Title: Advancing the Use of Machine Learning in the Development of Useable Tree Ring Proxy-Based Reconstruction of Streamflow

Subtitle: A Study of the Tennessee River Valley

Author(s) 1st Affiliation: Madi Brake1, Mahsa Mirzakhani2, Spencer Livingston2, Glenn Tootle1

Department: Civil engineering1, Geography, and the Environment2,

NSF NRT Water Research to Operations

Organization: The University of Alabama

City, Country: Tuscaloosa, Alabama, United States

Author(s) 2nd Affiliation: Andres Ramirez

Department: Department of Computer Science

Organization: The University of Alabama

City, Country: Tuscaloosa, Alabama, United States

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Abstract:

Water resources and planning management decisions are generally based on observed historic records. Therefore, it is crucial to investigate historic streamflow for water managers to make better water resources planning decisions. Tree-ring-based-proxies have been used to skillfully reconstruct streamflows with traditional regression techniques such as stepwise linear regression (SLR). Machine learning (ML) methods have the potential to improve streamflow reconstruction, but have not been widely applied as a skillful method. The Tennessee River Valley is a vital water source in the Southeast United States for varying applications including energy production (29 hydroelectric dams) and flood planning. Given the recent (and projected) variability in precipitation impacting watersheds across the Southeast US, the optimal allocation of streamflow is an ongoing challenge in the Tennessee Valley. USGS streamflow data for the Tennessee River Valley was collected for 11 sites (gauges) where average annual volumes were calculated for each gauge’s period of record, ~1920 to 2005 AD, and broken into seasonal volumes. Extending this period of record and increasing information about past (paleo) drought and pluvial (wet) periods would assist in quantifying risk to water managers and planners. This report evaluates novel AI/ML methods against traditional regression-based methods of streamflow reconstructions. The highest performing method for this preliminary investigation for reconstructing summer streamflow (June-September) in the Tennessee Valley is the random forest (RF) machine learning algorithm, which yielded significantly better performance than traditional, regression-based reconstructions of streamflow. The findings of this research exemplify machine learning as a viable and robust method of reconstructing streamflow using tree-ring based proxies, that can be utilized by water resources managers such as the Tennessee Valley Authority (TVA) to aid in informed decision-making.

1. Introduction: (MADI & MAHSA)
   1. Intro / Study Area / Background (MADI)

Water resources management decisions utilize observed records of climate variables such as temperature, precipitation, soil moisture, and streamflow. Numerous sectors from flood control to energy production are reliant upon historical streamflow data indicative of variation of water availability over time to make informed management decisions. Streamflow variability records on timescales of decades to centuries become increasingly critical for water management, as resource shortages are imposed by increasing demand and limited supply (Meko et al., 2010). Streamflow reconstructions using paleoclimatic proxies are among the most important tools for studying past climates, one of the most widely used being dendroclimatic proxies, or proxies established using tree-ring chronologies and the environmental conditions they reflect. Such information from tree rings has been used to reliably reconstruct a number of climate variables including streamflow on centennial to millennial scales (Cook et al., 2015). Dendroclimatic reconstructions provide information at higher temporal resolution than most other proxies, allowing modern climatic events to be placed into a longer-term context (Robeson et al., 2020).

