



University of
Massachusetts
Amherst

DEVELOPMENT OF A DECISION SUPPORT SYSTEM WEBTOOL FOR HISTORIC AND FUTURE LOW FLOW ESTIMATION IN THE NORTHEAST UNITED STATES WITH APPLICATIONS OF MACHINE LEARNING FOR ADVANCING PHYSICAL AND STATISTICAL METHODOLOGIES

Item Type	Dissertation (Open Access)
Authors	DelSanto, Andrew F
DOI	10.7275/36385583
Download date	2026-03-15 14:55:13
Link to Item	https://hdl.handle.net/20.500.14394/19438

**DEVELOPMENT OF A DECISION SUPPORT SYSTEM WEBTOOL FOR
HISTORIC AND FUTURE LOW FLOW ESTIMATION IN THE NORTHEAST
UNITED STATES WITH APPLICATIONS OF MACHINE LEARNING FOR
ADVANCING PHYSICAL AND STATISTICAL METHODOLOGIES**

A Dissertation Presented

By

ANDREW FRANCIS DELSANTO

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2024

Civil and Environmental Engineering Department

© Copyright by Andrew Francis DelSanto 2024

All Rights Reserved

**DEVELOPMENT OF A DECISION SUPPORT SYSTEM WEBTOOL FOR
HISTORIC AND FUTURE LOW FLOW ESTIMATION IN THE NORTHEAST
UNITED STATES WITH APPLICATIONS OF MACHINE LEARNING FOR
ADVANCING PHYSICAL AND STATISTICAL METHODOLOGIES**

A Dissertation Presented

By

ANDREW FRANCIS DELSANTO

Approved as to style and content by:

Richard N. Palmer, Chair

Konstantinos Andreadis, Member

Timothy Randhir, Member

John E. Tobiason, Department Head
Civil and Environmental Engineering Department

DEDICATION

In honor of my parents. Thank you for always believing in me, even when I did not.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Richard N. Palmer, for his many years of thoughtful guidance and support. I truly could not have done it without you. I also want to thank the other members of my committee, Dr. Konstantinos Andreadis and Dr. Timothy Randhir, for their continued advice and encouragement.

I am grateful to the current and past members of the UMass EWRE/Civil Engineering program that have generously contributed their time and knowledge to helping me succeed, including but not limited to Don Park, Nelson Da Luz, Craig Brinkerhoff, Aidan Provost, and Chinedum Eluwa.

I am thankful to the Northeast Climate Adaptation Science Center for funding my research and always providing a group of intelligent and kind-hearted colleagues to collaborate with.

Special thanks to Parker Norton and Jacob LaFontaine of the USGS for their assistance with extracting and running the National Hydrologic Model in the Precipitation Runoff Modeling System. Your assistance and guidance have helped me immensely.

Finally, thank you to my friends and family for the continued love and support throughout graduate school. For friends, I would like to thank Kiserian Jackson, Grant Monast, Philip Marchetti, Matthew Etman, Brett Riley, Greer Storey, Jack Grzechowiak, Luana Pego, Karina Hernandez, and Lauren Beckett. For family, I want to thank my dad Bill, my mom Cheryl, my stepdad Gene, my grandmothers Jane and Patricia, my sister Sarah, and my cat Wren. You all mean a great deal to me and that will never change.

ABSTRACT

DEVELOPMENT OF A DECISION SUPPORT SYSTEM WEBTOOL FOR HISTORIC AND FUTURE LOW FLOW ESTIMATION IN THE NORTHEAST UNITED STATES WITH APPLICATIONS OF MACHINE LEARNING FOR ADVANCING PHYSICAL AND STATISTICAL METHODOLOGIES

FEBRUARY 2024

ANDREW FRANCIS DELSANTO

B.S., ROGER WILLIAMS UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Emeritus Professor Richard N. Palmer

Droughts are a global challenge and anthropogenic climate change is expected to increase the frequency and severity of extreme low flow events. A major challenge for resource managers is how best to incorporate future climate change projections into low flow event estimations, especially in ungaged basins. Using both physically based hydrology models and statistical models, this dissertation contributes novel methodologies to three key challenges associated with 7-day, 10-year low flow (7Q10) estimation in the northeast United States.

Chapter 2 builds upon statistically based 7Q10 estimation in ungaged basins by comparing multiple machine learning algorithms to classical statistical methodologies. This chapter's objective is to identify a robust statistical methodology applicable for the entire northeast U.S. that includes statistically significant climate variables that allow for the incorporation of climate change. Results suggest that the random forest method can

provide regional 7Q10 estimates with similar errors to current, state-by-state 7Q10 estimates.

Chapter 3 tests the applicability of a novel machine learning algorithm, Fuzzy C-Means clustering, to calibrate rainfall-runoff models in ungaged basins for both daily streamflow and 7Q10 estimation. Future updates to national rainfall-runoff models, which can directly incorporate climate change projections into calculations, will allow these models to be created in ungaged basins, but they will require extensive calibration and/or verification. Results suggest that this methodology significantly improves daily streamflow estimation but fails to improve 7Q10 estimation.

Chapter 4 summarizes the development of a stakeholder-driven decision support system (DSS) web-application for calculating the 7Q10 at gages and estimating the 7Q10 in ungaged basins with projected climate changes. By incorporating the statistical model from Chapter 2 into the DSS and comparing the results to the physical modeling from Chapter 3, the DSS can estimate the impact of future temperature and precipitation changes on 7Q10s.

This work highlights advancements in physical and statistical modeling techniques for 7Q10 estimation in ungaged basins and assists resource managers in addressing a growing need for incorporating anticipated climate change into drought calculations.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
ABSTRACT	vi
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER 1	1
INTRODUCTION	1
CHAPTER 2	10
Low-Flow (7-Day, 10-Year) Classical Statistical and Improved Machine Learning Estimation Methodologies	10
2.1 Data and Study Area	11
2.2 Input Variables/Data	12
2.3 7Q10 Comparison Data	13
2.4 Materials and Methods	15
2.4.1 7Q10 Values at Each Site	15
2.4.2 Statistical Methods	16
2.4.3 Input Variables	21
2.4.4 Leave-One-Out Cross-Validation (LOOCV)	22
2.4.5 Error Metrics	23
2.5 Results and Discussion	25
2.5.1 Multiple Linear Regression	25
2.5.2 Logarithmic-Transformed Linear Regression	28
2.5.3 Random Forest	31
2.5.4 Neural Network	34
2.5.5 Generalized Additive Model	36
2.5.6 Comparisons to StreamStats Estimates	38
2.5.7 General Performance	39
2.6 Conclusions	43
CHAPTER 3	45
Fuzzy C-Means Clustering for Physical Model Calibration and 7-Day, 10-Year Low Flow Estimation in Ungaged Basins: Comparisons to Traditional, Statistical Estimates	45
3.1 Background	46

3.2 Data and Study Area	49
3.2.1 Study Area	49
3.2.2 Data	50
3.3 Methodology	53
3.3.1 Calibration of the NHM-PRMS Models	54
3.3.2 Fuzzy-C Means Clustering	59
3.3.3 Silhouette Analysis	61
3.3.4 Evaluation Metrics for Daily Streamflow and 7Q10 Estimates	62
3.4 Results and Discussion	63
3.4.1 Uncalibrated Model Performance vs Calibrated Model Performance	64
3.4.2 Uncalibrated vs. Calibrated Models for Gaged 7Q10 Estimation	65
3.4.3 Results from Silhouette Analysis with Fuzzy C-Means Clustering	67
3.4.4 Results from Clustering for Daily Streamflow Modeling	70
3.4.5 Procedure for Optimal Cluster Selection for Overall Model Performance	72
3.5 Conclusions	76
CHAPTER 4	77
A Stakeholder-Driven Decision Support System Web Tool for Calculating Historic and Future Projected 7-Day, 10-Year Low Flows in the Northeast United States	77
4.1 Introduction	78
4.2 Problem Definition and Knowledge Acquisition	82
4.2.1 Define Objectives	82
4.2.2 Identify Decision-Making Processes and Gather Requirements	83
4.2.3 Choose Technology	93
4.3 Data Processing and Algorithm Selection	93
4.3.1 Design Data Model	94
4.3.2 Acquire Data, Integrate with Existing Systems, and Apply Algorithms	95
4.4 User Interface Design	96
4.4.1 Evaluate Options, Integrate Visualization Tools, & Intuitive Design	96
4.4.2 Provide Documentation	97
4.5 Continuous Testing, Evaluation, and Feedback	98
4.5.1 In-House Alpha Testing, External Beta Testing, and Acquiring Feedback	98

4.5.2 Continuous Revisions and Re-Deployment	99
4.6 Continued Maintenance and Updates	102
4.7 Summary	103
CHAPTER 5	105
CONCLUSIONS	105
APPENDIX A	109
APPENDIX B	114
APPENDIX C	115
APPENDIX D	125
APPENDIX E	127
BIBLIOGRAPHY	130

LIST OF TABLES

Table 1: StreamStats 7Q10 estimation by state.....	14
Table 2: Input variables for 7Q10 estimation	21
Table 3: Multiple linear regression input variables and significance.	26
Table 4: Logarithmic-transformed linear regression input variables and significance.	29
Table 5: Random forest input variables and significance.	31
Table 6: Generalized additive model input variables and significance.....	36
Table 7: Multiple method comparisons to current estimates.	39
Table 8: Comparisons between statistical methods using leave-one-out cross-validation.	40
Table 9: RMSE comparisons between methods for specified size ranges.....	41
Table 10: Parameters calibrated in the physical hydrology model	54
Table 11: Predictor variables used in the Fuzzy C-Means analysis from the 94 test basins	61
Table 12: Error metrics for 7Q10 estimation (only StreamStats locations).....	66
Table 13: Error metrics for 7Q10 estimation (only StreamStats locations).....	66
Table 14: Highest 10 average silhouette coefficients.....	68
Table 15: Error metrics for 7Q10 estimation (only StreamStats locations).....	72
Table 16: Error metrics for 7Q10 estimation (only StreamStats locations).....	73
Table 17: Questions for Initial Interviews	88
Table 18: Module Steps for Calculating 7Q10 at Gages.....	99
Table 19: Module Steps for Ungaged Basin 7Q10 Estimation	101

LIST OF FIGURES

Figure 1: The 165 gages and their corresponding watersheds used for this analysis. Gages are displayed as blue triangles, the standard convention by the USGS for designating gaging stations. Watersheds are displayed as colors to facilitate differentiation.....	12
Figure 2: Schematic representation of the random forest algorithm.....	18
Figure 3: Schematic representation of a simple neural network with two hidden layers.	19
Figure 4: A standard linear model vs. a generalized additive model for the same data....	20
Figure 5: Example structure of leave-one-out cross-validation.....	22
Figure 6: Flowchart summarizing the methodology. Includes data from the USGS' Hydro-Climatic Data Network, USGS' StreamStats Reports, USGS' MA Low Flow Report, and the Livneh Climate Dataset.	25
Figure 7: Multiple linear regression estimates vs. actual historical 7Q10s (cfs).	27
Figure 8: Residuals for multiple linear regression (cfs).....	28
Figure 9: Logarithmic-transformed linear regression estimates vs. actual historical 7Q10s (cfs).	30
Figure 10: Residuals for logarithmic-transformed linear regression (cfs).....	31
Figure 11: Estimated out-of-bag error vs. number of trees applied.	32
Figure 12: Residuals for the random forest estimates (cfs).	33
Figure 13: Random forest estimated 7Q10 values vs. actual historical 7Q10s (cfs).	34
Figure 14: Neural network estimated 7Q10 values vs. actual historical 7Q10s (cfs).....	35
Figure 15: Residuals for the neural network model (cfs).	36
Figure 16: Generalized additive model estimated 7Q10 values vs. actual historical 7Q10s (cfs).	37
Figure 17: Residuals for the generalized additive model.....	38
Figure 18 (a–c): Raw 7Q10 estimates using each methodology.....	42
Figure 19: Summary of this study's experimental design.	48
Figure 20: 94 unimpaired gaged basins in the Northeast United States.	50
Figure 21: Parameter ranges after calibration and min-max normalization.	56
Figure 22: Example hydrograph for gage 01552000 that demonstrates the issue with calibrating daily rainfall-runoff models for low-flows. Monthly averages are calculated for 1980-2016, the full length of model runs.....	57
Figure 23 (a-b): Month of Historic Annual 7-Day Low Flows (a, left) and Corresponding 7-Day-10-Year Low Flow Months of Occurrence (b, right).....	58
Figure 24: Difference between hard clustering algorithms (e.g. K-Means) and soft clustering algorithms (e.g. Fuzzy-C Means).....	60
Figure 25: Results from calibration using KGE to evaluate daily streamflow estimation.	64
Figure 26a-b: Results from calibration for basins smaller than 100 mi ² (a, left) and basins larger than 100 mi ² (b, right) using KGE to evaluate daily streamflow estimation.	65
Figure 27: Silhouette Analysis for the Fuzzy C-Means clustering.	68
Figure 28: Daily streamflow KGE using FCM for calibration.	70
Figure 29: Results from calibration for basins smaller than 100 mi ² (a, left) and basins larger than 100 mi ² (b, right) using KGE to evaluate daily streamflow estimation.	71
Figure 30: Bias created from each model used, arranged by drainage area.	75

