stations classified as arid had the least variability, which is physically intuitive due to such areas generally being severely water limited.

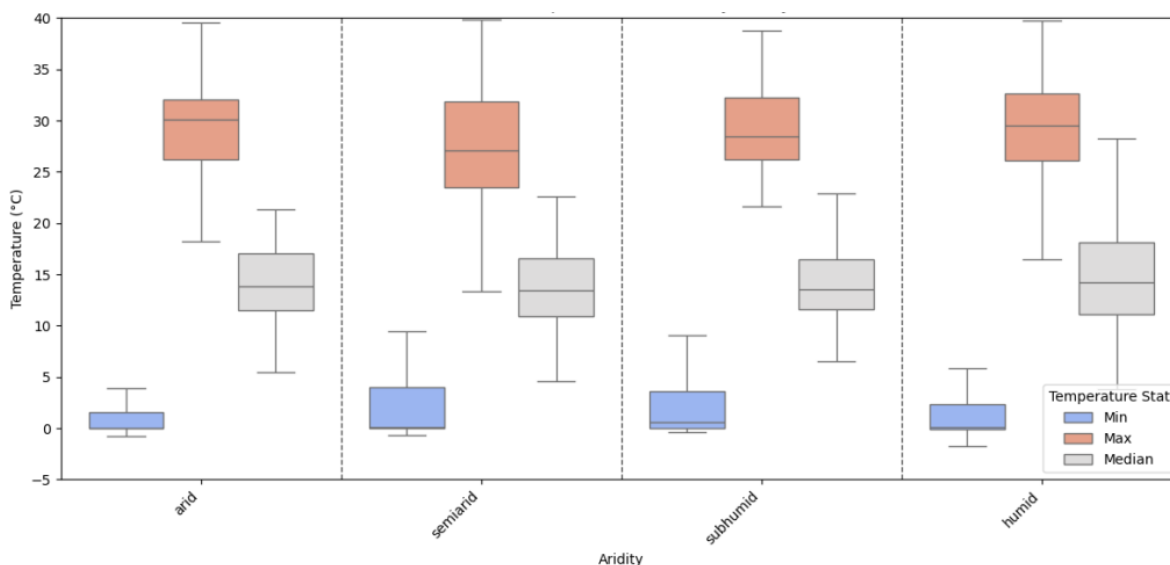


Figure 5. Box plot showing stream water temperature minimum (blue), median (gray), and maximum (orange) values in order of aridity index, from arid to humid regions.

In contrast to the minimum/maximum/median box plots of stream water temperature, figure 6 shows a proportional analysis of trend types by aridity index, by percentage of stations considered. Of note, increasing trends in stream water temperature are most dominant for stations defined as humid, with nearly 75% of stations considered to be in humid areas showed an increasing trend. Stations considered arid show the lowest percentage of increasing trends ~25%, followed by subhumid stations ~35% and semiarid stations ~40%. In contrast, decreasing trends in SWT range from less than 5% (arid) to ~8% (semiarid). The percentage of stations with no discernible trend in SWT ranged from ~10% (subhumid) to nearly ~25% (semiarid). Physically, it is understandable that the stations considered arid would have the lowest increasing trend as arid regions are generally water-limited, thereby impeding daily SWT variability (especially in the hot summer months), as there is likely little to no flowing surface water to be measured. At the same time, the higher percentage of increasing trends being experienced by semiarid regions compared to subhumid regions is interesting given that semiarid regions are also water-limited (though not as intensely as arid regions) relative to subhumid regions. We posit that this may be due to a greater presence of semiarid stations being at higher elevations, i.e., located in the Rocky Mountains or Cascade Mountains, where spatially close gages (that measure SWT) separated only by elevation may experience larger swings between colder and warmer temperatures, especially where affected by atmospheric rivers or other climate phenomenon, as was the case for small watersheds in the Sierra Nevada Mountains of California (Marshall et al., 2024).

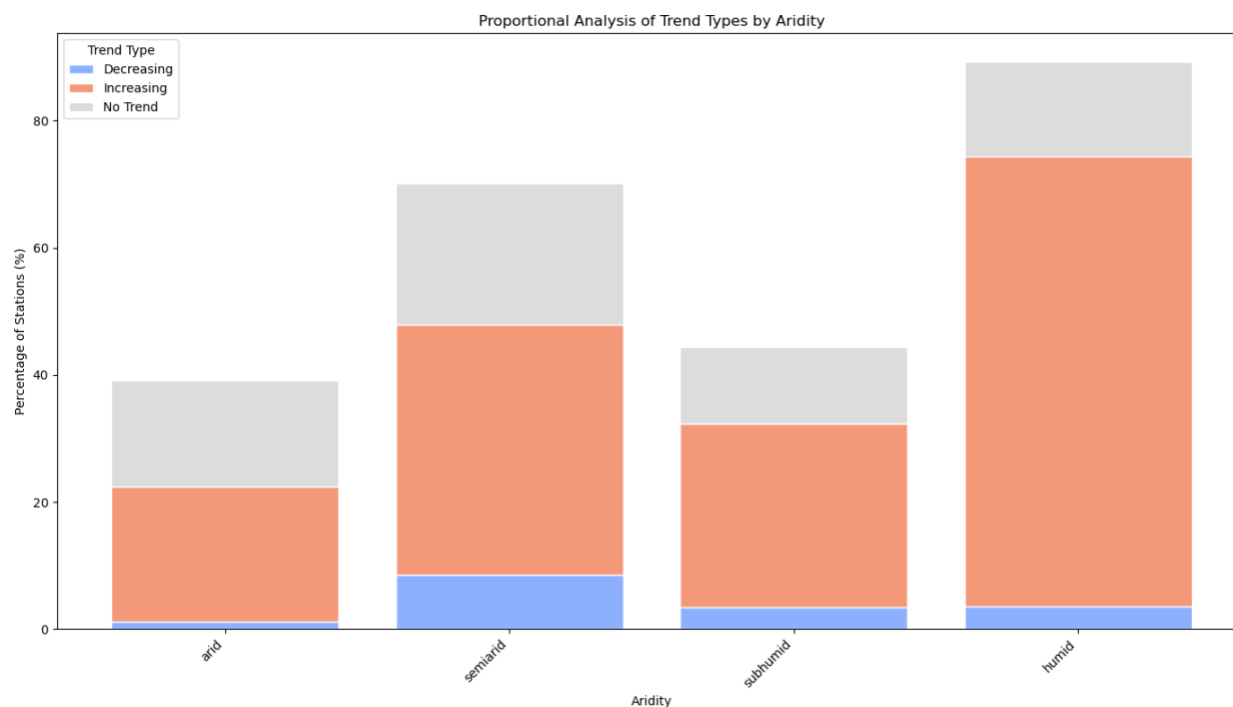


Figure 6. Bar plot shows proportional analysis of trend types, i.e., decreasing (blue), no trend (gray) and increasing (orange) by aridity index, organized from arid to humid.

3.4 SWT Trends by Stream Order

Figure 7 presents SWT statistics—minimum, median, and maximum—organized by stream order, distinguishing smaller-order streams as "lower" and larger-order streams as "higher." Surprisingly, the analysis reveals no significant trends in SWT variability across different stream orders. This outcome contradicts common hydrological assumptions. Typically, smaller streams, with less water volume, exhibit greater temperature fluctuations due to their limited heat capacity, leading to pronounced temperature swings. Conversely, larger streams, benefiting from greater water volumes, are expected to have more stable temperatures due to their enhanced heat buffering capacity (Leach et al., 2023).

The lack of observable variation in our continental-scale analysis suggests that the thermal dynamics of streams may not conform to simplified expectations based on stream order alone. This insight points to the potential oversimplification of complex hydrological and thermal interactions within stream networks at large scales. Consequently, there is a clear need for more detailed monitoring networks that cover both tributaries and main streams. Such networks would enhance our ability to detect localized changes and understand the nuances of thermal variability. Moreover, this finding underscores the importance of conducting more focused watershed-scale analyses. While continental-scale assessments provide broad overviews, they may mask critical subtleties and localized phenomena essential for effective water resource management and ecological conservation. Detailed analyses at the watershed scale could reveal significant variability and trends that continental assessments might overlook, offering more precise insights into the factors driving SWT variability. This approach would also facilitate more targeted



management strategies, tailored to the specific thermal and hydrological characteristics of individual watersheds.

