1. **Problem Statement**

Flooding results in loss of life and causes billions of dollars in damages annually in the United States. Flooding is expected to increase in temperate areas from increased precipitation due to human induced climate change and land use changes (Alfieri, et al. 2017). This variability in precipitation can have many effects on the management of streams and water infrastructure that were designed based on streamflow data from gauge stations. This data however, can have limitations based on the length of observation. One potential solution to this is using tree rings as a proxy for reconstructing streamflow. This can allow for extending the historical streamflow record to allow for better planning, flood prediction, and the evaluation of modern climate events in a long-term context. Advances in machine learning can also aid paleo streamflow reconstruction, allowing for extended streamflow records with potentially better skill than the traditional regression-based reconstruction methods.

Traditional reconstruction methods have been used for decades. Formetta et al. 2021 focused on the Adige River Basin (ARB), a vital water source for northern Italy, but with highly variable streamflow. Comprehending the historical variability of streamflow in the Adige River Basin is crucial for effective water resource management and planning in light of climate change. Traditional streamflow reconstruction methods utilize tree-ring proxies, which depend on the link between tree-ring growth and precipitation. These methods, however, are constrained by the availability of tree-ring data and the precision of the proxies. A novel approach employing the Old World Drought Atlas (OWDA) scPDSI as a reconstruction proxy has been developed and effectively applied in other European watersheds.

Traditional streamflow reconstructions using tree-ring based proxies such as scPDSI utilize stepwise linear regression or log-linear based models (Formetta et al, 2021). The implementation of machine learning for paleo streamflow reconstructions has not been largely explored, however, successful ML reconstructions of other climate variables have been completed in recent years. Paleo reconstructions of precipitation in the Sava River Basin were recently published using AI/ML/DL techniques which showed skill statistics higher than that of traditional reconstruction using SLR or log-linear modeling (Ramirez Molina et al, 2023).

This preliminary literature investigation helps build context of the previous studies on streamflow reconstructions using dendroclimatic proxies. This NRT research project aims to extend historical streamflow records with tree-ring proxies in the spatial extent of the Southeast United States using both traditional regression methods and novel AI/ML/DL techniques.

The use of tree-ring proxy-based streamflow reconstruction in the southeast US has the potential to provide valuable insights into past hydrological and climate conditions, which can inform water resource management decisions and help mitigate the impacts of climate change. However, there are several gaps, concerns, and opportunities that need to be addressed in order to fully realize the potential of this research.

One of the key gaps is the limited availability of streamflow records in the southeast US. This makes it difficult to assess long-term trends and variability in water resources and to develop accurate and reliable streamflow reconstructions (Patskoski et al. 2015). Another gap is the need for further development and evaluation of AI/ML approaches for streamflow reconstruction. While these methods have the potential to overcome some of the limitations of traditional regression techniques, they may also be more sensitive to the quality of the data used to train them. It is important to develop robust AI/ML models that can accurately capture complex relationships in streamflow data and that can be used to reconstruct streamflow records for a variety of conditions. Additionally, there's a lack of literature review or integration of more advanced machine learning techniques beyond regression models to enhance the accuracy or predictive capabilities of streamflow reconstructions.

In addition to these gaps, there are also several concerns that need to be addressed. One concern is that traditional regression techniques may not be able to capture complex relationships in streamflow data (Ridgeway et al. 1999), which could lead to inaccuracies in streamflow reconstructions. Another concern is that machine learning techniques may be overfitted if the models are not properly trained (Gharib & Davies 2021). It is important to develop methods for evaluating the performance of different streamflow reconstruction methods under a variety of conditions and to identify the most effective methods for different applications. Moreover, the development and application of AI/ML models require large amounts of high-quality data. Ensuring the availability and quality of tree-ring proxy data is crucial for the success of AI/ML-based streamflow reconstruction projects. Despite these gaps and concerns, there are also several opportunities for the use of tree-ring proxy-based streamflow reconstruction in the southeast US.

The goal of conducting this research is to extend historical streamflow records with tree-ring proxies, provide past information on water availability and variability, and assist water managers on making better water resources decisions in the southeast United States. The research objective is to utilize and compare AI/ML/DL models for streamflow reconstruction to traditional methods within the Tennessee Valley watershed using tree-ring based proxies. Because AI/ML/DL reconstruction techniques have not been employed on streamflow, our research is original in exploring these emerging techniques on a climate variable critical for making water resource management decisions.

1. **Research Plan Outline**
   1. **Data**

11 unimpaired sites (gauges) within the Tennessee River Valley were selected for reconstruction, with an observational record of 1920 to 2008. Streamflow data was gathered for each gauge from USGS gauge data, yielding monthly mean streamflow in both cubic feet per second (flow) and million cubic meters (volume).