Figure 31: Generalized process for creating an effective decision support system. Rectangular boxes highlight the major steps and hexagonal boxes represent substeps for each major step. 82

Figure 32: Data model for the first prototype of the decision support system. The first module utilizes daily streamflow data from the USGS’ Current Water Data for the Nation (U.S. Geological Survey, 2016) and the second utilizes the watershed delineation and corresponding physical watershed characteristics from the USGS’ StreamStats program (Ries et al., 2008). 95

Figure 33: Examples of graphics (created with ggplot2) and maps (created with leaflet) used in the decision support system webtool. 97

Figure 34: Documentation provided for the decision support system webtool. 98

Figure 35: Stakeholder-driven plans for maintaining the program. 102

CHAPTER 1

INTRODUCTION

Although annual and seasonal streamflow data have important roles in water resources planning, estimating the frequency and magnitude of hydrologic extremes (floods and droughts) is important for many reasons, including allocating water resources, improving flood warning systems, detecting the onset and severity of droughts, protecting aquatic species, and enhancing the recreational benefits of lakes and rivers (Wiche and Holmes, 2016; Blum et al., 2019). At locations where stream gages can provide long-term streamflow data, extreme hydrologic flow metrics are traditionally calculated directly from the daily streamflow time-series data. The most commonly applied low flow planning metric is the 7-day-10-year low flow (7Q10). This statistic is used for a wide range of management actions, including wastewater treatment plant design, aquatic species protection, and hydropower management (Blum et al., 2019). The 7Q10 is defined as the lowest streamflow for 7 consecutive days that occurs on average once every 10 years (EPA, 2018). At gaged locations with historic streamflow data, the 7Q10 is calculated by: 1) Determining a running average of 7-day flows from the historic daily streamflow, 2) Extracting the annual lowest 7-day average flow from each year, 3) Fitting a quantile distribution to the annual 7-day low flows, and 4) Extracting the 7Q10 by using the 7-day average flow at the 10% exceedance probability.

This procedure for calculating the 7Q10 assumes the factors impacting streamflow (meteorology, land-use, water management, etc.) remain statistically consistent over time (that is, are stationary). Recent studies support that global climate is not stationary but is changing (Milly et al., 2008; Salas et al., 2018). For instance,

Williams et al. (2022) estimate that the southwestern United States is experiencing its driest 22-year period since 800 A.D., with approximately 20% of the drought being attributed to recent anthropogenic changes. In contrast, recent studies in the Northeast United States have found that both average baseflows and 7-day summer baseflows are increasing with statistical significance (Ayers et al., 2022; Hodgkins and Dudley, 2011). In the Mid-Atlantic, Blum et al. (2019) found increasing 7Q10s in the northern part of the Mid-Atlantic (New York, Pennsylvania) and decreasing 7Q10s in lower Mid-Atlantic (Virginia, Maryland) (Blum et al., 2019). In addition, Blum et al. found that when a trend in the annual low flows is detected, using the most recent 30 years of the streamflow record reduces the error and bias in 7Q10 estimators compared to using the full record. This result implies that using the full record of historic streamflow to calculate the 7Q10 fails to account for these anthropogenic changes.

Two primary approaches have emerged in the literature for estimating extreme streamflows where no historic data exists: statistical based methods and methods that model the rainfall/runoff process (physically based models). Physically based approaches traditionally estimate streamflow by incorporating the most important features of the rainfall/runoff process and meteorological data in computer models to generate streamflow time-series data, in lieu of measured, historic streamflow time-series data (e.g. Selvanathan et al., 2016). These models use first principles equations that govern rainfall runoff and river hydrology (e.g. flow through porous media, conservation of mass and momentum equations and other physical processes applied to rivers) to estimate daily streamflow. From the simulated daily streamflow data, extreme flows are calculated. Physically based hydrologic studies typically use publicly available hydrologic modeling

software to generate streamflow data and recent peer-reviewed publications contain numerous studies that have utilized physically based models to describe the physical processes that influence hydrologic responses (e.g. Hirabayashi et al, 2013; Arnell and Lloyd-Hughes, 2014; Berghuijs et al., 2016).

Statistical methods can also be used to estimate extreme streamflows at ungaged locations. Contrary to most traditional rainfall-runoff models designed for daily streamflow estimation, statistical methods can be used for direct extreme flow estimation, avoiding the need to estimate daily streamflow completely (e.g. Ries et al., 2000). One common approach incorporates ample historic data from adjacent or similar locations to inform estimation at a location of interest with little or no historic data, known as hydrologic regionalization (e.g. Guo et al., 2021). Statistical methods are applied to the locations with ample historic data to determine patterns that can be transcribed to the location of interest, typically using a method that can be characterized as physical similarity-based, regression-based, or hydrologic signature-based (Guo et al., 2021). Recently, many studies have benefited from the use of machine/deep learning algorithms for hydrologic regionalization (e.g. Ferreira et al., 2021; Li et al., 2022; Wang et al., 2023). Though machine learning has been used in hydrology for several decades, application of this technique has accelerated with increased access to data and computational power (Kratzert et al., 2019a). Recent applications of the Long Short-Term Memory (LSTM) machine learning algorithm as a substitute for traditional, physically driven rainfall-runoff models has been shown to provide better Nash-Sutcliffe-Efficiency (NSE) for daily streamflow estimation than calibrated physical models at the same locations (Kratzert and Hernegger, 2021). Many other recent studies have benefitted from

using machine learning algorithms to improve streamflow estimation as well, including using Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF) (Zhang et al., 2016; Soleymani et al., 2016; Mosavi et al., 2018; Kratzert et al., 2019b; Tongal et al., 2020).

Putting aside the challenges of estimating anthropogenic impacts momentarily, there remains the need to estimate extremes flows at ungaged locations. For locations that have a streamgage nearby (or potentially downstream of the location of interest), the ratio of the ungaged watershed area to the gaged basin area can be used to scale up or down the attributed extreme flow (Asquith and Thompson, 2008). The authors note that this method is appropriate for similar watersheds and define similar as differing in area by less than 50 percent with locations less than 100 miles apart (Asquith and Thompson, 2008). For locations that do not satisfy these requirements, it is assumed that no data exists on the stream or downstream. Here, physical or statistical models can be applied for their desired purpose (daily streamflow estimation, extreme flow estimation, etc.). Even though physically based methods can more readily account for anthropogenic factors, physically based rainfall-runoff models are infrequently used for extreme flow estimation at ungaged locations for two main reasons:

- 1) Most rainfall-runoff models were developed to estimate daily streamflows.

Without proper calibration with daily streamflow data, it is difficult to verify that a rainfall-runoff model (designed for daily streamflow estimation) is performing adequately for extreme flow estimation.

- 2) Open-source programs are available that rely on simple statistical regression equations to provide quick and accurate estimates of historic extreme flows at

ungaged locations, making it unnecessary to create a complicated daily rainfall-runoff model at an ungaged location just for the purpose of calculating a single extreme flow metric.

For these reasons, statistical methods relying on regression are the most common method for estimating extreme flows at ungaged locations. The typical process begins with the derivation of the equation, following three main steps: 1) calculate the historic extreme flow metric (e.g. the 7Q10) at various gaged locations in a homogenous hydrologic area, 2) collect the physical characteristics (watershed area, elevation, slope, etc.) for each of the watersheds attributed to the gages used above, and 3) fit a multiple linear regression (usually in log-space) to relate the input variables (the watershed area, elevation, slope, etc.) to the variable of interest (e.g. the 7Q10). After the derivation, this equation can be applied to ungaged locations by following three additional steps: 1) delineate the watershed that is attributed to an ungaged location of interest, 2) calculate the physical characteristics (the same input variables) of the delineated watershed, and 3) apply the new physical characteristics from the ungaged, delineated watershed to the regression equation to estimate the desired extreme flow metric.

Though these six steps are straightforward, the derivation of the regression equation requires extensive analysis to determine the relationship between the input variables (physical watershed characteristics) to the variable of interest (e.g. the 7Q10). Additionally, there must be a way for practitioners to delineate ungaged watersheds and calculate their physical watershed characteristics for use in the equation. Open-source programs, such as the United States Geological Survey's *StreamStats* (Ries et al., 2008) are available that expedite this process and allow the quick and accurate estimation of

extreme flows at ungaged locations. *StreamStats*' protocol for estimating flows at ungaged locations varies slightly based on the region, state, and/or the watershed being analyzed, but it relies on regression equations that have been updated with some frequency to maintain accuracy (Ries et al., 2008). The regression equations used in the program were derived following the same procedure highlighted above, and additional programs such as the Environmental Protection Agency's *Basins* (US EPA, 2019) follow a similar procedure for estimating extreme streamflow in ungaged locations.

These types of programs are typically referred to as Decision Support Systems (DSS). There are several definitions of DSS, but standard definitions in water resources define DSSs as computational systems that integrate data and models on water and relevant drivers of change to aid in management decisions (e.g., Wardropper and Brookfield, 2022). There is a long history of the application of DSSs in water resources planning and management (Serrat-Capdevila, 2011). Some major studies in the United States document the successful creation of DSS for watershed management, including a system dynamics model in the Middle Rio Grande (Tidwell et al., 2004) and the evaluation of conservation measures in the Upper San Pedro Basin (Yalcin and Lansey, 2004). Some more recent applications of DSSs for water related issues and decision support include the creation of an integrated DSS for the urban food-water-energy nexus (Ghodsvali et al., 2023), a System Dynamic-DSS model of sustainable groundwater resources management using the water-food-energy security Nexus in Alborz Province (Rahmani et al., 2022), and an irrigation supporting system for water use efficiency improvement in precision agriculture (Bonfante et al., 2019).

In addition to a thorough literature review, as part of the knowledge acquisition process for this dissertation, interviews were conducted with resource managers throughout the U.S. to determine common ways in which resource managers are calculating and applying extreme flows in practice. Resource managers emphasized the frequent use of StreamStats in practice but mentioned some improvements that could be made to the standard procedure that StreamStats' extreme flow estimation relies on. For example, some resource managers in the northeast U.S. highlighted that StreamStats calculates the 7Q10 using the full, historic record of streamflow at all the representative streamgages in a homogenous area and then the regression equation is trained on those values. StreamStats does not allow the user to alter the regression equations, even if a smaller subset of streamflow is desired (e.g., following the procedure recommended in Blum et al., 2019 to better account for anthropogenic changes). Many managers also mentioned the drawback associated with only using historic data to calculate the 7Q10, highlighting that most methods frequently applied in practice (e.g., StreamStats) do not incorporate any type of future climate projections into calculations. Some resource managers acknowledged that physical rainfall-runoff modeling can also be used for 7Q10 estimation which would address some of the issues highlighted with the statistical regression-based methodology. However, they also emphasized the major drawbacks (discussed earlier) that these models are extremely complex, challenging to calibrate, and that it is difficult to validate that they are performing as intended without sufficient historic streamflow data to compare the model to. This has encouraged resource managers who frequently use the 7Q10 to seek improvements to both the most common statistical and physically based 7Q10 estimation methods.

Based on the literature review and interviews with resource managers, this dissertation advances the science related to 7Q10 estimation in ungaged locations by exploring the use of physical modeling techniques in conjunction with machine learning algorithms. This paper attempts to improve both statistical and physical modeling techniques for 7Q10 estimation by targeting some of the main issues identified by resource managers, including the ability to incorporate climate change projections into calculations. Chapter 2 of this dissertation builds upon statistically based 7Q10 estimation in ungaged basins by testing multiple machine learning algorithms against classical statistical methodologies. Neural networks, generalized additive models, and random forests are tested against the two most common methods for regression-based 7Q10 estimation: standard multiple linear regression and multiple linear regression in log space. Additionally, in our equation derivation, two landcover and two climate variables are specifically included as input variables for 7Q10 estimation. This allows for future climate projections to be applied to the equations later, following a procedure recommended by the Transportation Research Board given in NCHRP Project 15-61 (Kilgore et al., 2019). Chapter 3 expands physical modeling techniques by testing a novel machine learning algorithm, Fuzzy C-Means clustering, to calibrate rainfall-runoff models in ungaged basins. During the interviews, resource managers mentioned that future updates to rainfall-runoff models will likely make them easier to set up for ungaged basins, but it will still be difficult to trust these uncalibrated/unvalidated models for either daily streamflow estimation or extreme flow estimation. To test if this novel methodology can calibrate the physical rainfall-runoff models in ungaged locations as intended, the models are evaluated for both daily streamflow estimation and 7Q10

estimation. Chapter 4 describes the development of a DSS webtool, created in collaboration with resource managers throughout the northeast U.S., for calculating the 7Q10 at gages and estimating the 7Q10 in ungaged basins with or without projected climate changes. Because the physical model in Chapter 3 performed well for daily flow estimation but poor for historic 7Q10 estimation, the statistical model from Chapter 2 is incorporated into the DSS. Targeting the issues raised by resource managers, we include: 1) the ability to subset streamflow data when calculating the 7Q10, 2) regression equations for 7Q10 estimation that were trained on the last 30 years of streamflow data at every gage, and 3) the ability to apply projected climate changes to the equations. The dissertation concludes by briefly discussing drought management and informed decision making.