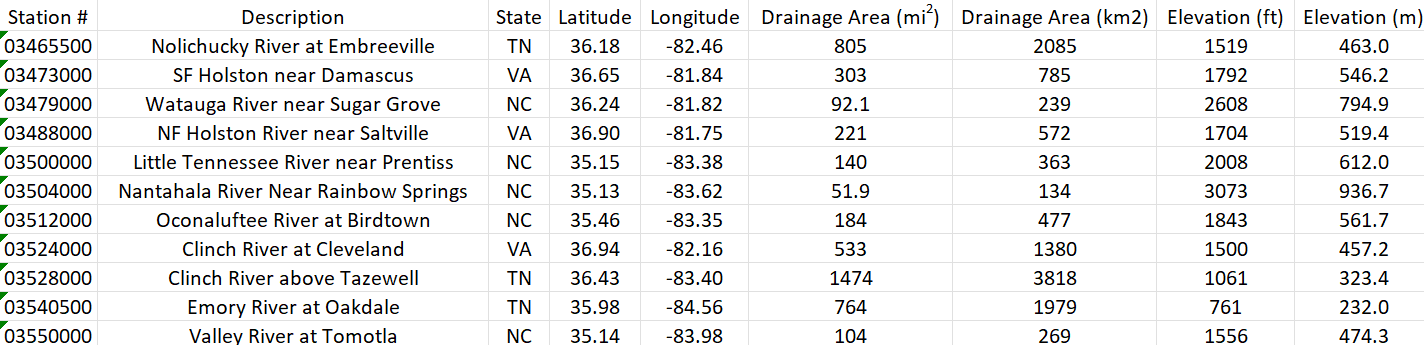
One of the key limitations of assessing long-term streamflow and water variability trends in the southeast United States, particularly in the Tennessee Valley, is the lack of unimposed stream gauges. Dendroclimatological research in the Southeastern U.S. has been notably limited compared to other regions over the past two decades, attributed to factors such as dam construction by the Tennessee Valley Authority (TVA) reducing the number of unimpaired stream gauges, and the region's complex topography and dry winters hindering tree-ring chronologies correlated with streamflow (Anderson et al., 2019). Despite lingering doubts about the applicability of dendroclimatology in the southeast United States, researchers have utilized tree-ring data to demonstrate that tree-ring patterns effectively predicted May–June precipitation in East Tennessee (Blasing et al., 1981) and regression-based reconstructions of summer streamflow using dendrochronological data are statistically significant (Anderson et al., 2019).

* 1. Goals and Objectives: (MAHSA)

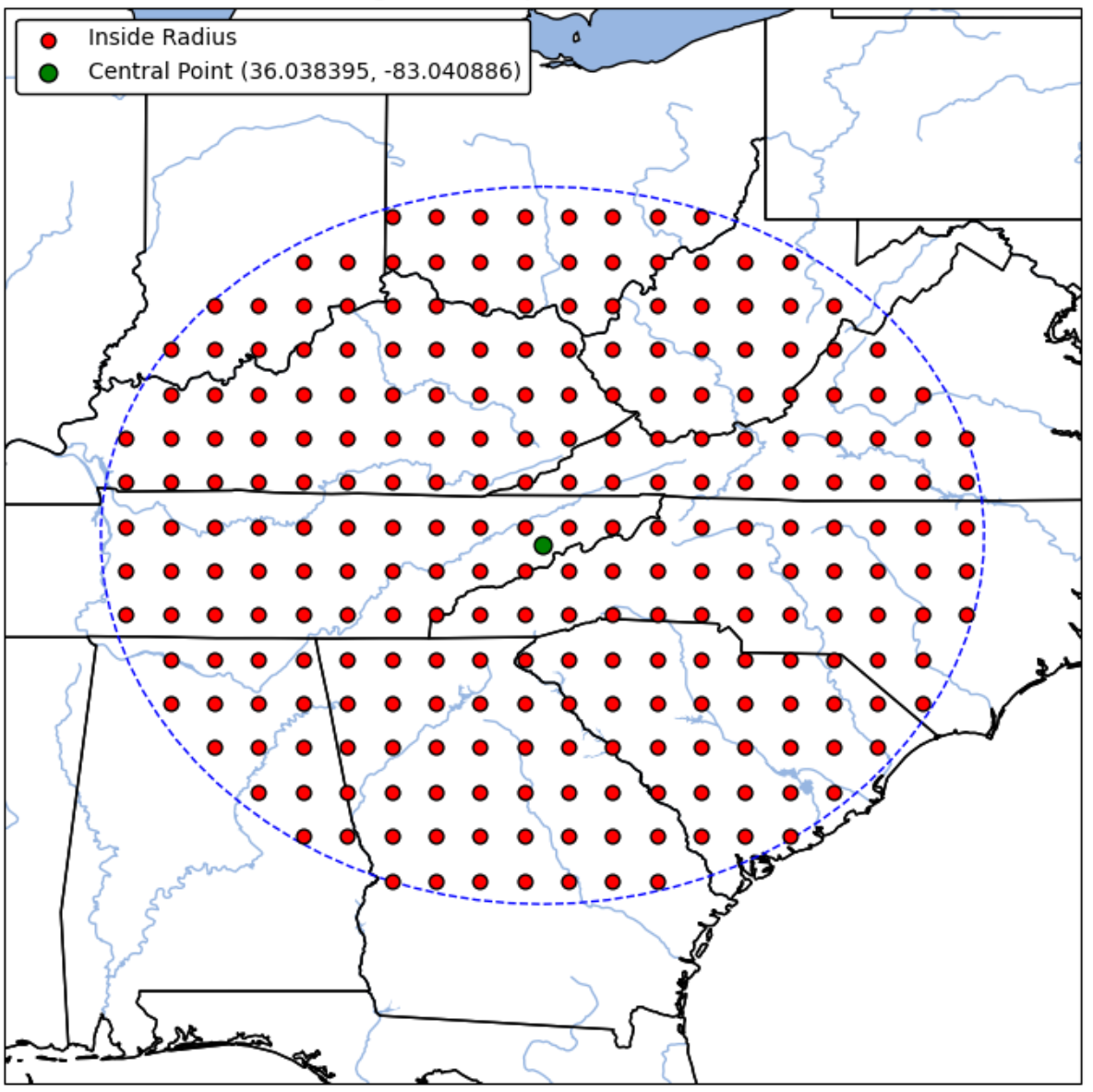
Advanced AI/ML techniques haven't been fully explored for streamflow reconstruction. Additionally, there's a lack of literature review regarding the use of machine learning techniques beyond regression models to enhance the accuracy or predictive capabilities of streamflow reconstructions. It is important 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. Ramírez Molina et al. 2023 investigate the use of AI and machine learning to enhance seasonal precipitation reconstructions in the Sava River Basin, finding that these methods outperform traditional regression techniques and provide valuable data for water management. The goal of this study is to explore how advanced artificial intelligence and machine learning techniques can be used to reconstruct streamflow data more accurately than traditional, regression-based methods. We compare these AI/ML approaches to traditional methods in the Tennessee Valley watershed using tree-ring proxies. This comparison helps us identify the most reliable models for capturing the complex dynamics of streamflow data, ultimately enabling dependable reconstructions across various conditions.