The North American Drought Atlas (NADA) provides Self-calibrating Palmer Drought Severity Index grid points within a 450-kilometer radius of the gauge in question. The scPDSI grid points provide annual values between -4 and 4 indicating the severity of drought and pluvial periods for the area, with a robust temporal extent based upon tree-ring chronologies. By utilizing the scPDSI grid points as a proxy for dendrochronology based streamflow reconstructions, we can extend historical streamflow records as far back as 0 AD.

* 1. **Traditional Methods**

Pre-screening is performed such that poorly correlated (less than 99%) scPDSI cells are eliminated. Additional pre-screening will be done to ensure that correlation between scPDSI and observed streamflow is stable for the overlapping period of record. The PDSI cells passing pre-screening will be used as predictors (independent variables) while observed streamflow will be used as the predictand (dependent variable) in applying forward and backward Stepwise Linear Regression (SLR). Quantile mapping bias correction will be applied to each SLR reconstructed streamflow vector.

* 1. **AI/ML Methods**

Novel AI/ML streamflow reconstruction methods will be employed using the no-code platform, RapidMiner Studio. For each gauge, the instrumental record and 450-km radius scPDSI values will be utilized for training and cross-validation. The ML algorithms used to reconstruct streamflow at each of the 11 gauges will include Generalized Linear Model (GLM), Random Forest (RF), Deep Learning (DL), and others as the research develops.

* 1. **Evaluation / Performance Metrics**

The output of the AI/ML models on tree ring proxies will be compared to the output of the traditional SLR models on tree ring proxies to see if there is any improvement. The output of these models will then be compared to the observed data from the unimpaired USGS gauges for accuracy as well. We anticipate using performance metrics of squared correlation and RMSE against the observed record to quantify differences in performances between the reconstruction methods.

1. **Anticipated Broader Impacts**

Streamflow reconstructions can be used to inform water resource management decisions by providing insights into past hydrological and climate conditions, as well as potential future trends. Stakeholders involved in water resources and infrastructure management, such as agencies like the Tennessee Valley Authority (TVA), will benefit from this research. Insights derived from the paleo streamflow reconstructions aid in optimizing resource utilization and resilience planning against climate-induced variability. resource utilization and resilience planning against climate-induced variability.

Reconstructing streamflow from tree ring data using machine learning will allow for TVA to have longer and more complete records. This would better inform their decisions when it comes to making improvements and repairs to infrastructure that manages water resources and flood control. This research could also aid the research and scientific community in understanding the changes in the water levels over time in the Tennessee Valley area. Leading to more complete research in ecological and hydrological impacts on resource management, energy production and ecological preservation.

With the threat of floods and droughts potentially increasing due to climate change, having a better understanding of historical streamflow in watersheds is more important than ever. Especially with people most at risk and with less means to mitigate their own risk being statistically overrepresented in flood plains, generating more complete streamflow records will allow for better planning to potentially avoid greater disasters (Tate et al., 2021).

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., Feyen, L., 2017. Global projections of river flood risk in a warmer world. Earth’s Future 5, 171–182. <https://doi.org/10.1002/2016EF000485>

Anderson, S., Ogle, R., Tootle, G., & Oubeidillah, A.(2019). Tree-ring reconstructions of streamflow for the Tennessee Valley. *Hydrology*,*6*(2), 34.

Formetta, G., Tootle, G., & Bertoldi, G. (2021). Streamflow Reconstructions Using Tree-Ring Based Paleo Proxies for the Upper Adige River Basin (Italy). *Hydrology*,*9*(1), 8.

Gharib, A., & Davies, E. G. (2021). A workflow to addresspitfalls and challenges in applying machine learning models to hydrology. *Advances in Water Resources*, *152*, 103920.

Patskoski, J., Sankarasubramanian, A., & Wang, H. (2015). Reconstructed streamflow using SST and tree-ring chronologies over the southeastern United States. *Journal of Hydrology*, *527*, 761-775.

Ramírez Molina, A.A.; Bezak, N.; Tootle, G.; Wang, C.; Gong, J. Machine-Learning-Based Precipitation Reconstructions: A Study on Slovenia’s Sava River Basin. *Hydrology* 202**3**, *10*, 207.<https://doi.org/10.3390/hydrology10110207>.

Ridgeway, G., Madigan, D., & Richardson, T. S. (1999, January). Boosting methodology for regression

problems. In *Seventh International Workshop on Artificial Intelligence and Statistics*. PMLR.

Tate, E., Rahman, M.A., Emrich, C.T., Sampson, C.C., 2021. Flood exposure and social vulnerability in the United States. Nat Hazards 106, 435–457.<https://doi.org/10.1007/s11069-020-04470-2>.