CHAPTER 2

Low-Flow (7-Day, 10-Year) Classical Statistical and Improved Machine Learning Estimation Methodologies

This chapter appeared as DelSanto, A., Bhuiyan, M. A. E., Andreadis, K. M., & Palmer, R. N. (2023). Low-Flow (7-Day, 10-Year) Classical Statistical and Improved Machine Learning Estimation Methodologies. *Water*, 15(15), 2813. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/w15152813>.

Water resource managers require accurate estimates of the 7-day, 10-year low flow (7Q10) of streams for many reasons, including protecting aquatic species, designing wastewater treatment plants, and calculating municipal water availability. StreamStats, a publicly available web application developed by the United States Geologic Survey that is commonly used by resource managers for estimating the 7Q10 in states where it is available, utilizes state-by-state, locally calibrated regression equations for estimation. This paper expands StreamStats' methodology and improves 7Q10 estimation by developing a more regionally applicable and generalized methodology for 7Q10 estimation. In addition to classical methodologies, namely multiple linear regression (MLR) and multiple linear regression in log space (LTLR), three promising machine learning algorithms, random forest (RF) decision trees, neural networks (NN), and generalized additive models (GAM), are tested to determine if more advanced statistical methods offer improved estimation. For illustrative purposes, this methodology is applied to and verified for the full range of unimpaired, gaged basins in both the northeast and mid-Atlantic hydrologic regions of the United States (with basin sizes ranging from 2–1419 mi²) using leave-one-out cross-validation (LOOCV). Pearson's correlation coefficient (R^2), root mean square error (RMSE), Kling–Gupta Efficiency (KGE), and Nash–Sutcliffe Efficiency (NSE) are used to evaluate the performance of each method. Results suggest that each method provides

varying results based on basin size, with RF displaying the smallest average RMSE (5.85) across all ranges of basin sizes.

2.1 Data and Study Area

The study area for this research is the northeast and mid-Atlantic United States, defined here as the states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, Pennsylvania, New Jersey, Delaware, Maryland, Virginia, and West Virginia. The USGS's Hydro-Climatic Data Network, HCDN-2009 (Lins, 2009) was used to identify unimpaired streams in gaged watersheds of varying sizes and physical attributes in the study area (ranging from 2.1 mi² to 1419 mi², Figure 1). Data for these 106 stations from the HCDN were downloaded from the USGS Current Water Data for the Nation: <https://waterdata.usgs.gov/nwis/rt> (accessed on 18 December 2020). After reviewing the watershed size distribution of the HCDN sites, we determined it lacked a sufficient number of gages in extremely small watersheds (<30 mi²) for training data. In addition to the HCDN sites, 59 small sites in Massachusetts, determined to be sufficient for 7Q10 training data (Ries, 2000), were added to the training data for a total of 165 sites throughout the area of study. Appendix A includes a table of all sites used, and Figure 1 displays the corresponding watersheds.

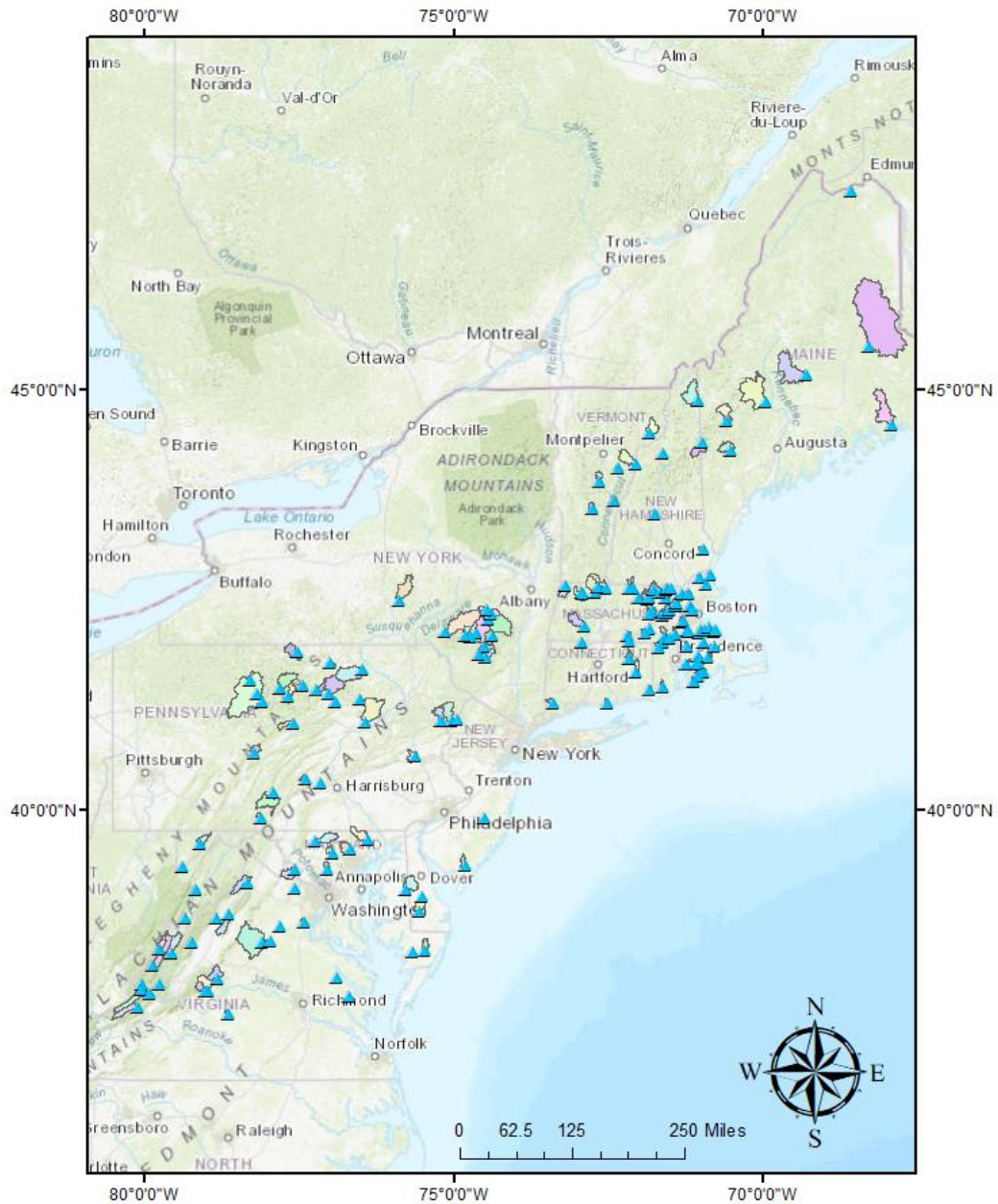


Figure 1: The 165 gages and their corresponding watersheds used for this analysis. Gages are displayed as blue triangles, the standard convention by the USGS for designating gaging stations. Watersheds are displayed as colors to facilitate differentiation.

2.2 Input Variables/Data

Daily precipitation and temperature data were extracted at each gage from the Livneh et al. (2015) hydrometeorological dataset (Livneh et al., 2015). This dataset contains air

temperature and precipitation data from approximately 20,000 weather stations monitored by GHCN-daily (U.S.), Environment Canada, and Servicio Meteorológico Nacional (Mexico) (Livneh et al., 2015). A minimum of 20 years of data was required for CONUS and Canadian stations, but due to the relative paucity of station data in Mexico, the authors followed the procedure recommended by Zhu and Lettenmaier (2007), which requires a minimum of 50 valid days of data in any given year for a station to be included (Zhu and Lettenmaier, 2007). From this station data, the data was interpolated using the SYMAP algorithm (Shepard, 1984), which employs statistical methods such as clustering analysis, regression analysis, and correlation analysis to identify patterns and relationships within spatial data. After interpolation, the authors followed procedures for quality control by computing a monthly coefficient of variation based on the standard deviation of the daily values compared to their monthly mean and removed months with a ratio of less than 0.18, determined empirically using 25 stations from 7 states with at least 15 years of data (Livneh et al., 2015). This dataset is publicly available, with gridded climate variables at $1/16^\circ$ horizontal resolution (~ 6 km) from 1950–2013 (Livneh, 2019). Static land data, including mean basin elevation, mean basin slope, forest and wetland percentages of the basin, and watershed area, were collected from USGS StreamStats Data-Collection Station reports: <https://streamstatsags.cr.usgs.gov/gagePages/html/> (accessed on 19 July 2021).

2.3 7Q10 Comparison Data

To compare 7Q10 estimates from this experiment to current statistical methodologies, we use the USGS's statistical estimation program StreamStats because of its wide usage in the states for which it has been developed in the northeast and mid-Atlantic. This

program uses logarithmic-transformed linear regression (LTLR) equations to estimate flow statistics (Ries et al., 2008). Where applied, different regression equations and variables are calculated for each state. Furthermore, states in the mid-Atlantic region use different equations (and in some cases, different variables) based on hydrologic regions within each state. Table 1 lists the candidate states and the corresponding variables used for 7Q10 estimation. Connecticut, Delaware, Maryland, New Jersey, New York, and Vermont are included last, as StreamStats 7Q10 estimates have not been developed for these states. Out of the 165 gaged sites used for training, raw 7Q10 estimates from StreamStats were available for 128 sites.

Table 1: StreamStats 7Q10 estimation by state.

State	Variables Used for 7Q10 Estimation
Massachusetts (Ries et al., 2000)	(1) Drainage area (2) Area of stratified-drift deposits per unit of stream length plus 0.1 (3) Mean basin slope (4) Indicator variable, 0 in the eastern region, 1 in the western region
Rhode Island (Bent et al., 2014)	(1) Drainage area (2) Stream density
New Hampshire (Flynn and Tasker, 2002)	(1) Drainage area (2) Mean annual temperature (3) Jun to Oct average gage precipitation
Maine (Dudley, 2004)	(1) Drainage area (2) Fraction of sand and gravel aquifers
Pennsylvania (Stuckey, 2006) Region 1 (Southeast) ¹ Region 2 (Central-east) ² Region 3 (Northwest) ³ Region 4 (Southwest) ⁴ Region 5 (Northeast) ⁵	(1) Drainage area ^{1,2,3,4,5} (2) Basin slope ¹ (3) Mean elevation ^{3,4} (4) Mean annual precipitation ^{2,3} (5) Stream density ² (6) Soil thickness ^{1,2} (7) Percent glaciation ⁵ (8) Percent carbonate bedrock ² (9) Percent forested area ⁵ (10) Percent urban area ¹

Virginia (Austin et al., 2011) Coastal Plain ¹ Piedmont ² Blue Ridge ³ Valley and Ridge ⁴ Appalachian Plateaus ⁵ Mesozoic Basins ⁶	(1) Drainage area ^{1,2,3,4,5,6}
West Virginia (Wiley, 2008) North ¹ South Central ² Eastern Panhandle ³	(1) Drainage area ^{1,2,3} (2) Longitude of basin centroid ¹
Connecticut, Delaware, Maryland, New Jersey, New York, Vermont	Unavailable

2.4 Materials and Methods

In this section, the materials and methods used in this research are described. This includes the calculation of the historical 7Q10 at each site based on the historical data, each statistical method being compared, the input variables included, a cross-validation procedure for testing, and the various efficiency/error metrics used to evaluate the performance of each method.

2.4.1 7Q10 Values at Each Site

The historical 7Q10 values based on historical data for each site were extracted from the USGS's StreamStats Data-Collection Station Reports. In addition, if at least 30 years of continuous, daily streamflow data is available for a site, the "fasstr" software package: <https://cran.r-project.org/web/packages/fasstr/index.html> (accessed on 19 July 2021) is used to calculate the 7Q10 directly from the daily streamflow data. This package fits a quantile distribution to daily streamflow data that allows for the efficient calculation of low-flow frequency analysis metrics, including the 7Q10. As expected, these 7Q10 values

were virtually identical to the 7Q10 values calculated by the USGS at each site. These values, noted as the “true 7Q10” values for each site, can also be found in Appendix A.