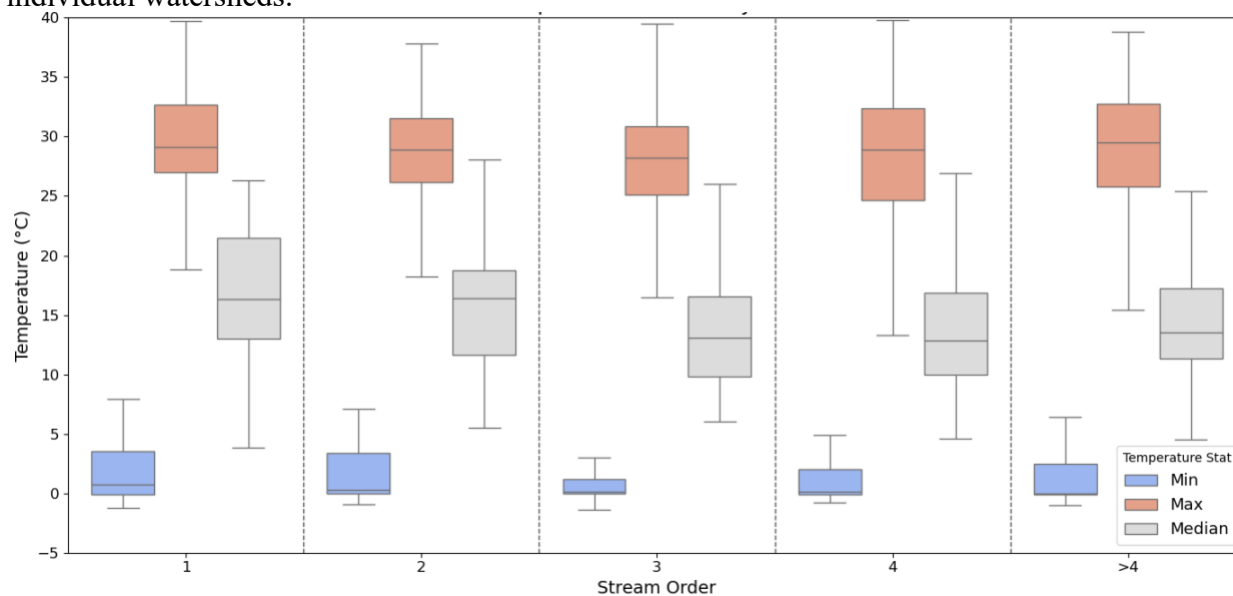


Figure 7. Box plot showing stream water temperature minimum (blue), median (gray), and maximum (orange) values organized by stream order, where a stream order of 1 indicates low-order (i.e., small) streams. Stream orders of > 4 indicate much larger streams (also called rivers) such as the Missouri river (7th order).

3.5 Link with climate oscillations

To explore the potential influence of large-scale climate patterns on stream water temperatures, our analysis focused on the correlation between maximum recorded SWT and the moving average over a three-month window of key climate indices—specifically ENSO, PNA, NAO, and AO. We selected three different stations located in the western United States, each exhibiting distinct trend patterns: increasing, decreasing, and no significant trend, all within the same Hydrological Unit.

Figure 8 illustrates the seasonal patterns in the relationship between maximum recorded SWT and selected climatic indices. Seasons marked with an asterisk denote the period preceding the water year of the recorded maximum SWT. Relationships extending outside the shaded region demonstrate significance at the 10% confidence level, with the highest correlation observed with the PNA for the October-November-December series. This suggests that Pacific patterns, particularly those associated with the PNA, significantly impact SWT variability.

The Pacific North American (PNA) pattern, known for causing strong fluctuations in the strength and location of the East Asian jet stream, plays a crucial role. During its positive phase, the PNA is associated with an enhanced East Asian jet stream and an eastward shift in the jet exit region toward the western United States, typically leading to increased precipitation rates and decreased SWT in these areas. Conversely, a negative phase tends to be associated with Pacific cold episodes which can increase SWT.

Our objective was to determine if, regardless of the local temperature trend, there was a consistent climatic signal affecting annual maximum SWT. The results were enlightening: we



found that a lagged summer signal from the PNA pattern, coupled with autumn signals from the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO), has the potential to serve as predictors for the annual maximum stream temperatures.

This linkage suggests that despite localized temperature trends, certain large-scale climate indices significantly impact SWT, offering valuable insights for predicting changes in stream temperatures based on global climate dynamics. The obtained negative correlation is expected given the influence of Pacific patterns on SWT variability, affirming the predictive power of these climatic indices.

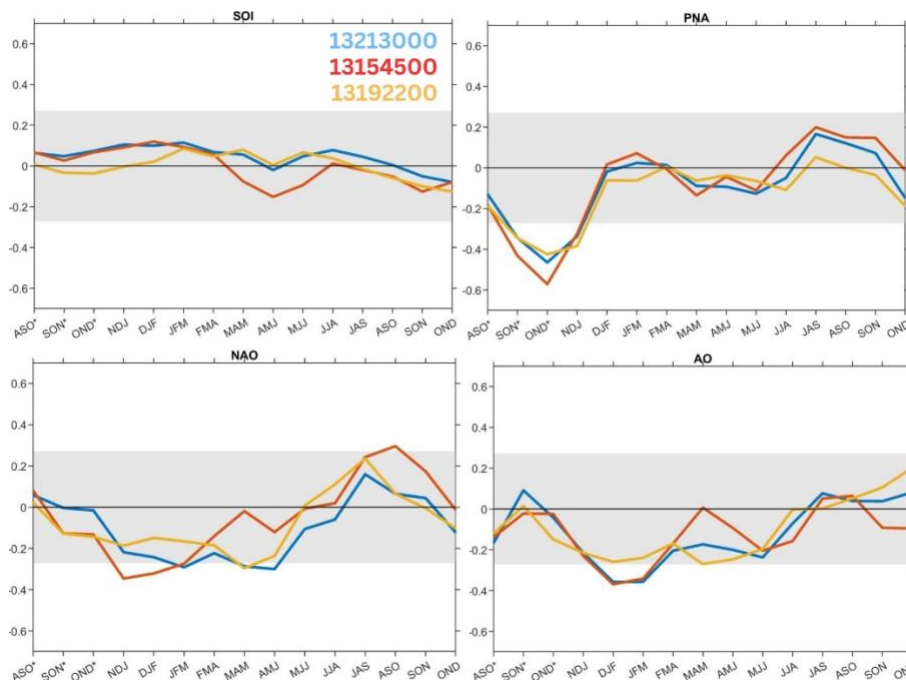


Figure 8. Seasonal patterns in the relationship between maximum recorded SWT and selected climatic indices. Seasons marked with an asterisk denote the period preceding the water year of the recorded maximum SWT. Relationships extending outside the shaded region demonstrate significance at the 10% confidence level.

4 Conclusions

Recent technological advances have improved our ability to measure SWT in a consistent manner across the U.S. (Caissie et al., 1998; Benyahya et al., 2007). The increase of available SWT data has also made data analysis of SWT possible and strengthened our ability as a scientific community to visually display data for public-use as well as identify physical patterns from the data in near real-time (Maheu et al., 2016; Siegel et al., 2023). To further our progress in data analysis of stream temperature, we introduce an interactive website dashboard for daily SWT, which to our knowledge was not previously in existence.

As our first project objective, the purpose of the web dashboard was to highlight observational variations in real time with color and comparative information on trends can be a useful tool for stakeholders needing a rapid, in-context understanding of real-time trends prior to making critical decisions such as dam water releases, agricultural effluent or industrial wastewater flow rate allowances. The second objective of the project was to analyze the 10+ years of stream



temperature observations for overall trends in SWT variability over time and determine possible influences on SWT variability by (a) HUC classification, (b) aridity index, and (c) stream order. In addressing the second objective, we identified increasing trends in SWT across the United States. By HUC classification, we found that increasing trends in SWT range from being present in 30% of stations (Great Basin region) to 90% of stations (South Atlantic-Gulf region). In terms of aridity index, we found that nearly ~75% of stations classified as being in humid areas experienced increasing trends in SWT variability over time, while ~25% of arid stations experienced increasing trends for SWT. In terms of stream order, there were no strong discernable trends in SWT variability over time, which is most likely due to data limitations, wherein smaller order streams have less representation in datasets relative to larger streams. Overall, the significance of this work is in its novel ability to organize national SWT data in a user-friendly manner to further support a comprehensive view of SWT variability across the United States, while also using the data to discover long-term trends that were previously not known in a larger spatial-temporal context. Possible future analysis will consider the inclusion of dams, non-point sources of pollution, and more in-depth examination on the influence of stream order on stream water temperature variability over time.

Acknowledgement

The authors express their gratitude to the WaterSoftHack team for their invaluable training and support throughout the Hackathon period.