1. Materials and Methods: (MADI & SPENCER)
   1. Data: (MADI)
      1. USGS Streamflow Data (MADI)

Streamflow data for 11 gauges within the Tennessee Valley were obtained from the United States Geological Survey (USGS). The accuracy and length of streamflow gauge records are critical components in streamflow reconstruction studies. While the USGS has been collecting streamflow data since 1887, not all gauge stations have complete records, often due to technical issues (Anderson et al., 2019). However, for this particular study, only USGS gauge stations with complete and acceptable records were used. These stations provided monthly cumulative flow data in million cubic meters (MCM), with three seasonalities evaluated using traditional regression-based reconstruction techniques: annual (Jan-Dec), summer (June-September), and winter (Dec-March). Each station had unique characteristics such as elevation and drainage area, which are important factors influencing streamflow.



* + 1. Gridded Self-Calibrating Palmer Drought Severity Index (MADI)

Precipitation-based tree-ring chronologies are used to quantify the Palmer Drought Severity Index (PDSI) which dates back thousands of years (Ho et al., 2016). The Old-World Drought Atlas (OWDA) provides self-calibrating Palmer Drought Severity Index (scPDSI) values for grid points across Europe from 0 to 2012 AD (Formetta et al., 2021). Similarly, the North American Drought Atlas (NADA) provides Self-calibrating Palmer Drought Severity Index (scPDSI) grid points, with annual values between -4 and 4 indicating the severity of drought and pluvial periods (Anderson et al., 2019). The gridded scPDSI data provides a robust temporal extent, extending as far back as 365 AD, based upon moisture-sensitive tree-ring chronologies for the Tennessee Valley spatial region. For both the traditional, regression-based models and the novel machine learning models used to reconstruct streamflow in the Tennessee Valley, annualized, gridded scPDSI data within a 450-kilometer radius of the gauges was utilized. 