2.4.2 Statistical Methods

In this analysis, five statistical methods are applied. Two classical statistical methods, namely multiple linear regression and logarithmic-transformed linear regression, are tested alongside three machine learning algorithms, namely random forest decision trees, neural networks, and generalized additive models. For the machine learning algorithms, feature scaling is applied to the input variables before method application using min-max normalization:

$$x = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

2.4.2.1 Multiple Linear Regression

Multiple linear regression (MLR) is a simple, common methodology that takes the general form of

$$Y_i = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i \quad (2)$$

where Y_i is the estimate of the dependent variable for site i , X_1 to X_n are the n independent variables, b_0 to b_n are the $n + 1$ regression model coefficients, and ε_i is the residual error for site i . Assumptions for use of MLR are (1) the relationship displays linearity, (2) the mean of ε_i is zero, (3) the variance of the ε_i is constant and independent of X_n , (4) the ε_i are normally distributed, and (5) the ε_i are independent (Iman and Conover, 1983). For this study, we force the intercept b_0 to be 0 since a basin with 0 area should have 0 flow.

2.4.2.2 Logarithmic-Transformed Linear Regression

Logarithmic-transformed linear regression (LTLR) is the most used method for 7Q10 estimation because it can correct for spatial correlation and differences in streamflow record lengths (Tasker and Stedinger, 1989). In addition, streamflow and basin characteristics used in hydrologic regression have been found to be log-normally distributed, with residuals (calculated by subtracting the estimated values from the observed values) that were not randomly distributed when multiple linear regression was applied, suggesting that the variables should be transformed to log space (Ries, 2000). This results in a model of the form:

$$\log Y_i = b_0 + b_1 \log X_1 + b_2 \log X_2 + \dots + b_n \log X_n + \varepsilon_i$$

Using base 10, the equation takes the general form of

$$Y_i = 10^{b_0} (X_1^{b_1}) (X_2^{b_2}) \dots (X_n^{b_n}) 10^{\varepsilon_i} \quad (3)$$

Though theory suggests that LTLR is the preferred method for 7Q10 estimation, in practice, both MLR and LTLR have been found to perform similarly (Vogel and Kroll, 1990).

2.4.2.3 Random Forest

The random forest (RF) algorithm applied here is a non-parametric, tree-based regression model (Breiman, 2001). RFs use bootstrap aggregation, where bootstrap samples are randomly chosen with substitution seeking a lower test error by variance reduction. RFs consist of numerous decision trees (Figure 2).

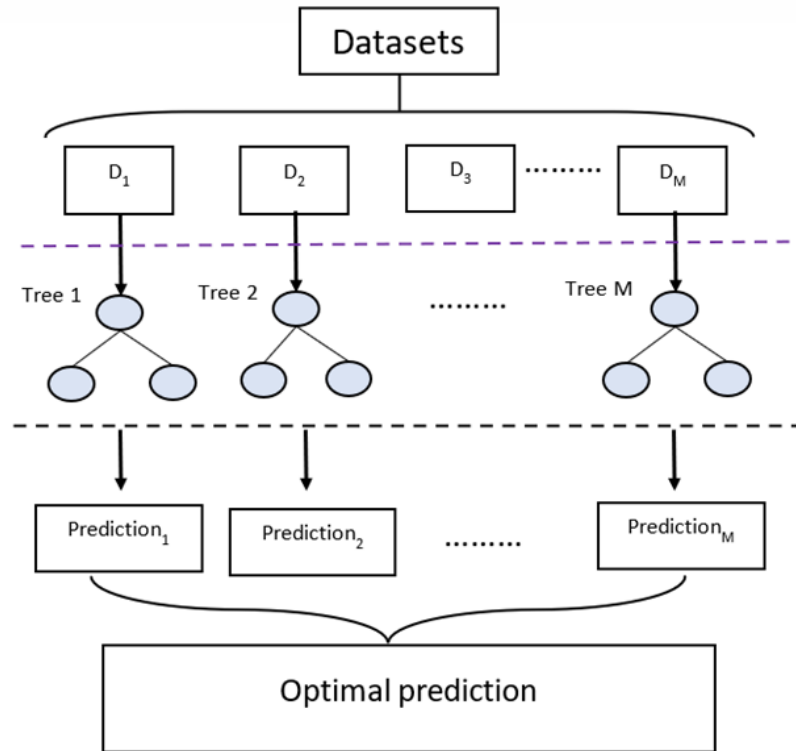


Figure 2: Schematic representation of the random forest algorithm

The RF model is optimized by tuning or calibrating its three major hyperparameters: (1) “ m_{try} ”, the number of predictors that will be randomly sampled at each split when creating the tree models; (2) “ n_{trees} ”, the number of decision trees contained in the ensemble; and (3) the minimum size of terminal nodes, “ n_t .” All parameters were manually tuned to create a stable model using the package “randomForest”: <https://cran.r-project.org/web/packages/randomForest/index.html> (accessed on 20 July 2021) in R.

2.4.2.4 Neural Networks

Neural networks (NN) are a class of machine learning algorithms inspired by the structure and function of the human brain (McCulloch and Pitts, 1943). The three primary types of layers are the input layer, one or more hidden layers, and the output layer. Each neuron takes inputs, performs a weighted sum of these inputs, applies an activation

function to produce an output, and then passes this output to neurons in the next layer. The connections between neurons have associated weights, which the network learns from data during the training process. The network adjusts its weights iteratively using optimization algorithms to minimize the difference between its predictions and the actual targets. A simple neural network structure is highlighted in Figure 3.

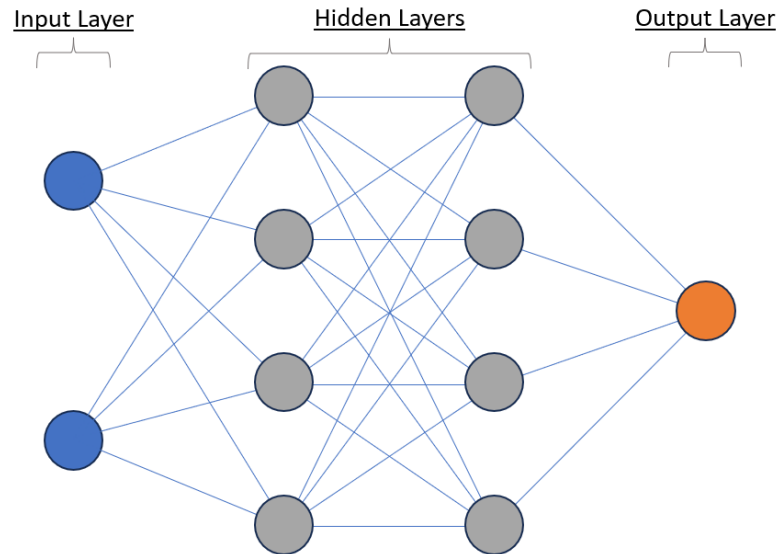


Figure 3: Schematic representation of a simple neural network with two hidden layers.

Neural networks are optimized by tuning major parameters, including the number of hidden layers, the limiting threshold for the partial derivatives of the error function as stopping criteria, and the maximum allowable steps for training. Additionally, the number of neurons per hidden layer, initial weights, activation functions, and learning rate can be customized for different scenarios. All parameters were manually tuned to create a stable model that converges using the “neuralnet” package:

<https://www.rdocumentation.org/packages/neuralnet/versions/1.44.2/topics/neuralnet>

(accessed on 21 July 2021) in R.

2.4.2.5 Generalized Additive Models

Generalized additive models (GAM) (Hastie and Tibshirani, 1986) represent a flexible extension of traditional linear models. GAMs capture non-linear dependencies of data through the application of smoothing functions which allows for finding complex relationships between variables. By employing smoothing functions such as cubic regression splines, GAMs accommodate non-linear patterns and mitigate issues of model misspecification. This property of GAMs makes it well-suited for scenarios where linear models may fall short in capturing intricate patterns. Furthermore, GAM does not impose assumptions on the underlying distribution of the response variable, enabling it to incorporate various response distributions appropriately. An example of a standard linear model and a GAM applied to the same data is provided in Figure 4.

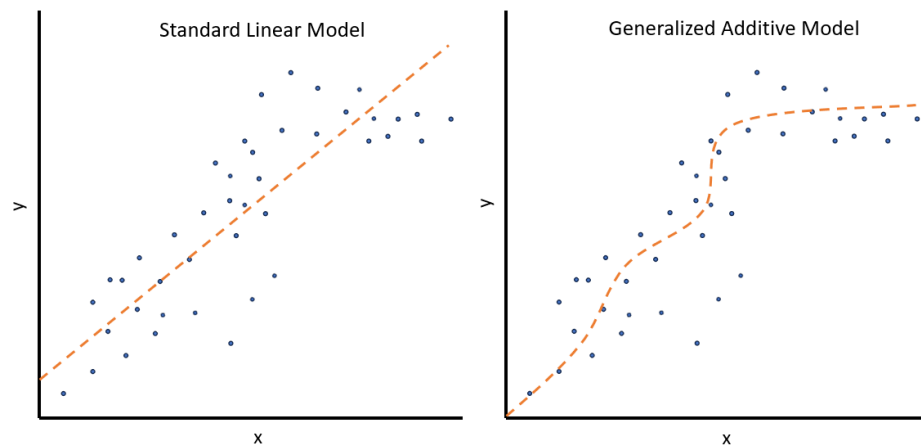


Figure 4: A standard linear model vs. a generalized additive model for the same data.

GAMs are optimized by tuning several parameters, including “gamma” to increase smoothing, “family” to specify the distribution to be used, and “weights” to designate prior weights on the contribution of the data to the log-likelihood. All parameters were manually tuned to create a stable model using the GAM function from the “mgcv”

package: <https://www.rdocumentation.org/packages/mgcv/versions/1.9-0/topics/gam> (accessed on 17 August 2021) in R.

2.4.3 Input Variables

Stational land data, including mean basin elevation, percent mean basin slope, percent landcover considered wetland and forest, and basin area, were collected from the USGS’s StreamStats Data-Collection Station Reports. These data are direct inputs into the statistical models. In addition, timeseries of daily precipitation and maximum temperature were extracted at each of the gages from Livneh et al. (2015). A running cumulative 30-day precipitation value was calculated, as well as the corresponding average 30-day maximum daily temperature. Attempting to isolate when a 7Q10 flow would occur, we extracted the lowest 30-day cumulative precipitation limited to only 30-day periods of high temperatures (>90th percentile). The 30-day cumulative precipitation and corresponding high average temperature were recorded. A list of all input variables is included in Table 2.

Table 2: Input variables for 7Q10 estimation

Variable	Description
Area (mi ²)	Watershed area
Mean Elevation (ft)	Average elevation of the watershed
Slope (%)	Average basin slope
Percent Wetland (%)	Wetland percentage of the watershed
Percent Forest (%)	Forest percentage of the watershed
Min 30-day Cumulative Precipitation (mm)	Lowest 30-day cumulative precipitation, limited to abnormally hot periods ($X > 90$ th percentile temperatures)
Average 30-day High Temperatures (C)	Average 30-day temperature during the corresponding period of low cumulative precipitation

2.4.4 Leave-One-Out Cross-Validation (LOOCV)

The leave-one-out cross-validation (LOOCV) method is an extreme version of K-fold cross-validation where $K = N$ (Molinaro et al., 2005). It is an iterative process that is executed the same number of times as the number of data points. Only one value is used as the test set, while all other values are used as the training set. This iterative process is run for every value so that there is a test set value for every value in the dataset, which allows for a new test set to be created with all of the individual test set values (Figure 5).

The new test set can then be evaluated using traditional error metrics and analysis.

LOOCV can be computationally intensive, but for relatively small datasets, it can provide better performance than K-fold cross-validation due to the largest possible training set being used to estimate each test set value (Molinaro et al., 2005).

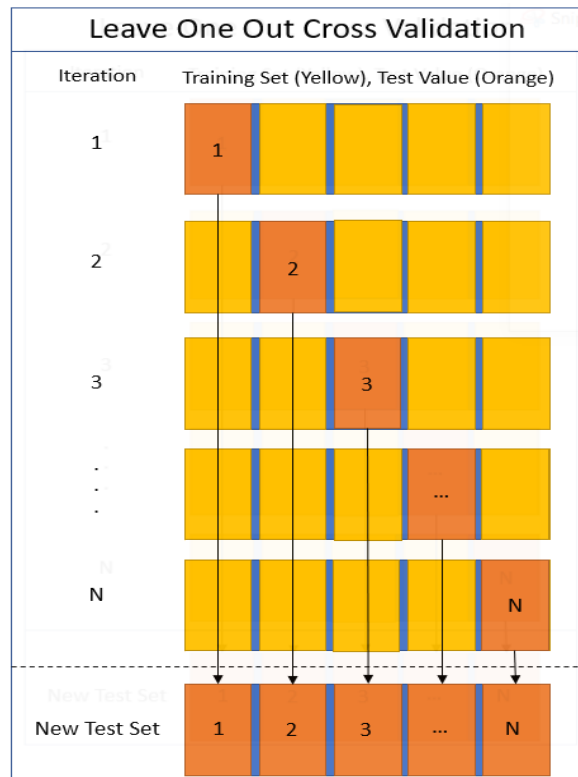


Figure 5: Example structure of leave-one-out cross-validation.