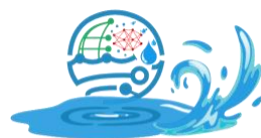


References

- Arora, V.K., 2002. The use of the aridity index to assess climate change effect on annual runoff. *Journal of hydrology*, 265(1-4), p.164-177. [https://doi.org/10.1016/S0022-1694\(02\)00101-4](https://doi.org/10.1016/S0022-1694(02)00101-4)
- Barbarossa, V., Bosmans, J., Wanders, N., King, H., Bierkens, M.F.P., Huijbregts, M.A.J., Schipper, A.M., 2021. Threats of global warming to the world's freshwater fishes. *Nat Commun* 12, 1701. <https://doi.org/10.1038/s41467-021-21655-w>
- Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T.B.M.J., Bobée, B., 2007. A Review of Statistical Water Temperature Models. *Canadian Water Resources Journal* 32, 179–192. <https://doi.org/10.4296/cwrj3203179>
- Caissie, D., El-Jabi, N., St-Hilaire, A., 1998. Stochastic modelling of water temperatures in a small stream using air to water relations. *Canadian Journal of Civil Engineering* 25.
- Fuller, M.R., Detenbeck, N.E., Leinenbach, P., Labiosa, R., Isaak, D., 2023. Spatial and temporal variability in stream thermal regime drivers for three river networks during the summer growing season. *J American Water Resour Assoc* 1752–1688.13158. <https://doi.org/10.1111/1752-1688.13158>
- Gocic, M., Trajkovic, S., 2013. Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. *Global and Planetary Change* 100, 172–182. <https://doi.org/10.1016/j.gloplacha.2012.10.014>
- Helsel, D.R., Hirsch, R.M., 2002. Chapter A3: Statistical Methods in Water Resources, in: *Techniques of Water Resources Investigations, Book 4*. U.S. Geological Survey, p. 522.
- Johnson, S.L., Jones, J.A., 2000. Stream temperature responses to forest harvest and debris flows in western Cascades, Oregon 57.
- Kendall, M.G., 1975. *Rank Correlation Methods*, 4th Edition. ed. Charles Griffin, London.
- Leach, J.A., Kelleher, C., Kurylyk, B.L., Moore, R.D. and Neilson, B.T., 2023. A primer on stream temperature processes. *Wiley Interdisciplinary Reviews: Water*, 10(4), p.e1643.
- Lettenmaier, D.P., 1988. Multivariate nonparametric tests for trend in water quality. *JAWRA Journal of the American Water Resources Association* 24, 505–512.
- Maheu, A., Poff, N.L., St-Hilaire, A., 2016. A Classification of Stream Water Temperature Regimes in the Conterminous USA. *River Research & Apps* 32, 896–906. <https://doi.org/10.1002/rra.2906>
- Marshall, A.M., Abatzoglou, J.T., Rahimi, S., Lettenmaier, D.P. and Hall, A., 2024. California's 2023 snow deluge: Contextualizing an extreme snow year against future climate change. *Proceedings of the National Academy of Sciences*, 121(20), p.e2320600121.
- Miller, M.P., Burley, T.E., McCallum, Brian E., 2022. USGS National Water Dashboard (Fact Sheet No. 2022–3003), *Water Priorities for the Nation*. United States Geological Survey, Reston, VA.
- Patra, R.W., Chapman, J.C., Lim, R.P., Gehrke, P.C., Sunderam, R.M., 2015. Interactions between water temperature and contaminant toxicity to freshwater fish. *Enviro Toxic and Chemistry* 34, 1809–1817. <https://doi.org/10.1002/etc.2990>
- Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28, 126–135.
- Risley, J.C., Constantz, J., Essaid, H., Rounds, S., 2010. Effects of upstream dams versus groundwater pumping on stream temperature under varying climate conditions: Upstream Dam and Groundwater Pumping Impacts. *Water Resour. Res.* 46. <https://doi.org/10.1029/2009WR008587>



- Rogers, J.B., Stein, E.D., Beck, M.W., Ambrose, R.F., 2020. The impact of climate change induced alterations of streamflow and stream temperature on the distribution of riparian species. PLoS ONE 15, e0242682. <https://doi.org/10.1371/journal.pone.0242682>
- Siegel, J.E., Fullerton, A.H., FitzGerald, A.M., Holzer, D., Jordan, C.E., 2023. Daily stream temperature predictions for free-flowing streams in the Pacific Northwest, USA. PLOS Water 2, e0000119. <https://doi.org/10.1371/journal.pwat.0000119>
- U.S. Geological Survey. 2016. “National Water Information System Data Available on the World Wide Web (USGS Water Data for the Nation).” <https://doi.org/10.5066/F7P55KJN>
- Ward, J.V., 1998. Riverine landscapes: Biodiversity patterns, disturbance regimes, and aquatic conservation. Biological Conservation 83, 269–278. [https://doi.org/10.1016/S0006-3207\(97\)00083-9](https://doi.org/10.1016/S0006-3207(97)00083-9)
- Westra, S., Alexander, L.V., Zwiers, F.W., 2013. Global Increasing Trends in Annual Maximum Daily Precipitation. Journal of Climate 26, 3904–3918. <https://doi.org/10.1175/JCLI-D-12-00502.1>



Appendix

Table A1. HUC-2 regions with their respective total stream gaging stations, and separation of stations by aridity index. See figure A1 for a United States map showing the HUC-2 regions.

HUC 2	Total Stations	Arid	Semiarid	Subhumid	Humid
New England Region (01)	37	0	0	0	37
Mid Atlantic Region (02)	96	0	0	0	96
South Atlantic-Gulf Region (03)	152	0	0	0	152
Great Lakes Region (04)	48	0	0	0	48
Ohio Region (05)	107	0	0	0	107
Tennessee Region (06)	11	0	0	0	11
Upper Mississippi Region (07)	26	0	0	0	26
Lower Mississippi Region (08)	17	0	0	0	17
Souris-Red-Rainy Region (09)	10	0	0	8	2
Missouri Region (10)	38	0	16	15	7
Arkansas-White-Red Region (11)	38	0	22	0	16
Texas-Gulf Region (12)	22	0	6	8	8
Rio Grande Region (13)	9	8	1	0	0
Upper Colorado Region (14)	31	6	25	0	0
Lower Colorado Region (15)	4	4	0	0	0
Great Basin Region (16)	17	12	5	0	0
Pacific Northwest Region (17)	151	0	19	3	129
California Region (18)	69	6	39	18	6

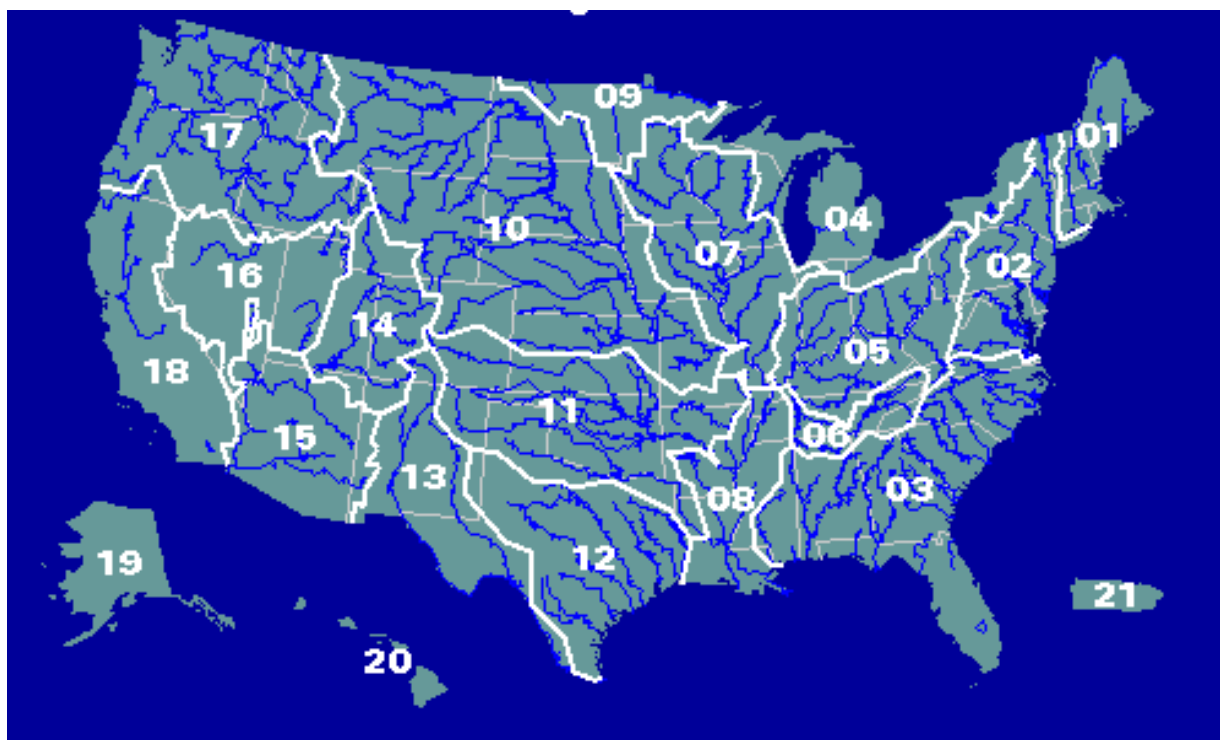


Figure A1. Map of the United States showing the HUC-02 regions directly obtained from the following U.S. Geological Survey web address: <https://water.usgs.gov/GIS/regions.html>.