* 1. Traditional Methods : (SPENCER)
     1. Pre-Screening

Two different methods for pre-screening were used for the data to ensure it was useful for doing reconstructions. The first method was correlation using the Pairwise Pearson Correlation test, where observed discharge was compared to the values from the scPDSI cells. Cells that were correlated less than 99% with the overlapping record of observed streamflow were eliminated. The next pre-screening method was the stability test. This was done by using an 11-year moving correlation window between the overlapping period of observed streamflow and scPDSI cells. From the temporal stability analysis, if there is any negative correlation between a PDSI cell and observed streamflow, the cell is deemed unstable and removed from the model (Tootle et al., 2023)

* + 1. Modeling

With scPDSI cells that were not removed during the pre-screening process, forwards and backwards (standard) stepwise linear regression (SLR) is performed on the observed streamflow and the corresponding scPDSI cells, and performs the Durbin-Watson test to test for autocorrelation, further filtering the usable scPDSI cells in the reconstruction of streamflow. The regression equation created from this is then used for modeling streamflow in unobserved years where the regressor data is available (Tootle et al., 2023). Statistical measures of performance, such as coefficients of determination (R²) and root mean square error (RMSE) values are computed for each reconstruction.

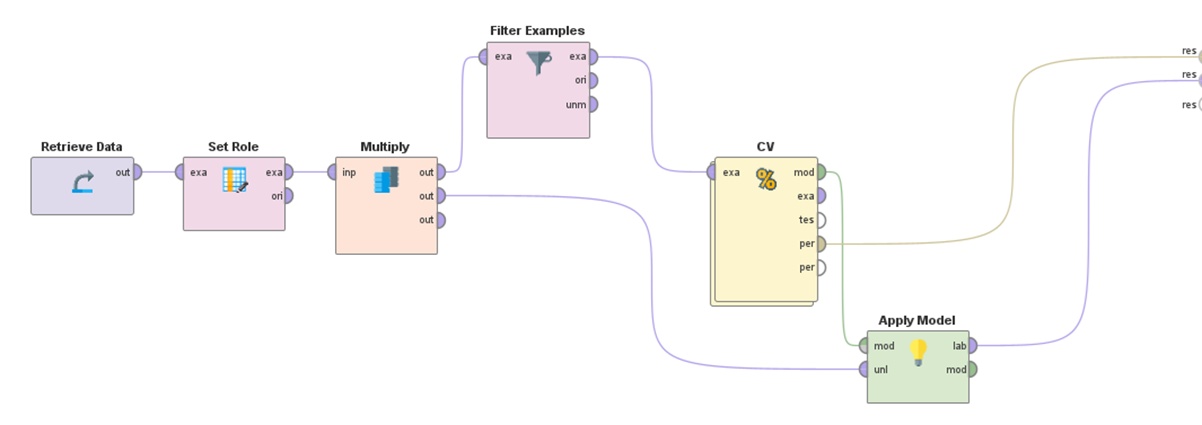
* + 1. Bias Correction (MADI)

Quantile mapping is the bias correction method used to align the modeled cumulative distribution function (CDF) with the observed values, involving comparing the empirical quantiles of observed and modeled data and adjusting the modeled (reconstructed) values accordingly (Robeson et al., 2020). RQUANT, a specific implementation of quantile mapping in the R *qmap* package, is utilized for the bias correction of the regression-based (SLR) streamflow reconstructions in the Tennessee Valley. While it produces a close match between observed and modeled CDFs, it reduces the risk of overfitting by not preserving every individual quantile difference (Robeson et al., 2020).

* 1. Machine Learning Methods: (SPENCER)

Three different machine learning algorithms were utilized to reconstruct streamflow on the highest performing (R²) gauges, in relation to season, including Random Forest (RF), Generalized Linear Model (GLM), and Deep Learning (DL). Random Forest is an ensemble method that builds multiple decision trees to improve prediction accuracy. It reduces overfitting by averaging the results of trees trained on random subsets of the data, used for both classification and regression tasks. Generalized Linear Models extend linear regression to accommodate non-normal response distributions. They link a linear predictor to the response variable through a specific link function, suitable for various data types including binary and count data. Deep Learning involves neural networks with multiple layers that learn from vast amounts of data. It's effective in pattern recognition across unstructured data like images and text (Ramirez et al., 2023).

Using the platform RapidMiner Studio, processes were developed to train and test the learning algorithms on all data, including observed streamflow and all scPDSI data within a 450-kilometer radius, to reconstruct streamflow. The performance of these three ML models, RF, GLM, and DL, were evaluated using 40-, 5-, and 10-fold cross-validations, respectively, using R² and RMSE as the metrics (Ramirez et al., 2023).