2.4.5 Error Metrics

R^2 and RMSE are used to directly evaluate the random and systematic error of each method. In addition, the Nash–Sutcliffe Efficiency (NSE) and the Kling–Gupta Efficiency (KGE) are also included because of their frequent usage for evaluating streamflow models.

The widely known coefficient of determination, R^2 (Wright, 1921), will be one of the metrics for evaluating model performance. Values range from 0 (no correlation) to 1 (perfect correlation). This value is calculated using the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Here, y_i is the observed 7Q10, \hat{y} is the model predicted 7Q10, and n is the number of samples used in the calculation.

RMSE is used to evaluate error because it is among the most used indicators for evaluation of model performance (Shortridge et al., 2016). Similar studies have also chosen RMSE over MAE for its sensitivity to outliers (Mekanik et al., 2016). The general equation is given by:

$$RMSE = \sqrt{\frac{1}{n} * \sum_{1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Hydrologists commonly use the Nash–Sutcliffe Efficiency (Nash and Sutcliffe, 1970) and Kling–Gupta Efficiency (Gupta et al., 2009) for streamflow modelling evaluation. The Nash–Sutcliffe Efficiency (NSE) signals a model’s ability to predict variables different from the mean. NSE is calculated given:

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (6)$$

NSE values range from negative infinity (indicating a poor model) and 1 (indicating a perfect fit between observed and predicted values). Negative values indicate that the mean is a better predictor of the observed values than the model.

Furthermore, the Kling–Gupta Efficiency (KGE) (Gupta et al., 2009) is widely used for hydrologic applications (Formetta et al., 2011; Beck et al., 2016). KGE provides three components, the general correlation (r term), the bias (beta term), and the relative variability (alpha term), between the modelled and observed values. KGE is calculated using the following formula:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (7)$$

$$\alpha = \frac{\sigma_m}{\sigma_o}$$

$$\beta = \frac{\mu_m}{\mu_o}$$

where σ_m is the standard deviation of model, σ_o is the standard deviation of reference, μ_m is the mean of model, and μ_o is the mean of reference. Like NSE, KGE values range from negative infinity (poor performance) to 1 (best performance). Here, r is Pearson's correlation coefficient, α represents the variability error, and β indicates the bias error.

The methodology of this paper is summarized in Figure 6. This figure includes all datasets, variables, and methods utilized.

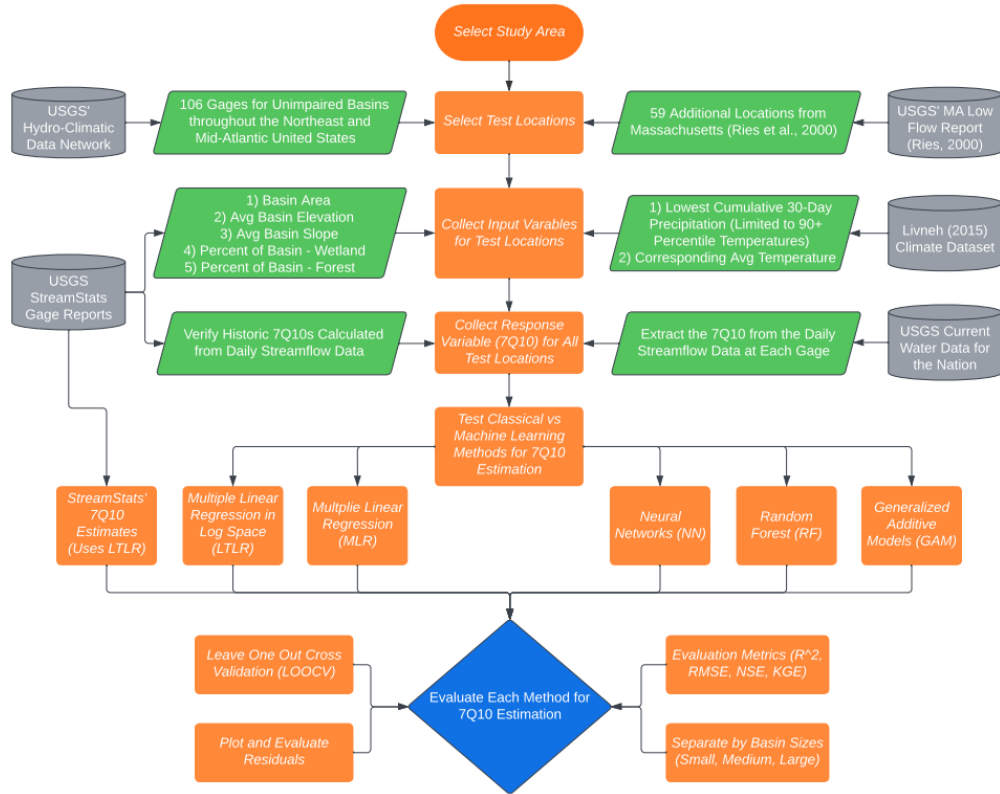


Figure 6: Flowchart summarizing the methodology. Includes data from the USGS' Hydro-Climatic Data Network, USGS' StreamStats Reports, USGS' MA Low Flow Report, and the Livneh Climate Dataset.

2.5 Results and Discussion

In this section, the development and general performance of each statistical model are evaluated. For all methods excluding neural networks where it is not applicable, we present the significance of variables using the standard p -value, with four thresholds given by minorly significant ($0.1 > X > 0.05$), moderately significant ($0.05 > X > 0.01$), largely significant ($0.01 > X > 0.001$), and extremely significant ($X < 0.001$).

2.5.1 Multiple Linear Regression

Applying multiple linear regression to the input variables, with the constraint that the intercept \mathbf{b}_0 is set to 0, gives the following results (Table 3).

Table 3: Multiple linear regression input variables and significance.

Variable	Estimate	p-Value	Significance
Area (mi ²)	0.0579833	2×10^{-16}	<0.001
Mean Elevation (ft)	0.0014461	0.05343	<0.1
Slope (%)	-0.2804801	0.00197	<0.01
Percent Wetland (%)	-0.0618991	0.27690	No significance
Percent Forest (%)	0.0048811	0.86061	No significance
Min 30-day Cumulative Precipitation (mm)	0.3421899	0.00246	<0.01
Average 30-day High Temperatures (C)	-0.0466718	0.58980	No significance

As expected, the area was found to be extremely significant at the 0.001 level. In addition, slope and precipitation were found to be largely significant at the 0.01 level. Elevation was found to be minimally significant at the 0.1 level but did show significance and improved results. Neither landcover variable, the percentage of basin considered to be forest or wetland, showed significance. The resulting equation, only including statistically significant variables, results in the following:

$$7Q10 = 0.0579833 * (\text{Area}) + 0.0014461 * (\text{Elevation}) \\ - 0.2804801 * (\text{Slope}) + 0.3421899 * (\text{Precip})$$

$$(\text{R Square} = 0.7481, \text{Residual Standard Error} = 7.693, p\text{-Value} = 2 \times 10^{-16})$$

The sign of each variable aligns intuitively with the expected relationship between that variable and the 7Q10. Increasing area and precipitation allows for additional water during low flows, leading to positive coefficients. Elevation and slope relate directly to baseflows, and as baseflows play a large role in low flows, the relationship between

baseflow and 7Q10s should be similar. High elevations are typically associated with higher baseflows (Rumsey et al., 2015), leading to an increasing relationship between 7Q10s and elevation. The relationship between slope and 7Q10 was expected to be positive, but as noted by Rumsey et al. (2015), “positive correlations between slope and baseflow are expected to be related to effects of elevation, but slope steepness is known to affect rates of groundwater transmission and determines whether groundwater will reach a channel network or be retained in the soil” (Rumsey et al., 2015), making it reasonable that increasing slope may not increase 7Q10s.

Figure 7 presents the MLR 7Q10 estimates and the actual and historical 7Q10s. The line represents a perfect fit, and the points closest to the line indicate lower bias.

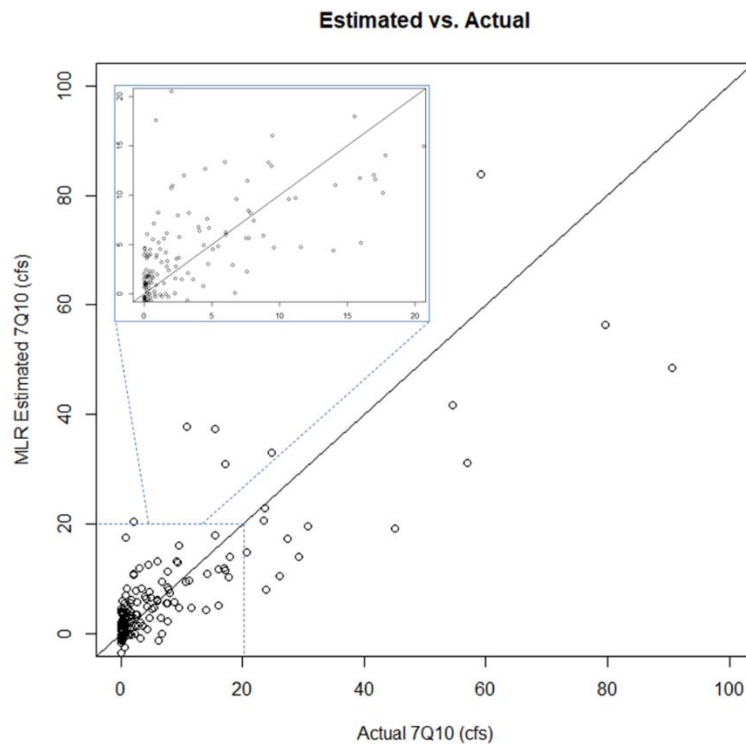


Figure 7: Multiple linear regression estimates vs. actual historical 7Q10s (cfs).

A known weakness of applying MLR for extremely low flow estimation is that the residuals are not randomly distributed, as noted in most StreamStats low flow reports

cited earlier. Plotting the residuals, from the smallest to largest observed 7Q10, confirms this (Figure 8).

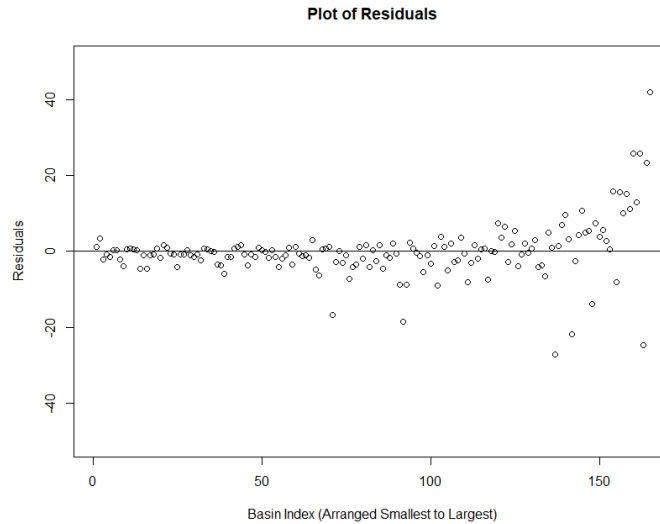


Figure 8: Residuals for multiple linear regression (cfs).

Though the MLR fit is statistically significant, the plot of residuals suggests that MLR may not be the appropriate method for this case. Because of this, the next method, logarithmic-transformed linear regression (LTLR), is the traditional method for estimating 7Q10s in small basins. This method is traditionally used by StreamStats, though StreamStats reports refer to it as generalized least squares regression with a logarithmic transformation. This methodology accounts for the drawbacks of simple MLR, including the non-random residuals, spatially distributed correlation, and differences in record lengths.

2.5.2 Logarithmic-Transformed Linear Regression

Similarly, applying LTLR to the same input data leads to the following results, given in Table 4.

Table 4: Logarithmic-transformed linear regression input variables and significance.

Variable	Estimate	p-Value	Significance
Intercept	4.27157	0.1590	No significance
Area (mi ²)	1.31308	2×10^{-16}	<0.001
Mean Elevation (ft)	-0.11573	0.2908	No significance
Slope (%)	-0.19413	0.0303	<0.05
Percent Wetland (%)	-0.02036	0.7542	No significance
Percent Forest (%)	0.22437	0.0489	<0.05
Min 30-day Cumulative Precipitation (mm)	0.31049	0.0160	<0.05
Average 30-day High Temperatures (C)	-4.37462	0.0309	<0.05

In log space, only area was found to be extremely significant at the 0.001 level, while slope, precipitation, and temperature are moderately significant at the 0.05 level.