Clear Data, Clearer Waters: An Interactive Dashboard for Visualizing Stream Turbidity Trends

Abiodun Ayo-Bali¹; Payton Davis²; Sergio Mandoza³; Galina Shinkareva⁴; Srilani Wickramasinghe⁵

* All authors contributed equally, and order is set alphabetically by surname

¹ University of Pittsburgh, Pittsburgh, Pennsylvania; aea69@pitt.edu

² Clemson University, Clemson, South Carolina; pbdavis@clemson.edu

³ Southern Oregon University, Ashland, Oregon; mendozas1@sou.edu

⁴ Illinois State University, Normal, Illinois; galina.shinkareva@gmail.com

⁵ Purdue University, West Lafayette, Indiana; swickra@purdue.edu

Corresponding Authors: Payton Davis, pbdavis@clemson.edu; Srilani Wickramasinghe swickra@purdue.edu

Advisors: Carlos E. Ramirez, Vidya Samadi, and Demir Ibrahim

Abstract

Turbidity is a crucial physical parameter for assessing water quality, yet open-source datasets containing turbidity data are scarce. Moreover, the process of identifying water stations with turbidity data, extracting the data, and formatting the data into a usable and interpretable form is often time-consuming and labor-intensive. This creates a significant obstacle, resulting in a time lag between the identification of stream quality issues and the implementation of protective measures for water quality. To address this issue, our team developed a turbidity dashboard using HydroLang that presents real-time turbidity data from United States Geological Survey (USGS) stream gauges using Portland Oregon as our point of interest. The dashboard retrieves data from the USGS instantaneous values data repository, to create a visual representation of the data, enabling faster detection of water quality issues and more timely interventions to safeguard water ways. Additionally, the dashboard uses data from USGS to estimate current temporal trends and develop a forecasting model using a simple univariate time series xgboost model which helps to identify and predict future turbidity issues. This paper describes the process of designing and implementing the data visualization dashboard, highlighting the potential of the dashboard to streamline water quality monitoring and facilitate access to real-time updates of turbidity data reported by the USGS. Initial user testing indicates that the dashboard significantly reduces the time required to access and visualize turbidity data. As future work one can expand upon the dashboards functions, including the addition of other water quality parameters.

Keywords: Turbidity, Water quality, Dashboard, HydroLang, Data visualization



1. Introduction

Water is considered one of the most important resources on earth, essential for all forms of life.^{1,2} However, as human population and urbanization continue to rise, so do water use and water pollution, placing increasing pressure on water quality. Water quality encompasses the physical, chemical, and biological conditions of water relative to its intended use or purpose.³ Various parameters are used to assess water quality, with turbidity being a key physical indicator.³

Turbidity measures the cloudiness of water caused by suspended particles.^{3,4,5} More specifically, turbidity is measured by using a turbidimeter, a device that passes light through water and detects the amount of light that is reflected back to the sensor.^{4,5} Turbidity is particularly important to aquatic life as high turbidity can lead to warmer water temperatures, decreased dissolved oxygen levels, and reduced food availability.³ Additionally, suspended particles can clog the gills of fish, further threatening aquatic ecosystems.³

Turbidity measurements can also provide valuable insights into other water quality parameters and help identify pollution sources.⁶ For example, Meyer et al., (2019) observed a strong correlation between turbidity levels and total phosphorus concentrations at the river Biles in Germany. When turbidity levels increased, the total phosphorus levels also increased.⁶ Meyer et al., (2019) concluded that the increase in turbidity and total phosphorus was caused by agricultural runoff from nearby farmlands. This is just one example that demonstrates how turbidity data, when analyzed in conjunction with other water quality parameters, can be a powerful tool for identifying non-point source pollution.

Furthermore, water quality regulatory organizations such as the U.S. Environmental Protection Agency (US EPA) identifies turbidity as an indicator for water safety and filtration effectiveness (see <https://www.epa.gov/ground-water-and-drinking-water/>). Water utility services must ensure that the water meets regulatory standards for turbidity prior to being distributed for industrial & domestic water use. Exceeding these thresholds raises water utility concerns and public health concerns, as higher turbidity levels could indicate contamination by disease-causing microorganisms (e.g. viruses, bacteria).⁷

Given the importance of turbidity as an indicator of water quality, it is crucial to have a reliable method for monitoring and analyzing turbidity data as data-driven decision-making is a critical aspect of water-quality management.⁸ This highlights the need for a turbidity dashboard that allows users to visualize turbidity patterns effectively. Traditionally, manually accessing comprehensive water quality records from multiple stations has been an effort-intensive and time-consuming process. However, the development of a turbidity dashboard enables stakeholders to quickly detect anomalies, uncover patterns, and respond to potential pollution incidents effectively. Modern software tools, such as turbidity dashboards, are essential for optimizing and expediting water quality monitoring processes while making data more accessible to the public.^{9, 10} These tools not only streamline the work of regulatory organizations but also enhance the transparency in water quality management.

Problem Statement: Only a limited number of United States Geological Survey (USGS) stations record stream turbidity data. Manually, filtering out these stations and accessing their daily data-



updates is laborious & time consuming. Our team identified the lack of an automated, easy to access system dedicated to stream turbidity, that can expedite the above process.

The objectives of this paper include (1) to develop and design a web-based dashboard that integrates instantaneous turbidity data from on-site automated recording equipment installed at multiple USGS sites using Hydrolang, and (2) to create a visualization component in the dashboard that allows users to visualize turbidity data, ensuring easy readability and an intuitive user interface.

2. Methodology

2.1 Study Area:

We focused our study within a 20 mile radius of Portland, Oregon (Fig. 1). Portland Oregon is influenced by the Mediterranean climate, experiencing warm and dry summers and cool and rainy winters. With climate change on the rise, Portland has experienced more intense weather extremes. For example, in 2022 Portland experienced record breaking high temperatures throughout the summer and fall months.¹¹ Additionally, in the same year, Portland experienced an intense and prolonged period of rainfall, increasing turbidity levels in local watersheds.¹¹



Figure 1: A map showing the study area Portland, Oregon with its USGS stream gauges marked by the gray pointers (source:<http://www.usgs.gov/>)

2.2 Data Retrieval:

We created a dashboard that displays current turbidity data from the USGS API, utilizing HydroLang for data retrieval and visualization. Figure 2 below displays the work flow of developing the dashboard.



HydroLang is an open source web-based toolkit for hydrological data retrieval, analysis and visualization.^{12, 13} HydroLang was used to connect to the USGS API and fetch instantaneous turbidity data. For turbidity measurements, USGS uses a data sampling method involving unfiltered water samples illuminated by monochrome near-infrared LED light with a wavelength range of 780-900 nm. The detection angle is set at 90 ± 2.5 degrees. Turbidity levels are recorded in Formazin Nephelometric Units (FNU, variable code 63680) and are commonly recorded at a fixed interval of 15- to 60-minutes and transmitted to the USGS every hour. The types of data retrieved included turbidity values from stations, date and time records, station name and station number. Data visualization with HydroLang helped create basic charts and graphs of time-series trends in turbidity and display them on an interactive dashboard. It includes several data visualization elements such as an interactive map displaying stream gauges recording turbidity values, an interactive graph showing the current trends in turbidity and data tables with basic statistics (median values, min & max). Python was used for statistical analysis and trend detection along with a simple 1D neural network model for forecasting. Both the Python analysis and the forecasting model were integrated into the dashboard. The pipeline (Fig. 2) and the turbidity dashboard (Fig. 3) will allow end users to access instantaneous turbidity data & statistics and analyze general turbidity trends.

2.3 Data cleaning and analysis:

Due to data gaps at each station when measurements were unavailable, we performed data pre-processing and cleaning. Large periods with missing data were excluded from the analysis. For smaller gaps, we used interpolation within the pandas package for Python, utilizing the 'time' parameter to fill NaN values. This method effectively interpolates data at daily and higher resolutions.

We performed a boxplot analysis for monthly and seasonal data, categorizing seasons as follows: December-January as winter, March-May as spring, June-August as summer, and September-November as fall. Additionally, we applied a time series decomposition using an additive model to analyze seasonal trends. The results are obtained by first estimating the trend by applying a convolution filter to the data. The trend is then removed from the series and the average of this detrended series for each period is the returned seasonal component.