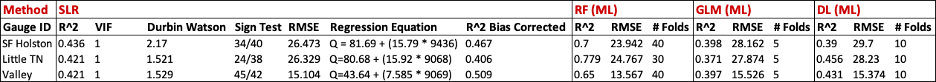
1. Results (Mahsa):
   1. Regression Models

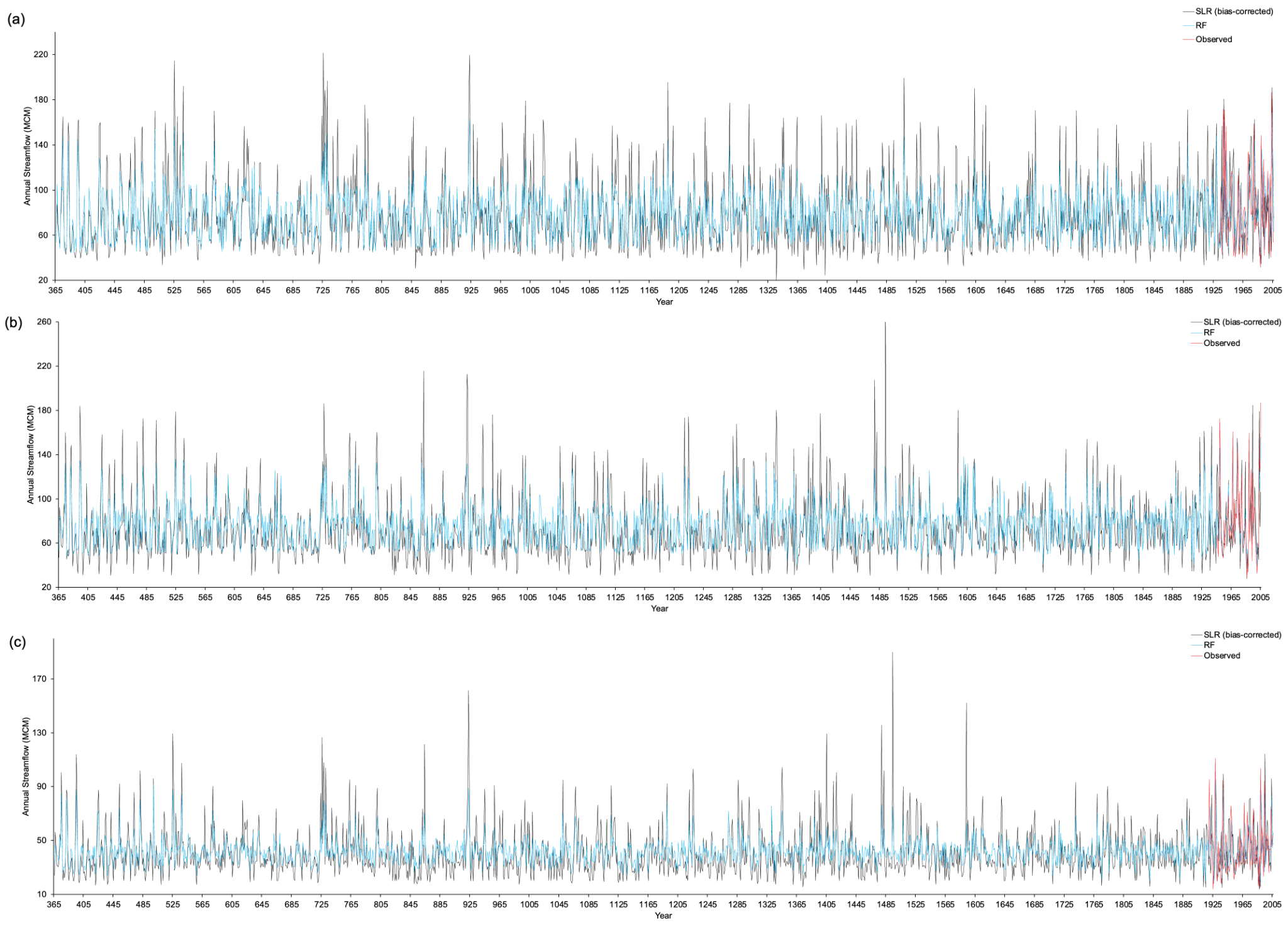
The analysis of regression models, particularly Stepwise Linear Regression (SLR), has yielded significant insights into the historical reconstruction of streamflow. For the SF Holston gauge, SLR exhibited a coefficient of determination (R²) of 0.467 after bias correction, indicating a moderate fit to the observed data. Similarly, the Little TN and Valley gauges showed R² values of 0.421. Despite the decent explanatory power of the SLR models, the RMSE values of 26.473 for SF Holston, 26.329 for Little TN, and 15.104 for Valley suggest that the SLR method's predictions deviate from the observed values to some extent. These findings highlight the limitations of traditional regression techniques in capturing the complex behaviors inherent in streamflow data over extended periods.