Additionally, the percentage of the basin considered to be forest was found to be moderately significant at the 0.05 level in log space when it was not found to be significant using basic MLR. Once again, the percentage of the basin considered to be wetland was not found to be significant, but surprisingly, elevation was not found to be significant in log space, while it was just above the threshold for moderate importance (p -value~0.05) using MLR. Only including the significant variables leads to the following equation:

$$\log(7Q_{10}) = 1.31308 * \log(\text{Area}) - 0.19413 * \log(\text{Slope}) + 0.22437 * \log(\text{Forest}) + 0.31049 * \log(\text{Precip}) - 4.37462 * \log(\text{Temp})$$

R Square = 0.67, Residual Standard Error = 0.6139, p -Value = 2×10^{-16} .

Once again, plotting the estimated values vs. the actual historical 7Q10s (but this time, in log space) is displayed in Figure 9.

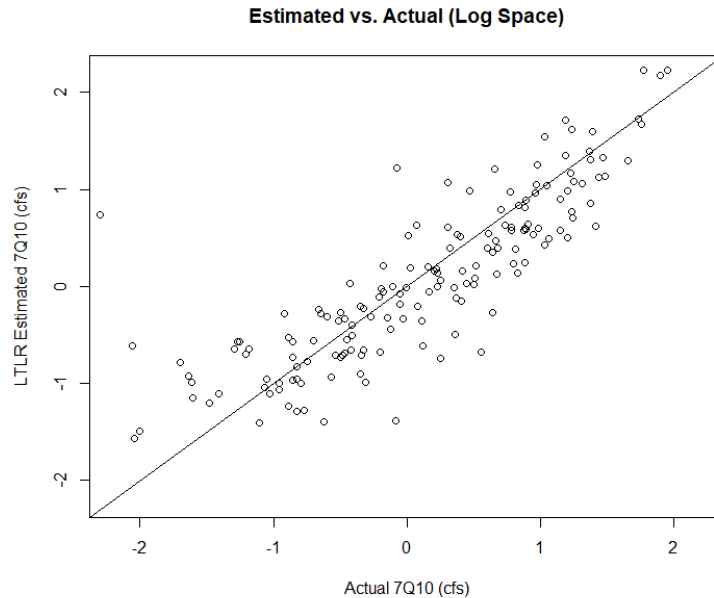


Figure 9: Logarithmic-transformed linear regression estimates vs. actual historical 7Q10s (cfs).

This fit is noticeably more linear but moves away from linearity for extremely small 7Q10 values (Actual 7Q10s $< 10^{-1}$ cfs). These extremely small points can be ignored due to significant figures, as these very small numbers suggest that the stream's 7Q10 is essentially 0 flow (ephemeral streams). More importantly, a goal of this experiment is to include much larger basin areas (and their corresponding 7Q10s) than are traditionally accounted for in regression equations. The largest 7Q10s, which correspond with the largest basins in the analysis, are found in the top right of Figure 9 and seem to continue to fit the general trend of linearity in log space. However, it should be noted that even though those points are similar distances from the line in log space, the difference is much larger than Figure 6 suggests. The three points correspond to actual 7Q10s of 79.67, 59.16, and 90.51 cfs, with their corresponding estimates to be 151.21, 170.09, and

167.49 cfs, respectively. To further examine this, Figure 10 displays the residuals in log space, once again arranged from smallest to largest 7Q10.

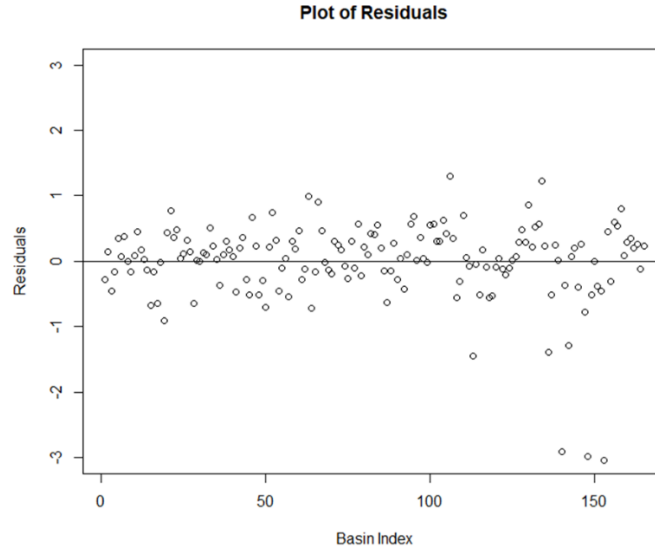


Figure 10: Residuals for logarithmic-transformed linear regression (cfs).

Besides the three points in the bottom right corner of Figure 10, which were discussed previously, the residuals in Figure 10 appear to be more consistently distributed than the residuals in Figure 8, suggesting that using LTLR is the preferred method over general MLR for 7Q10 estimation.

The final equation, translated back into standard space, is given below.

$$7Q10 = Area^{1.31308} Slope^{-0.19413} Forest^{0.22437} Precip^{0.31049} Temp^{-4.37462}$$

2.5.3 Random Forest

Applying the random forest (RF) machine learning algorithm to the input data yields the following results in Table 5.

Table 5: Random forest input variables and significance.

Variable	% Included MSE	p-Value	Significance
Area (mi ²)	57.580982	0.0099	<0.01

Mean Elevation (ft)	6.911398	0.03465	<0.05
Slope (%)	2.257348	0.07228	<0.1
Percent Wetland (%)	-2.099959	0.9901	No significance
Percent Forest (%)	3.335443	0.08911	<0.1
Min 30-day Cumulative Precipitation (mm)	7.726635	0.04275	<0.05
Average 30-day High Temperatures (C)	1.265636	0.6634	No significance

Area was found to be significant at the 0.01 level, with elevation and precipitation significant at the 0.05 level, and both slope and percent forest significant at the 0.1 level.

Temperature was not found to be significant using the random forest model or the multiple linear regression model, making it only significant in log space. In all three cases so far (MLR, LTLR, and RF), the percentage of the basin considered wetland was not found to be significant. Figure 11 displays the estimated out-of-bag error as a function of the number of decision trees.

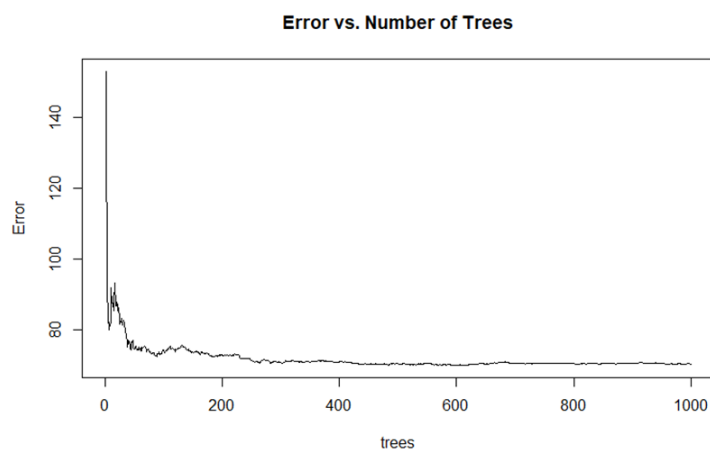


Figure 11: Estimated out-of-bag error vs. number of trees applied.

Figure 11 displays that the model stabilized around 100 trees. Though the RF method does not make assumptions about normality, a plot of the residuals given in Figure 12, once again arranged from the smallest to largest 7Q10, shows that they are not randomly distributed.

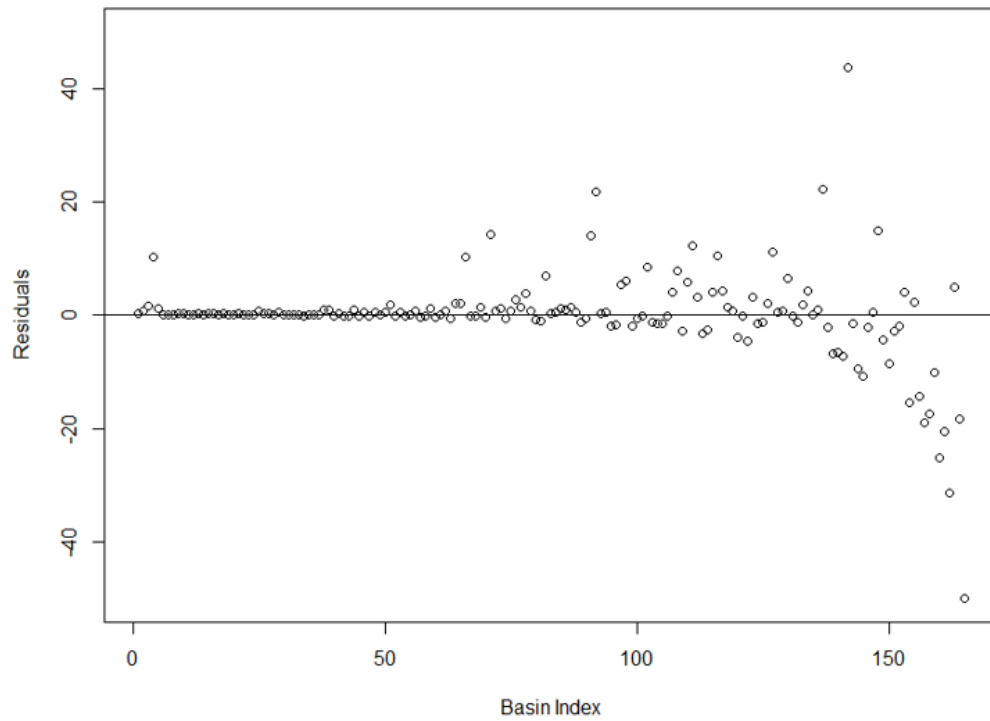


Figure 12: Residuals for the random forest estimates (cfs).

Plotting the 7Q10 values estimated using the random forest method vs. the actual historical 7Q10s is given in Figure 13.

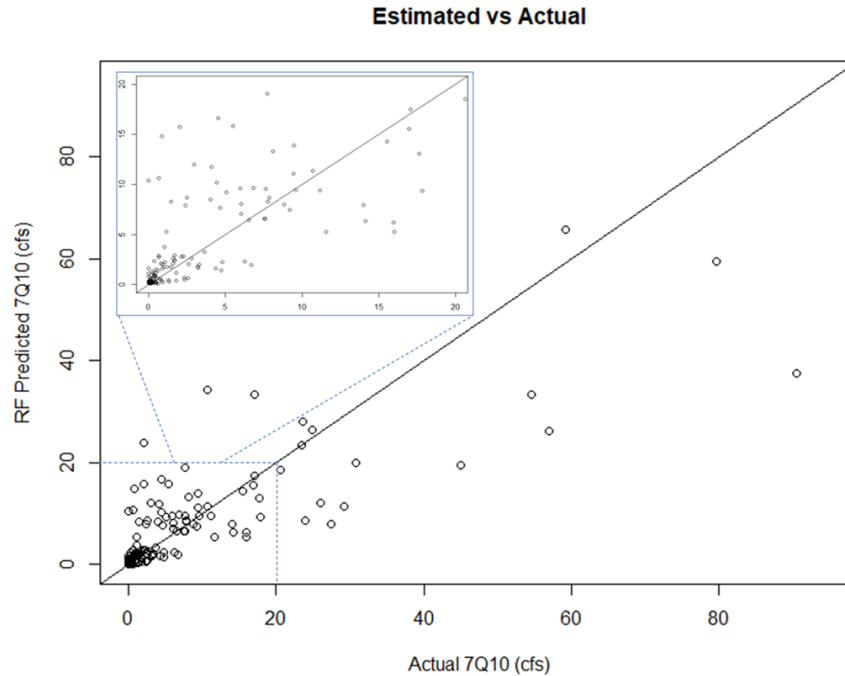


Figure 13: Random forest estimated 7Q10 values vs. actual historical 7Q10s (cfs).

For relatively smaller 7Q10s, RF estimates are similar to the other methods. However, RF underestimates actual historical 7Q10s that are over 20 cfs. For this range, there are only 3 points above the line (overestimation) and 11 points under the line (underestimation). This range is specifically difficult to estimate because there are very few unimpaired watersheds in the study area that are large enough to have 7Q10s in this range, limiting the available training data.