All graphical materials for seasonal trends analysis and forecasting were obtained with Seaborn and Matplotlib packages for Python.

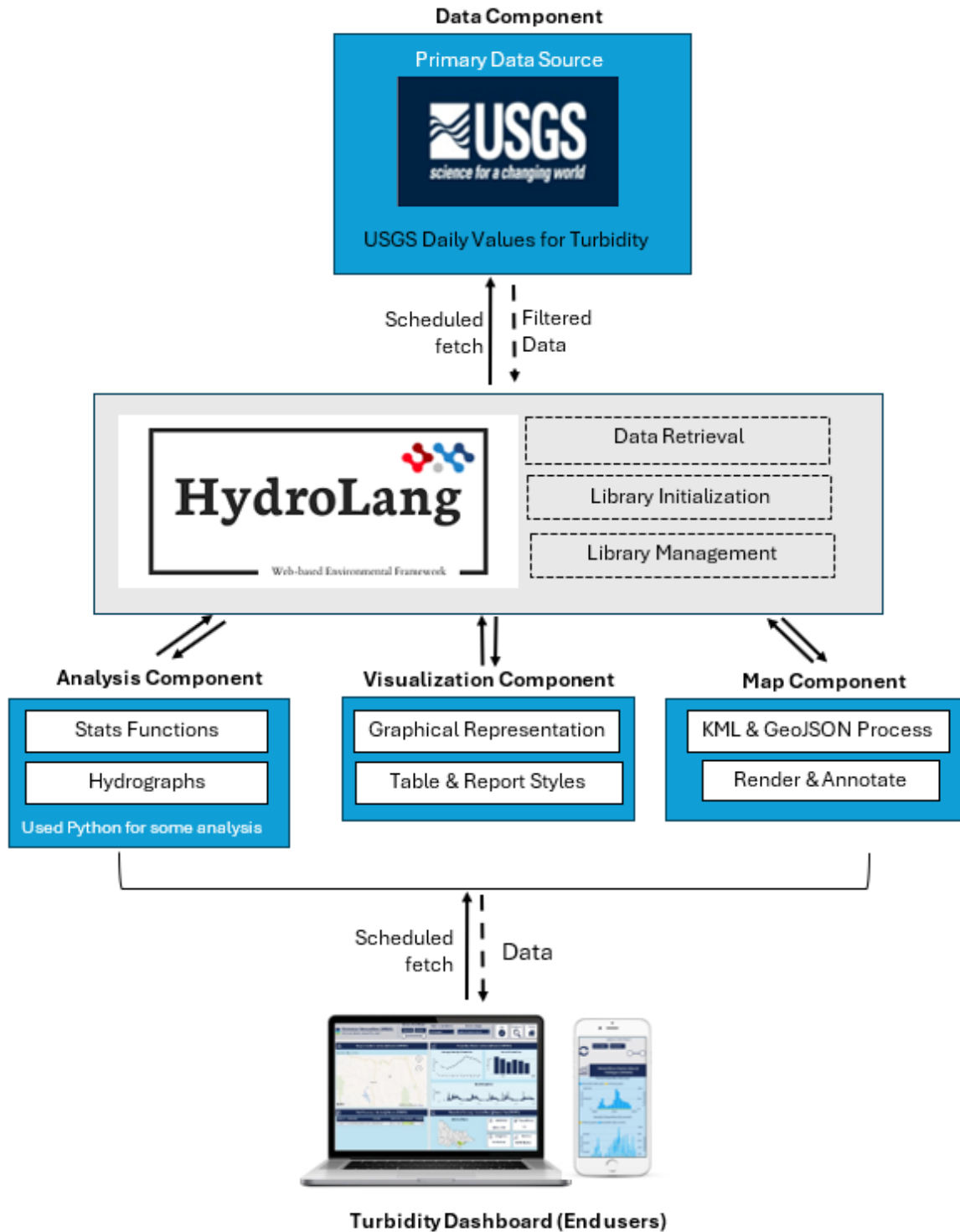


Figure 2. Diagram illustrating the dashboard design work flow from the data source (USGS) to the turbidity dashboard