* 1. Machine Learning Models

Three different machine learning techniques—Random Forest (RF), Generalized Linear Model (GLM), and Deep Learning (DL)—were evaluated against traditional Stepwise Linear Regression (SLR) for their ability to reconstruct streamflow using tree-ring proxies. The effectiveness of each model was evaluated based on their ability to reconstruct streamflow patterns for eleven gauges within the Tennessee River Valley, with a particular focus on the summer season (JJAS), winter season, and on an annual basis. However, during the analysis, it became apparent that three gauges—SF Holston, Little TN, and Valley—exhibited superior performance for the summer season reconstructions. Consequently, to refine our study, we concentrated exclusively on these three gauges and excluded the remaining eight from further consideration. For the SF Holston gauge, RF demonstrated an improved coefficient of determination (R²) of 0.7 after bias correction, a notable enhancement from the SLR method's R² of 0.436. The GLM method showed a less robust performance with an R² of 0.398, while DL displayed an R² of 0.39. However, RF yielded the lowest RMSE of 23.942, indicating a stronger predictive accuracy among the models tested. Similar improvements were observed for the Little TN gauge, where the RF method's R² increased to 0.779, substantially surpassing the SLR's R² of 0.421. GLM and DL models also showed improvements but were outperformed by RF, with R² values of 0.371 and 0.456, respectively. The Valley gauge exhibited consistent results, with the RF method improving to an R² of 0.65, and the GLM and DL methods showing R² values of 0.397 and 0.431, respectively, after bias correction. The RF method's predictive accuracy, as indicated by the lowest RMSE of 13.567, confirmed its superiority (Table 1 and Figure 2).

Table 1: Comparative Performance of Machine Learning Techniques Versus Stepwise Linear Regression for Streamflow Reconstruction (JJAS) Skill Statistics (365-2005 AD) in the Tennessee River Valley



Figure 2: JJAS streamflow reconstructions *(a)* SF Holston Gauge, (b) Little TN Gauge, and (c) Valley Gauge – SLR vs RF from 365-2005 AD in the Tennessee River Valley.

1. Discussion and Broader Impacts: (ALL)
   1. Analysis of Results / Conclusions (MADI / SPENCER)

The analysis compared the performance of traditional Stepwise Linear Regression (SLR) with three machine learning techniques—Random Forest (RF), Generalized Linear Model (GLM), and Deep Learning (DL)—in reconstructing streamflow patterns in the Tennessee Valley using the tree-ring based proxy of gridded scPDSI data. While SLR provided moderate fits with coefficients of determination (R²) ranging from 0.437 to 0.421 across the 11 different gauges after bias correction, its predictive accuracy was limited, evidenced by relatively high root mean square error (RMSE) values. These R² and RSME values in the SLR methods were inline with the findings of Anderson in this region.

The study focused on refining the analysis by concentrating exclusively on three gauges—SF Holston, Little TN, and Valley for machine learning method comparisons. The machine learning models, particularly RF, showed significant enhancements in performance and yielded notably higher R² values ranging from 0.7 to 0.779 for the SF Holston and Little TN gauges for the summer (JJAS) seasonality. Despite some improvements with GLM and DL, RF consistently outperformed them across all three gauges, highlighting the efficacy of machine learning techniques over traditional regression methods in capturing the complexities of streamflow data.

These findings underscore the limitations of traditional regression techniques in capturing the intricate behaviors inherent in streamflow data over extended periods, while highlighting the effectiveness of machine learning methods, particularly Random Forest, in improving predictive accuracy and providing valuable insights into historical streamflow reconstruction.