2.5.4 Neural Network

Neural networks were applied to the input data using a variety of tuning parameters. The addition of multiple hidden layers increased computation time, caused failure to converge in some cases, and did not improve model performance, so the final neural network described only included one hidden layer, an associated convergence threshold of 0.01, and a maximum step of 1×10^5 . In this section, no table of variable importance and significance is included, as calculating p -values for neural networks is not common

practice. Neural networks are highly complex models with multiple weights and parameters. When calculating p -values for each weight or parameter, it is effectively conducting multiple hypothesis tests for each. This introduces the risk of the multiple comparisons problem (Holm, 1979), where the probability of obtaining false positives (significant p -values) increases, which can lead to misleading results. Instead, we display the general results in Figure 14.

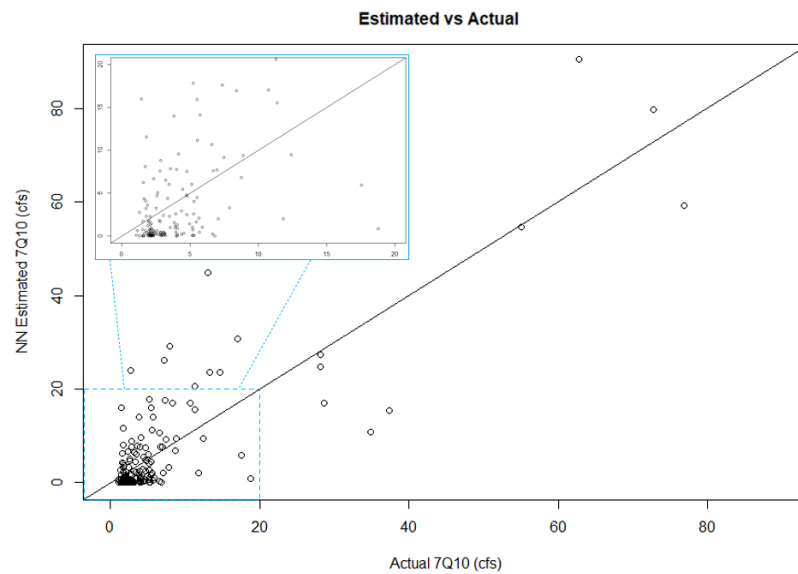


Figure 14: Neural network estimated 7Q10 values vs. actual historical 7Q10s (cfs).

Figure 14 displays that the NN model overestimates smaller 7Q10s (especially in the 0–20 cfs range) and overestimates 7Q10s larger than 20 cfs (7 points below the line, as opposed to 2 above). This should be corroborated by the residuals, which are displayed in Figure 15, once again organized from smallest actual 7Q10 to largest.

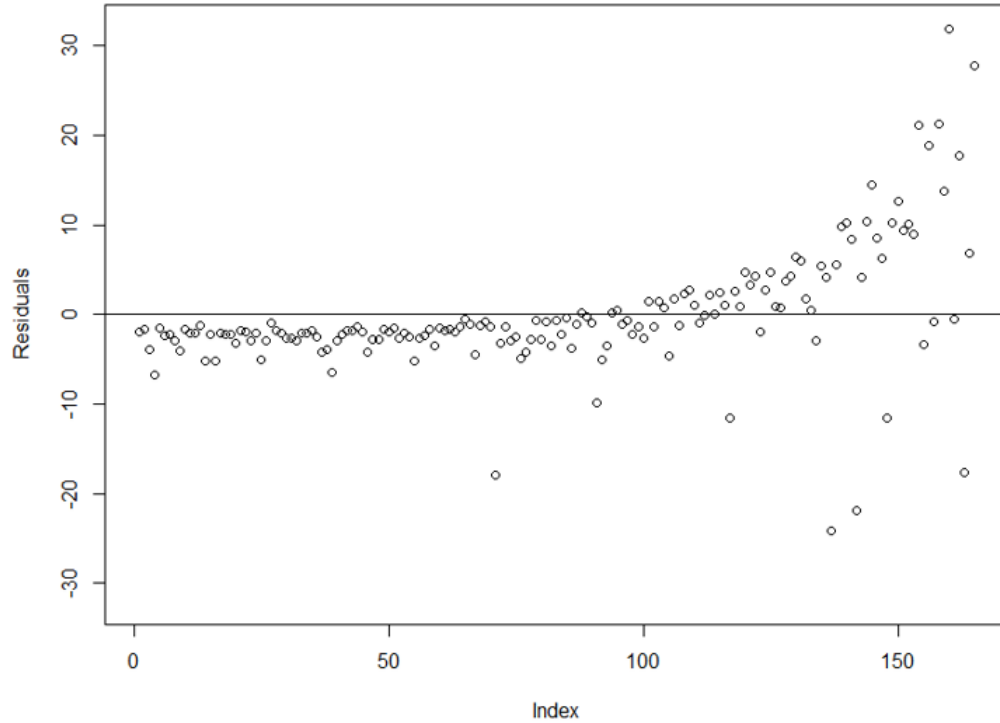


Figure 15: Residuals for the neural network model (cfs).

Figure 15 suggests that the neural network consistently overestimates the smaller 7Q10 values and underestimates the larger 7Q10 values. Even if this model proves to have the smallest error metrics, this is a significant drawback.

2.5.5 Generalized Additive Model

GAM was applied to the input data with a variety of tuning parameters. No initial weights or scale parameters were given, and the optimal model was found using the Gaussian distribution with generalized cross-validation (GCV). The optimal GAM model yields the following results in Table 6.

Table 6: Generalized additive model input variables and significance.

Variable	Estimated Degrees of Freedom	<i>p</i> -Value	Significance
Area (mi ²)	8.407	0.000000	<0.001
Mean Elevation (ft)	8.143	0.000608	<0.001
Slope (%)	1.000	0.230267	No significance

Percent Wetland (%)	1.315	0.160834	No significance
Percent Forest (%)	3.548	0.484701	No significance
Min 30-day Cumulative Precipitation (mm)	1.661	0.017755	<0.05
Average 30-day High Temperatures (C)	1.466	0.007338	<0.01

Area and elevation were both found to be extremely significant, while precipitation and temperature were found to be moderately and largely significant. The only other methodology where temperature was found to be significant was LTLR (linear regression in log space), suggesting that there may be a subtle importance of this variable that was not detected in standard space by MLR or RF. In addition to slope, neither of the landcover variables were found to be significant using GAM. The estimated 7Q10 values vs. the actual historical values are presented in Figure 16.

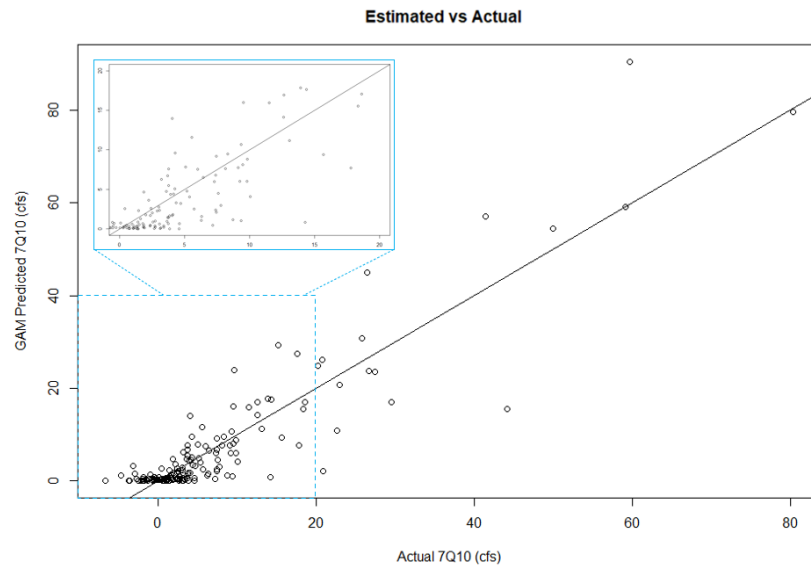


Figure 16: Generalized additive model estimated 7Q10 values vs. actual historical 7Q10s (cfs).

Figure 16 does not suggest an obvious pattern in residuals. To further analyze the model fit, we plot the residuals in Figure 17, once again organized from smallest actual 7Q10 to largest.

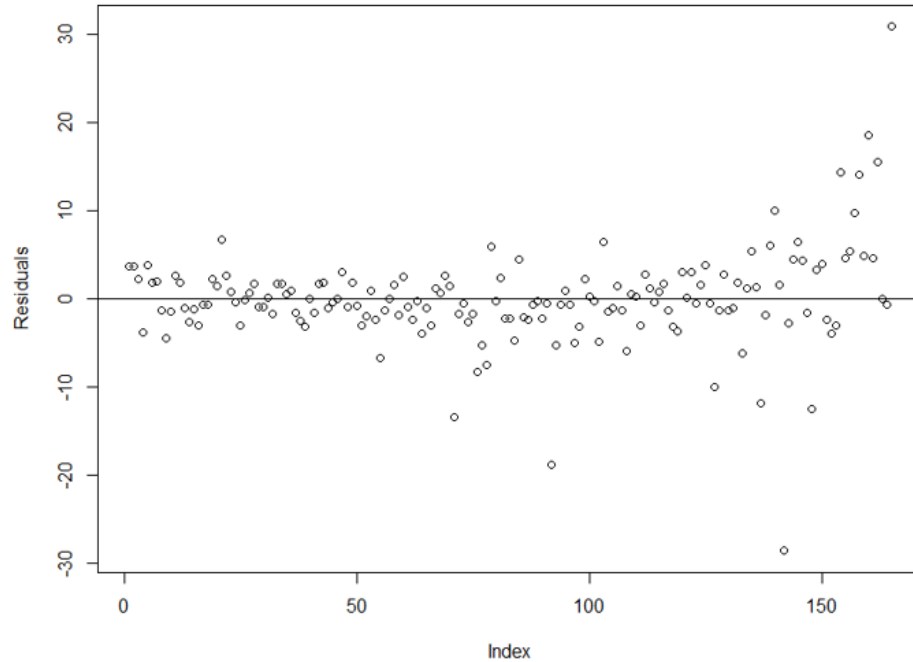


Figure 17: Residuals for the generalized additive model.

The residuals for the larger basins seem to deviate from the perfect fit line, but the residuals in Figure 17 seem to be much more randomly distributed than the neural network model that displayed a clear pattern in Figure 15.

In the following sections, we will evaluate how each method performs in comparison to StreamStats and in relation to each other. Additionally, all methods seem to perform worst for large 7Q10s, so the final section will highlight under what range of basin sizes each method performs best, using RMSE as the main metric compared to StreamStats estimates.

2.5.6 Comparisons to StreamStats Estimates

Current 7Q10 estimates were derived using the USGS's StreamStats program, discussed earlier. Estimates are only available for some states in the domain, leaving 128 data points for comparison out of the 165 total used for training. In addition, no test set was

used in the StreamStats derivations, so comparing exact estimates from each method without using a validation procedure is given in Table 7.

Table 7: Multiple method comparisons to current estimates.

Method	R²	KGE	NSE	RMSE
StreamStats Estimates	0.66	0.66	0.65	9.88
Log-Transformed Linear Regression	0.67	0.54	0.62	13.50
Multiple Linear Regression	0.70	0.80	0.63	7.14
Random Forest	0.63	0.76	0.53	7.97
Neural Network	0.73	0.82	0.67	6.79
Generalized Additive Model	0.84	0.91	0.83	5.19

Most methods perform similarly, but GAM displays the best R², KGE, NSE, and RMSE by far out of all the methods. Because of the high flexibility of GAMs, they are prone to overfitting, and this success will be further tested in Section 4.7 with LOOCV. The NN model also displays a high R² and KGE, but that is with the drawback that it overestimates small 7Q10s and underestimates large 7Q10s, as highlighted in Figure 15. MLR and RF outperform current estimates by an average of 12% in terms of KGE, as well as 25% in terms of RMSE, which is arguably the most important metric as it directly measures error. RF’s success could also be due to overfitting, which is tested in the next section. MLR, however, is a classical method for 7Q10 estimation and is not prone to overfitting. Given its success compared to StreamStats for the exact same locations that were derived using state-by-state equations, results suggest that a single, generalized methodology is appropriate.

2.5.7 General Performance

Comparing each method using leave-one-out cross-validation results in the following metrics, given in Table 8.

Table 8: Comparisons between statistical methods using leave-one-out cross-validation.

Method	R ²	KGE	NSE	RMSE
Log-Transformed Linear Regression	0.72	0.50	0.62	15.24
Multiple Linear Regression	0.60	0.73	0.47	8.53
Random Forest	0.61	0.69	0.41	8.39
Neural Network	0.53	0.69	0.36	9.41
Generalized Additive Model	0.53	0.65	0.52	12.15

Table 8 confirms that the success displayed by both NN and GAM in the previous section was due to overfitting. They display the two worst R²s and have RMSEs larger than both RF and MLR. With the addition of a test set, the RF method performance only declined 3.17% for R² (0.61, down from 0.63), 9.21% for KGE (0.69, down from 0.76), 22.64% for NSE (0.41, down from 0.53), and increased 5.27% for RMSE (8.39, up from 7.97). The average decline for MLR and LTLR was similar, at 3.41% for R², 8.08% for KGE, 12.70% for NSE, and an increase of 16.18% for RMSE. This suggests that RF's previous success was not due to overfitting, as it displayed similar declines to LTLR and MLR, which utilize straight lines for fitting.