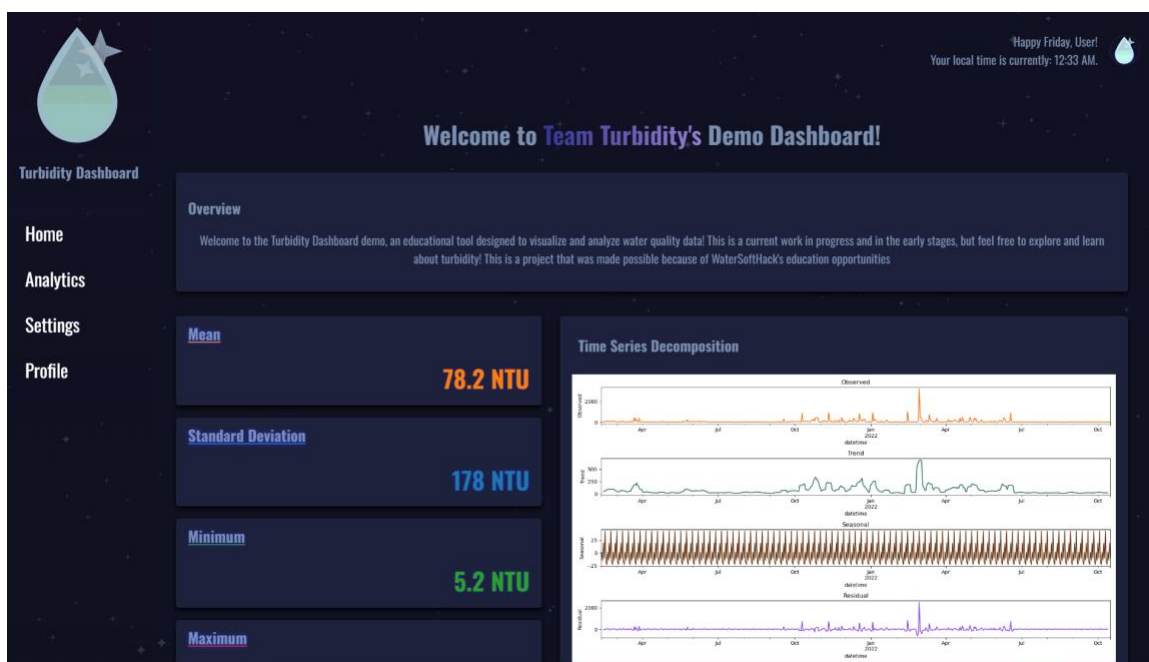


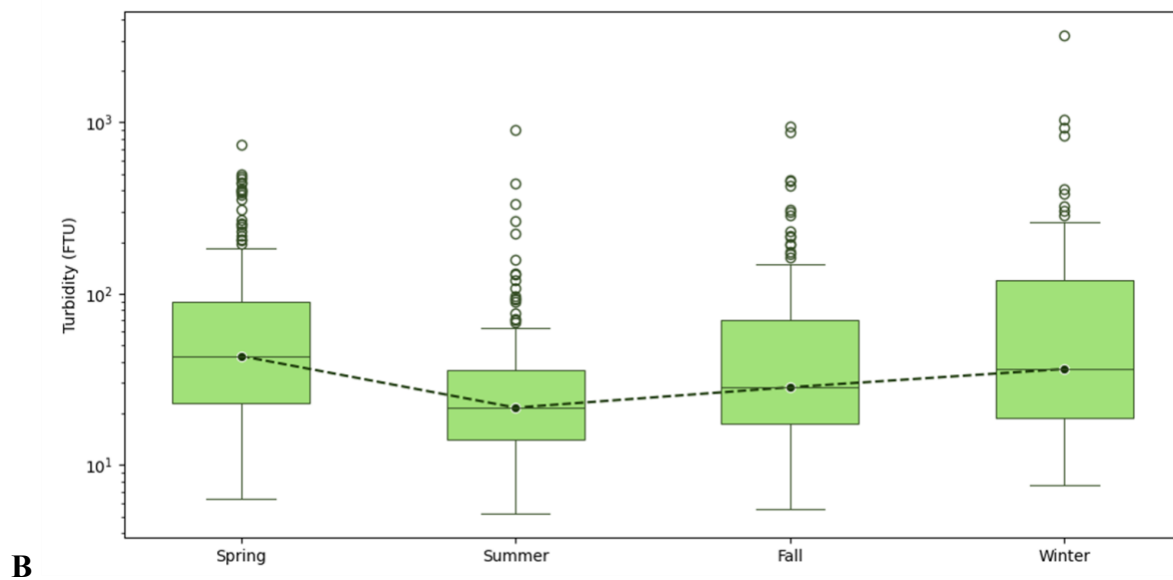
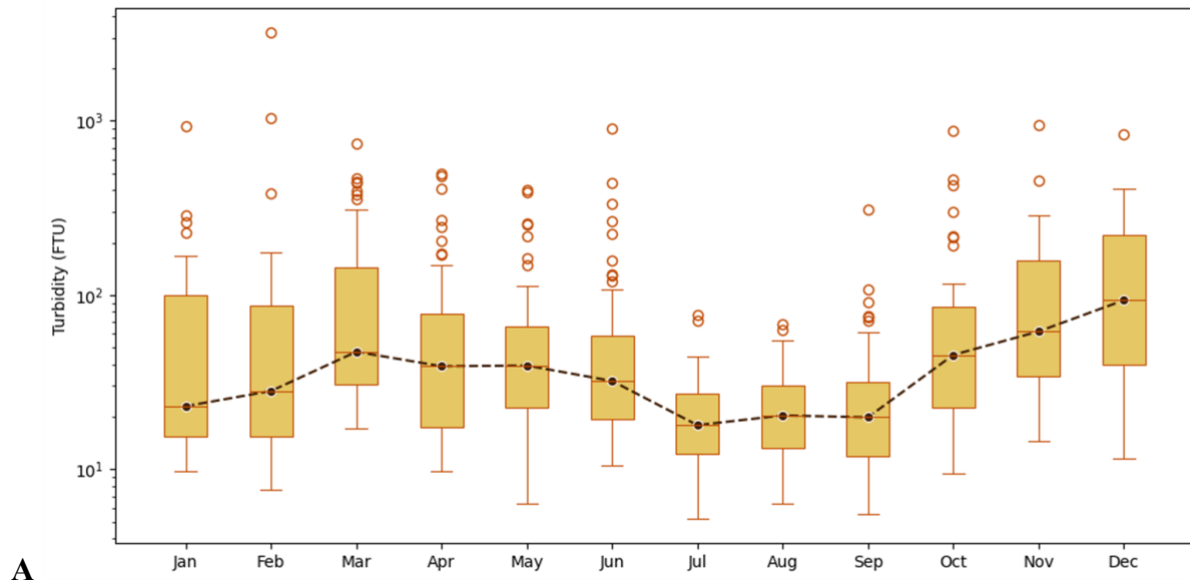
Figure 3. The dashboard developed for this project visualizes many different statistical analyses of turbidity, including a clear visual roadmap of next steps in its development.

3. Results and Discussion

3.1 Exploratory Data

The preliminary results from using the turbidity dashboard (Fig. 3) demonstrated its high efficiency in time series analysis and forecasting. There are visible trends in seasonal turbidity changes in the Portland area with lower turbidity values at the end of summer low water period until early fall (July-September), and higher values during winter months, such as December and January (Fig. 4), which usually receive the highest amount of precipitation. Time decomposition analysis also demonstrated turbidity peaks during winter and spring months. The seasonal trend indicates that turbidity values are generally higher in the middle of the month compared to the beginning and end. This pattern may be influenced by several factors, such as increased agricultural activities, runoff from periodic rainfall events, or human activities that peak mid-month.

Furthermore, the year 2022, for which USGS data was available for the area of interest, exhibited extreme weather conditions, including snowfall in spring (April) and a prolonged heat wave with 90-degree-plus days during the summer period.¹⁴ Portland had 1.6 inches of snow on April 11 2022, which was the first measurable snow in April at Portland International Airport since records started there in 1940. The same year Portland saw its third longest heat wave in recorded history with eight consecutive days of temperatures of 90 degrees or higher (from July 24 to July 31). Severe weather conditions with extreme high and low temperatures may have had a great impact in streamflow and turbidity levels at streams in the Portland area.



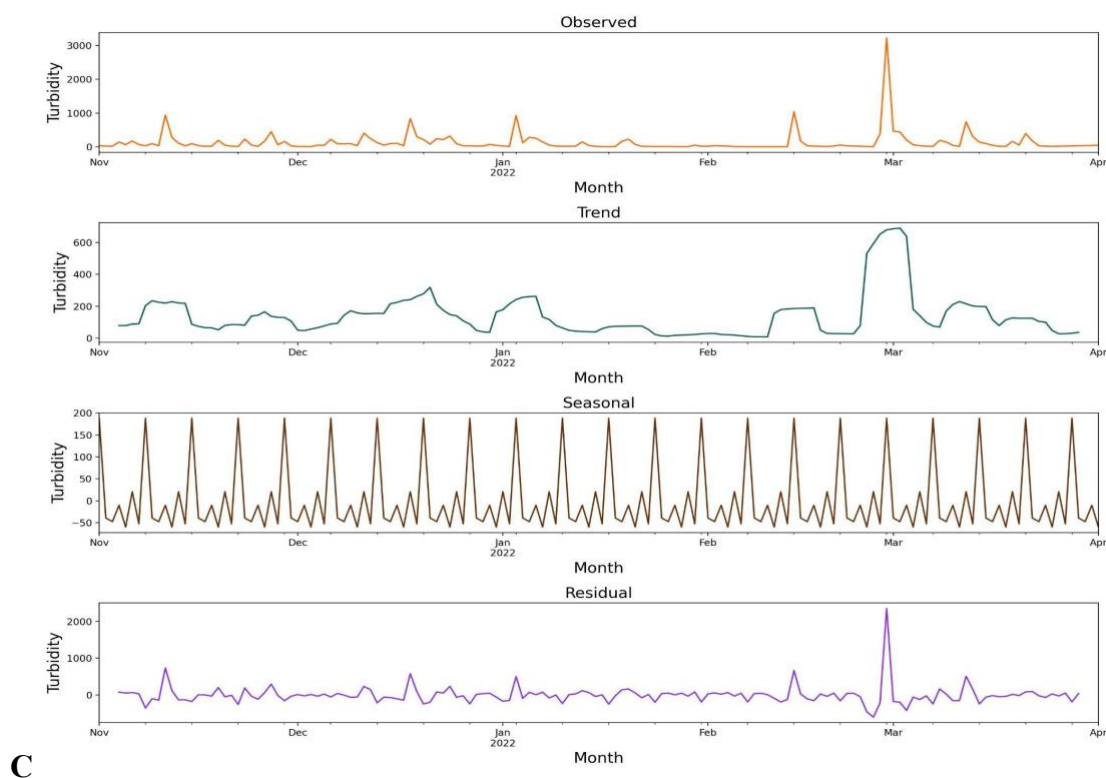


Figure 4. Seasonal turbidity trends at the example station of Johnson Creek at Pleasant Home, OR: (A) boxplot for monthly values; (B) boxplot seasonal values; (C) Time series decomposition plot.

3.2 XGBoost model

The USGS water quality datasets provide a strong foundation for model development, enabling accurate predictions.^{15, 16} However, missing data is a common challenge that can significantly impact model performance.^{17, 18} Bridging these gaps often requires costly data acquisition processes, involving data cleaning, integration, and potentially the collection of new information. Moreover, machine learning approaches have been instrumental in addressing difficulties in short-term load forecasting by integrating meteorological and climate data, utilizing big data, and employing hybrid models.¹⁹ In this study we adopted the extreme gradient boosting (XGBoost) for univariate time series model which was implemented through an optimized library created by Chen and Guestrin.²⁰ Figures 5 and 6 below display the results of using XGBoost.

The time series dataset is processed through two techniques for turbidity prediction. Data acquired from one station was split into training and testing datasets. Mar-2021 to Mar-2022 and from Jun-2022 to Nov-2022 for training and testing respectively. The error metrics of the model were evaluated using root mean squared error (RMSE), Mean Absolute Percentage Error (MAPE) and mean absolute error (MAE).