* 1. Broader Impacts (MAHSA)

Reconstructing streamflow history is crucial for understanding past hydrological patterns and climate conditions. These reconstructions provide a powerful tool for water managers by extending their datasets beyond the limitations of recorded data. For example, stakeholders like the Tennessee Valley Authority can use these reconstructions to gain a more robust and comprehensive understanding of streamflow trends within the Tennessee Valley. The detailed picture painted by these extended records is instrumental in optimizing water resource management, ensuring that the region's water allocation is both efficient and sustainable. Additionally, by reflecting on past hydrological events, these reconstructions enable a proactive approach to resilience planning, preparing the region for varying climatic scenarios that could impact water availability and distribution. Moreover, extending the temporal span of hydrological records through these reconstructions facilitates the education of the public and policymakers about the historical context of current water challenges. Such knowledge is indispensable for fostering public awareness and shaping policy that is forward-looking, resilient, and sustainable. It also provides a solid foundation for academic research, offering datasets that can be used to validate hydrological models, test hypotheses about past climate change impacts, and train the next generation of hydrologists and climate scientists. On a socio-economic level, these reconstructions offer invaluable insights for agricultural planning, urban development, and industrial water use. Farmers, urban planners, and industrial stakeholders stand to benefit from improved seasonal forecasts and risk assessments, which can inform irrigation practices, urban water supply planning, and industrial water usage. This can lead to more strategic crop scheduling, reduced risk of water shortages for cities, and optimized industrial operations that depend on water availability. Furthermore, this research has significant implications for infrastructure management and natural disaster preparedness. Detailed streamflow data empowers managers to make informed decisions regarding the inspection and upkeep of critical water infrastructure, ensuring its integrity and functionality. When faced with potential floods and droughts, the ability to draw upon a more extensive historical record allows for the evaluation of control measures under diverse conditions, increasing the effectiveness of response strategies. Moreover, by placing contemporary climate events within a historical framework, it becomes possible to discern patterns and assess the impacts of climate change with greater clarity. Simultaneously, the ongoing exploration and refinement of machine learning methodologies in hydrologic forecasting signal a transformative step forward, promising to elevate the precision and predictive capabilities of future hydrological models.

References:

Anderson, S.; Ogle, R.; Tootle, G.; Oubeidillah, A. Tree-Ring Reconstructions of Streamflow for the Tennessee Valley. Hydrology 2019, 6, 34. <https://doi.org/10.3390/hydrology6020034>

Blasing, T.J.; Duvick, D.N.; West, D.C. Dendroclimatic calibration and verification using regionally averaged and single station precipitation data. Tree-Ring Bull. 1981, 41, 37–43.

Edward R. Cook et al., Old World megadroughts and pluvials during the Common Era.Sci. Adv.1,e1500561(2015).DOI:10.1126/sciadv.1500561

D.M. Meko, C.A. Woodhouse, K. Morino, Dendrochronology and links to streamflow, Journal of Hydrology, Volumes 412–413, 2012, Pages 200-209, ISSN 0022- 1694, <https://doi.org/10.1016/j.jhydrol.2010.11.041>

Formetta, G.; Tootle, G.; Bertoldi, G. Streamflow Reconstructions Using Tree-Ring Based Paleo Proxies for the Upper Adige River Basin (Italy). Hydrology 2022, 9, 8. <https://doi.org/10.3390/hydrology9010008>

Ho, M., Lall, U., and Cook, E. R. (2016), Can a paleodrought record be used to reconstruct streamflow?: A case study for the Missouri River Basin, Water Resour. Res., 52, 5195–5212, doi:10.1002/2015WR018444.

Ramírez Molina, A. A., Bezak, N., Tootle, G., Wang, C., & Gong, J. (2023). Machine-Learning-Based Precipitation Reconstructions: A Study on Slovenia’s Sava River Basin. Hydrology, 10(11), 207.

Robeson, S. M., Maxwell, J. T., & Ficklin, D. L. (2020). Bias correction of paleoclimatic reconstructions: A new look at 1,200+ years of Upper Colorado River flow. Geophysical Research Letters, 47, e2019GL086689. <https://doi.org/10.1029/2019GL086689>