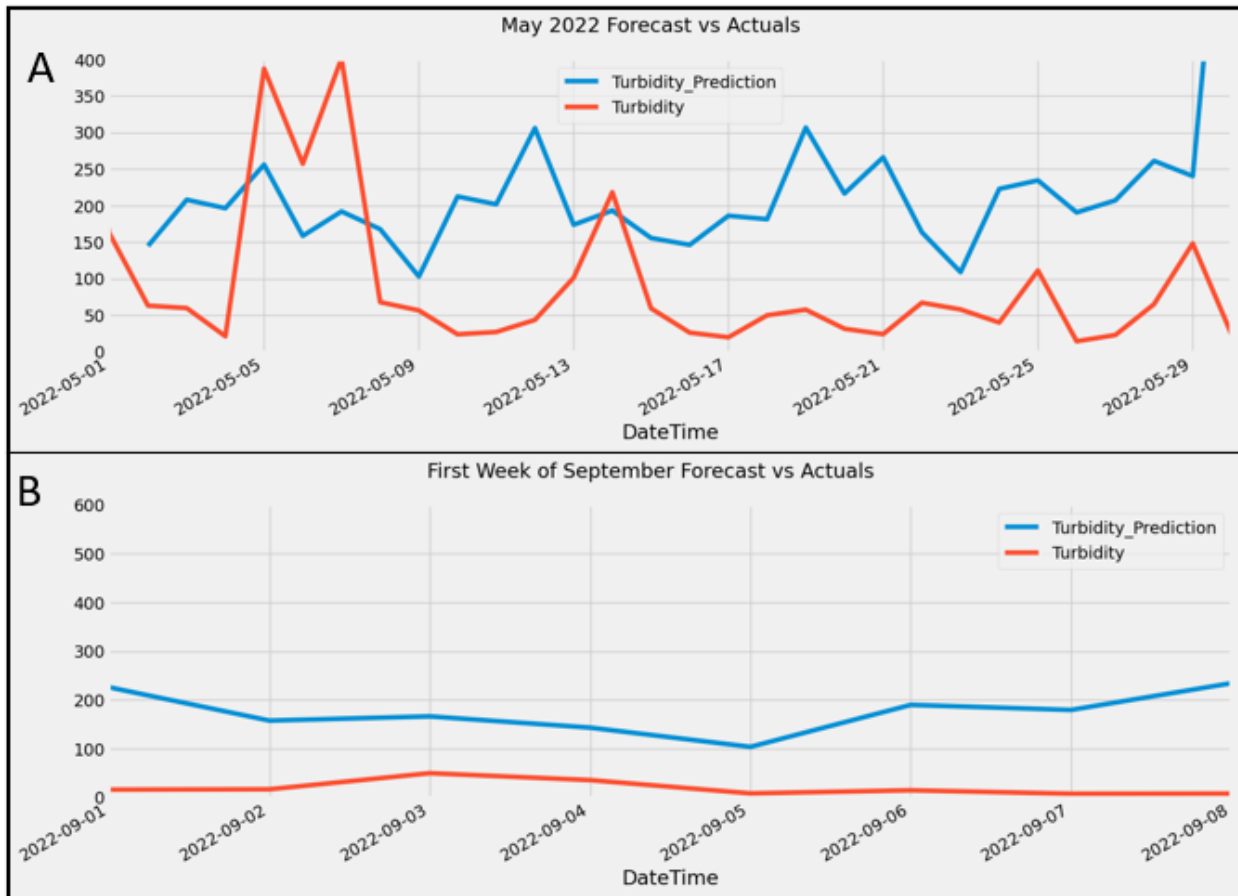


Figure 5. Univariate XGBoost turbidity forecasting for Badger Creek at Rugg Road Near Gresham, OR. (A) forecast vs. actual plot of turbidity of month May 2022. (B) forecast vs. actual plot of turbidity of first week of September May 2022.

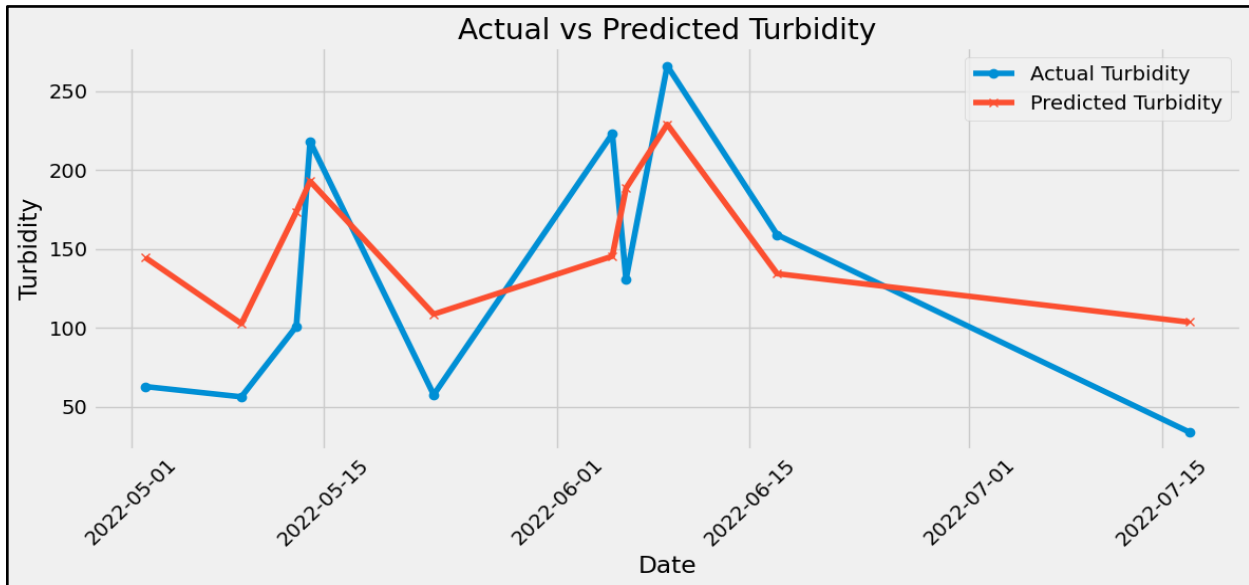


Figure 6. Best predicted days by the model.



Error metrics on the test dataset shows an RMSE of 51619 which suggests that, on average, the model's predictions deviate from the actual values by approximately 51619 units. This is quite large, indicating substantial errors in predictions. Measured MAE of 192.70 suggests that, on average, the model's predictions are off by 192.70 units from the actual values. While a MAPE of 970.78% is extremely high, suggesting that, on average, the predictions are off by 970.78% from the actual values.

Overall, high RMSE and extremely high MAPE indicate that the model's predictions are far from the actual values, both in absolute terms and relative to the actual values. In addition, the disparity between RMSE and MAE suggests the presence of some large errors (outliers) that are significantly affecting the RMSE value. To improve the accuracy of performing forecasting models with these datasets, we recommend the following:

1. Thorough examination of the retrieved data for outliers or anomalies that might be disproportionately affecting the error metrics.
2. Optimizing the model by exploring different algorithms, hyperparameter tuning, or incorporating more relevant features.
3. Turbidity, as an optical characteristic of water, can not be studied in isolation, as it can be influenced by both internal and external factors of the system.

3.3 Challenges

Throughout the completion of this project some challenges were encountered. The first challenge that arose was that the USGS turbidity data were sparse and not easily accessible. Additionally, among the limited number of stations that had turbidity data, we observed inconsistencies in how different stations recorded and stored this data. Both of these challenges complicated the data retrieval and data cleaning procedure, making it a time-consuming process. It is also important to note that turbidity data alone is difficult to interpret. Additional watershed and stream parameters are needed for a comprehensive analysis of water quality and ecosystem health.

4. Conclusions

Finding and interpreting data for research can be challenging. To address this, we developed a web dashboard that facilitates the retrieval and analysis of USGS turbidity data. Accessing turbidity data through a dashboard is arguably more palatable and convenient compared to the laborious and time-consuming manual data retrieval. While this tool provides a quick summary, seasonal trends and forecasts on instantaneous USGS data, interpreting turbidity data alone is difficult. To achieve a comprehensive analysis of water quality and ecosystem health, it is essential to include additional watershed and stream parameters. Thus, incorporating more variables into the dashboard would significantly enhance its utility. Future work will also require expanding the dashboard's functionality, including the integration of other water quality parameters.

References

1. Aulenbach, D.B., 1967. Water--Our Second Most Important Natural Resource. *BC Indus. & Com. L. Rev.*, 9, p.535.
2. Boyd, C.E., 2019. *Water quality: an introduction*. Springer Nature.



3. Omer, N.H., 2019. Water quality parameters. *Water quality-science, assessments and policy*, 18, pp.1-34.
4. Sampedro, Ó. and Salgueiro, J.R., 2015. Turbidimeter and RGB sensor for remote measurements in an aquatic medium. *Measurement*, 68, pp.128-134.
5. Kitchener BG, Wainwright J, Parsons AJ. A review of the principles of turbidity measurement. *Progress in Physical Geography*. 2017 Oct;41(5):620-42.
6. Meyer, A.M., Klein, C., Fünfroeken, E., Kautenburger, R. and Beck, H.P., 2019. Real-time monitoring of water quality to identify pollution pathways in small and middle scale rivers. *Science of the Total Environment*, 651, pp.2323-2333.
7. De Roos, A.J., Gurian, P.L., Robinson, L.F., Rai, A., Zakeri, I. and Kondo, M.C., 2017. Review of epidemiological studies of drinking-water turbidity in relation to acute gastrointestinal illness. *Environmental health perspectives*, 125(8), p.086003.
8. Aram, S., Rivero, M. H., Pahuja, N. K., Sadeghian, R., Paulino, J. L., Meyer, M., & Shallenberger, J., 2020. Multi-Environmental Parameters Dashboard for Susquehanna River Basin using Machine Learning techniques. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* IEEE. p. 697-700.
9. Zainuddin, Z.Q.M., Yahya, F., Gubin Mounq, E., Mohd Fazli, B., Abdullah, M.F., 2023. Effective dashboards for urban water security monitoring and evaluation. *Int. J. Electr. Comput. Eng. IJECE* 13, 4291. <https://doi.org/10.11591/ijece.v13i4.pp4291-4305>
10. Monschein, C., Layman, L., 2024. SoutheastCon 2024 Designing, Implementing, and Evaluating a Municipal Water Quality Dashboard, in: SoutheastCon 2024. Presented at the SoutheastCon 2024, pp. 113–118. <https://doi.org/10.1109/SoutheastCon52093.2024.10500297>
11. Portland Water Bureau (2023) 2023 Drinking Water Quality Report. Available at: <https://www.portland.gov/water/water-quality/2023-drinking-water-quality-report> (Accessed: 1 August 2024).
12. Erazo Ramirez, C., Sermet, Y., Molkenthin, F., Demir, I., 2022. HydroLang: An open-source web-based programming framework for hydrological sciences. *Environ. Model. Softw.* 157, 105525. <https://doi.org/10.1016/j.envsoft.2022.105525>
13. Erazo Ramirez, C., Sermet, Y., Demir, I., 2023. HydroLang Markup Language: Community-driven web components for hydrological analyses. *J. Hydroinformatics* 25, 1171–1187. <https://doi.org/10.2166/hydro.2023.149>
14. Oregon’s extreme weather in 2022 included April snow and a record hot October. The Oregonian. Advance Local Media LLC. Web portal. Date accessed: 8/1/024. URL: <https://www.oregonlive.com/weather/2022/12/oregons-extreme-weather-in-2022-included-april-snow-and-a-record-hot-october.html>
15. Hirsch, R.M., Archfield, S.A. and De Cicco, L.A., 2015. A bootstrap method for estimating uncertainty of water quality trends. *Environmental Modelling & Software*, 73, pp.148-166. <https://doi.org/10.1016/j.envsoft.2015.07.017>
16. Srebotnjak, T., Carr, G., de Sherbinin, A., & Rickwood, C. (2012). A global Water Quality Index and hot-deck imputation of missing data. *Ecological Indicators*, 17, 108–119. <https://doi.org/10.1016/j.ecolind.2011.04.023>
17. Read, E. K., Carr, L., De Cicco, L., Dugan, H. A., Hanson, P. C., Hart, J. A., Kreft, J., Read, J. S., & Winslow, L. A. (2017). Water quality data for national-scale aquatic



- research: The Water Quality Portal. *Water Resources Research*, 53(2), 1735–1745.
<https://doi.org/10.1002/2016WR019993>
18. Sprague, L. A., Oelsner, G. P., & Argue, D. M. (2017). Challenges with secondary use of multi-source water-quality data in the United States. *Water Research*, 110, 252–261.
<https://doi.org/10.1016/j.watres.2016.12.024>
19. Reddy, B.K., Ayyagari, K.S., Medam, R.R. and Alhaider, M., 2023. Application of Machine Learning Techniques in Modern Hybrid Power Systems—A Case Study. In *IoT, Machine Learning and Blockchain Technologies for Renewable Energy and Modern Hybrid Power Systems* (pp. 173-204). River Publishers.
<https://doi.org/10.1109/CONIT59222.2023.10205638>
20. Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>

Appendix

<https://www.epa.gov/ground-water-and-drinking-water/>



zoom.us Meeting View Edit Window Help Recording A closed captioning transcript is being generated. Leave meeting OK Mon Jul 22 11:09 AM

WaterSoftHack Cybertraining (2024 - 2026)

WaterSoftHack Team: Vidya Samadi, Bijaya Adhikari, Matthew Boyer, Anthony Castronova, Ibrahim Demir, Krishna Panthi, & Carlos E. Ramirez

July 22, 2024

Zoom meeting interface showing a grid of participants on the right side of the screen.

zoom.us Meeting View Edit Window Help Recording A closed captioning transcript is being generated. Leave meeting OK Mon Jul 22 11:10 AM

INTRODUCTION

Our Team

<p>Vidya Samadi Assistant Professor, Clemson University Principal Investigator</p>	<p>Matthew Boyer Research Associate Professor, Clemson University Co-Principal Investigator</p>	<p>Demir Ibrahim Associate Professor, The University of Iowa Co-Principal Investigator</p>	<p>Carlos E. Ramirez Graduate Student, The University of Iowa Research Assistant</p>
<p>Bijaya Adhikari Assistant Professor, The University of Iowa Co-Principal Investigator</p>	<p>Anthony Castronova Senior Research Hydrologist at CUAHSI Co-Principal Investigator</p>	<p>Krishna Panthi Graduate Student, Clemson University Graduate Research Assistant</p>	

Zoom meeting interface showing a grid of participants on the right side of the screen.



zoom.us Meeting View Edit Window Help Recording A closed captioning transcript is being generated. Leave meeting OK Mon Jul 22 11:20 AM


WaterSoftHack ACTIVITIES July 22-August 02, 2024

Week 1


Day	11-13:00 EST	13-14:00 EST	14-15:00 EST
Week 1			
Monday (July 22)	Project Introduction / Project Milestone	Lunch	FAIR - Liberians(UT, Clemson/IOWA)
Tuesday (July 23)	Training on Data Retrieval (API/Web Service)	Lunch	Hands-on Workshop
Wednesday (July 24)	Training on Data Analytics (Analysis Tools)	Lunch	Hands-on Workshop
Thursday (July 25)	Training on Data Analytics (Visualization Tools)	Lunch	Hands-on Workshop
Friday (July 26)	Training on Scientific Communication (Alda Center - 2hr training)	Lunch	CUAHSI(DEI and Science)

zoom.us Meeting View Edit Window Help Recording A closed captioning transcript is being generated. Leave meeting OK Mon Jul 22 11:25 AM


Machine Learning Training



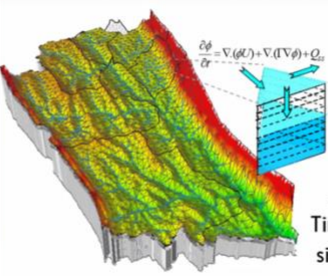
Vidya Samadi
Clemson University



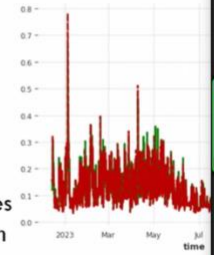
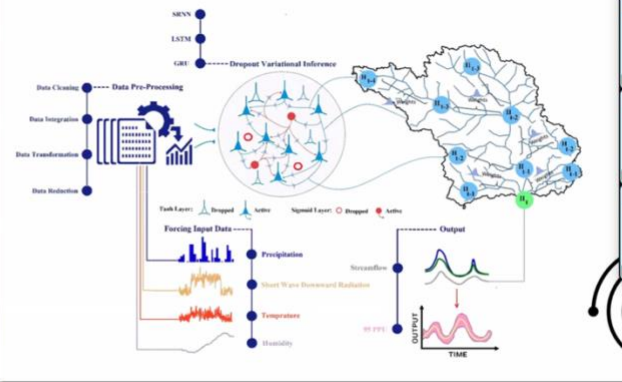
Tony Castronova
CUAHSI



Krishna Panthi
Clemson University



Time series simulation

Krishna Panthi



The image shows two screenshots from a Zoom meeting. The top screenshot displays a slide titled "EXPECTATIONS" with the following text:

- Be Inclusive:** Treat your peers with respect regardless of race, gender, sexual orientation, abilities/disabilities, beliefs, technical skills, or position.
- Listen and be Considerate:** Master the skill of listening and deepen your relationships with your peers.
- Critique Respectfully and Constructively:** Engage in respectful discourse with your peers whose perspectives differ from your own, recognizing that differences of opinion do not signify rightness or wrongness.

The slide also features an NSF logo in the bottom left corner. The Zoom interface shows a list of participants on the right, including Krishna Panthi, Vidya Samadi (Clemson University), Ibrahim Demir, Biyya Adhikari, Anthony Castronova (Ishim), and Erizo Ramirez, Carlos.

The bottom screenshot shows a presentation slide for "HydroSuite" with a yellow "IOWA" logo in the top right and a yellow button labeled "WaterSoftHack '24" at the bottom. The Zoom interface shows a recording notification and a participant list on the right including Krishna Panthi, Ibrahim Demir, Vidya Samadi (Clemson University), and Carlos Erizo Ramirez.