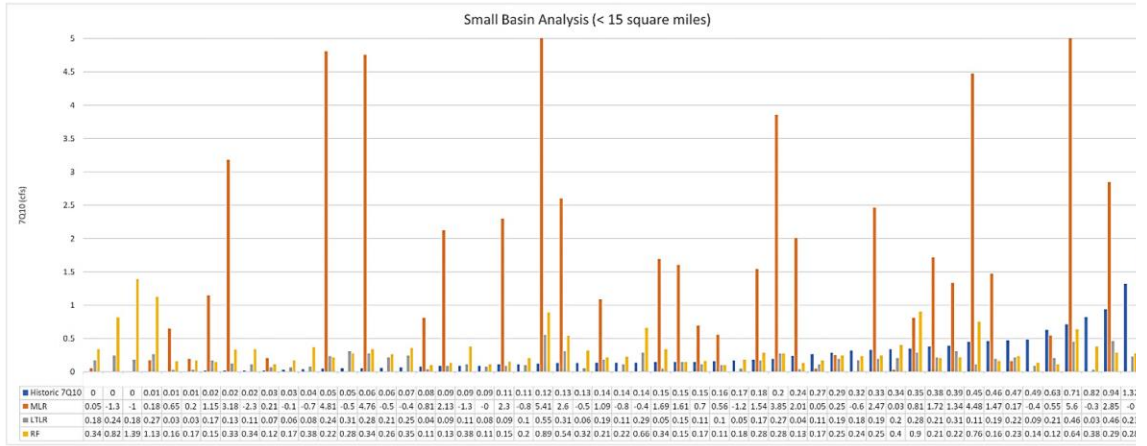
Each method performs differently based on the evaluation metric used. One explanation for this is the wide range of basin sizes included in this analysis. This is highlighted in the plots of residuals earlier, which demonstrated that many models perform poorer for larger 7Q10s. Splitting the data into three distinct subsets based on basin size will allow us to further examine why some methods display high R² but poor RMSE. For this analysis, small basins are defined as basins under 15 mi², while medium basins are basins between 15 and 70 mi², and large basins are basins larger than 70 mi². These thresholds do not have any physical meaning and are simply selected to divide the full dataset into three equally sized subsets for further analysis. In Table 9, we provide the RMSE for each

method applied to each basin size range. Especially for these subsets, RMSE is the best metric to base success on, as it measures the error of estimates vs. the actual values, which is the most important metric for resource managers who need accurate 7Q10 estimates.

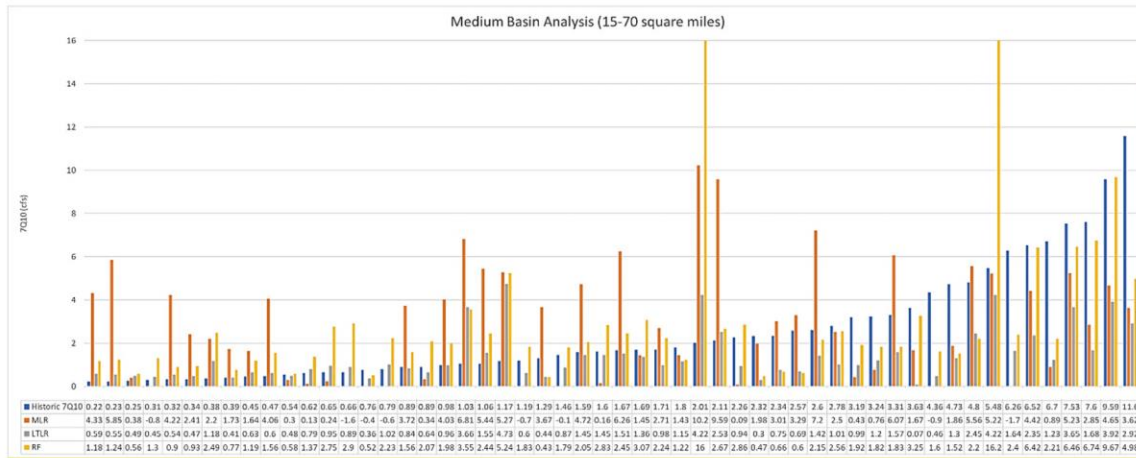
Table 9: RMSE comparisons between methods for specified size ranges.

Subset	RMSE for Each Methodology				
	MLR	LTLR	RF	NN	GAM
Small Basins (<15 mi ²)	2.11	0.34	0.44	2.77	2.97
Medium Basins (15–70 mi ²)	3.96	2.83	3.09	3.87	4.66
Large Basins (>70 mi ²)	14.02	26.23	14.01	15.60	20.32
Average	6.70	9.80	5.85	7.41	9.31

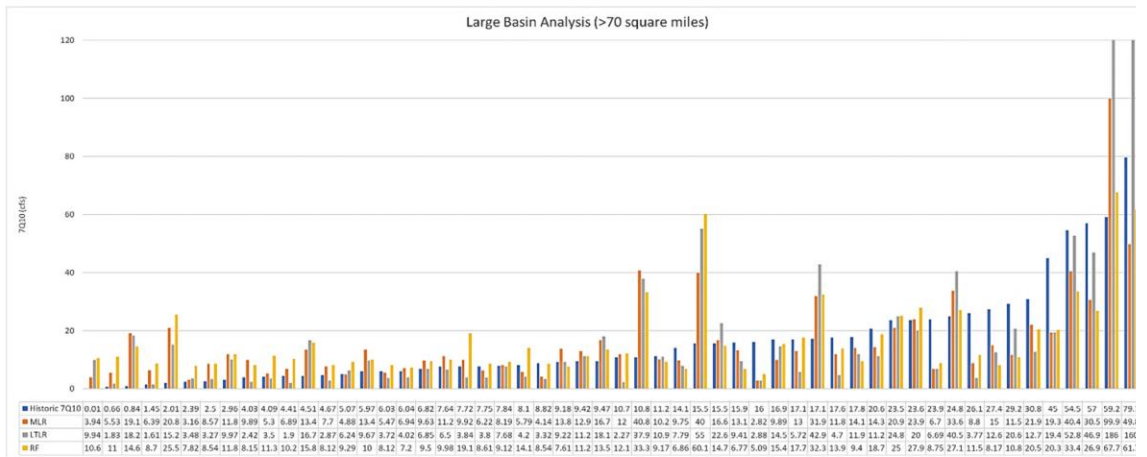
Table 9 demonstrates that LTLR and RF perform similarly well for 7Q10 estimation in both small- and medium-sized basins, greatly outperforming MLR, NN, and GAM. However, for large basins, LTLR performs poorly because of the extreme overestimation discussed earlier, while MLR, RF, and NN perform similarly well (RMSE = 14.02, 14.01, and 15.60, respectively). Based on this experiment, RF performs the best across all ranges of basin sizes (Avg. RMSE = 5.85), while LTLR performs similarly well in small- and medium-sized basins where it is traditionally used, and MLR performs similarly well in large basins. Based on RF’s success for all ranges of basin sizes, we include Figure 18 to compare raw 7Q10 estimates for the RF method to both classical methods, MLR and LTLR. Figure 18a–c displays the actual 7Q10, as well as the LTLR, MLR, and RF estimates, arranged from the smallest historical 7Q10 to the largest.



(a)



(b)



(c)

Figure 18 (a–c): Raw 7Q10 estimates using each methodology.

The results in Figure 18a–c corroborate the results from Table 9. Logarithmic-transformed linear regression performs best overall in terms of R^2 and NSE but performs poorly in KGE and RMSE due to its poor estimates in large basins, reflected in Figure 18c. LTLR estimates the largest historical 7Q10, which is 90 cfs, to be 181 cfs, more than double the actual value. Similarly, LTLR estimates the next two largest 7Q10s, which are 59.2 cfs and 79.7 cfs, to be 186 cfs and 160 cfs, respectively. Though the LTLR vs. actual 7Q10 graph previously suggested that LTLR may be able to be expanded for large basins because the largest 7Q10s seemed to maintain a constant distance from the perfect fit line, Figure 9 is in log space, and differences in log space are amplified for larger numbers when returning to standard space. Conversely, multiple linear regression performs well in large basins but poorly in small basins, as it attempts to minimize the overall error and gives more weight to larger 7Q10s. Lastly, the random forest method performs well overall, especially in larger basins, suggesting that this flexible machine learning algorithm may be able to account for the drawbacks of both LTLR and MLR.

2.6 Conclusions

This research improves upon current methodologies for statistically estimating the 7Q10 by analyzing multiple statistical methods, testing various topographical, landcover, and climate variables for significance, and widening the geographical and watershed size ranges of current methodologies. Results support that a single, generalized methodology can be used for 7Q10 estimation throughout the entire northeast and mid-Atlantic, with similar R^2 , RMSE, NSE, and KGE compared to current state-by-state StreamStats' estimates while only requiring one equation/model. Estimates from StreamStats display an R^2 , KGE, NSE, and RMSE of 0.66, 0.66, 0.65, and 9.88, respectively, for the

unimpaired gages where it is available in the study area, while the random forest method displays an R^2 , KGE, NSE, and RMSE of 0.61, 0.69, 0.41, and 8.39, respectively, for the full range of gages even after cross-validation was applied. Two other machine learning algorithms (neural networks and generalized additive models) were tested as well but displayed significantly worse R^2 s, KGEs, NSEs, and RMSEs (0.53, 0.69, 0.36, and 9.41 and 0.53, 0.65, 0.52, and 12.15, respectively) with the addition of the cross-validation test set.

Future work could involve testing other advanced statistical methods and/or machine learning algorithms for better low-flow estimation. Because we were able to successfully apply this methodology to such a large geographical footprint, other future work may include determining the boundaries at which assuming hydrologic homogeneity is no longer satisfied. Additional future work directly related to this study may involve landcover and climate-altered futures. The inclusion of climate and landcover input variables, which were both found to be statistically significant, can be used prescriptively in conjunction with physical hydrology models to test how changing landcover and climate conditions affect 7Q10 estimates. Because of additional stakeholder input, future work may also involve only using 7Q10 data derived from the last 30 years of streamflow data at each site for use as the actual historical 7Q10. Blum et al. (2019) found that using the most recent 30 years of streamflow record to derive the “true” 7Q10 when a trend is detected reduces error and bias in 7Q10 estimators compared to using the full record of streamflow [2]. This may account for recent climatic and hydrologic conditions that will be more representative of future 7Q10 conditions at a particular site, but additional studies must be completed to confirm this relationship.

CHAPTER 3

Fuzzy C-Means Clustering for Physical Model Calibration and 7-Day, 10-Year Low Flow Estimation in Ungaged Basins: Comparisons to Traditional, Statistical Estimates

In the northeast U.S., resource managers commonly apply 7-day, 10-year (7Q10) low flow estimates for protecting aquatic species in streams. In this paper, we evaluate the efficacy of process-based hydrology models for estimating 7Q10s compared to the United States Geological Survey's (USGS) widely applied web-application StreamStats, which uses traditional statistical regression equations for estimating extreme flows. To generate the process-based estimates, the USGS's National Hydrologic Modeling (NHM-PRMS) framework is applied with 36 years of forcings from the Daymet climate dataset to a representative sample of ninety-four unimpaired gages in the Northeast and Mid-Atlantic U.S. The process-based model is calibrated to the measured streamflow at each gage using the recommended NHM-PRMS calibration procedure and evaluated using Kling-Gupta Efficiency (KGE) for daily streamflow estimation. To evaluate the 7Q10 estimates made by the process-based model and StreamStats, a multitude of error metrics are applied, including median relative bias (cfs/cfs), Root Mean Square Error (RMSE) (cfs), Relative RMSE (RRMSE) (cfs/cfs), and Unit-Area RMSE (UA-RMSE) (cfs/mi²). The calibrated process-based model displays improved daily streamflow estimation (median KGE improving from 0.30 to 0.52) and 7Q10 estimation (smaller median relative bias, RMSE, RRMSE, and UA-RMSE, especially for basins larger than 100mi²). We extend the success of calibration to ungaged locations using the machine learning algorithm Fuzzy C-Means (FCM) clustering, finding that traditional K-Means clustering (FCM clustering with no fuzzification factor) is the preferred method for model regionalization

based on 1) Silhouette Analysis, 2) daily streamflow KGE, and 3) 7Q10 error metrics. The optimal model created with clustering shows improvement for daily streamflow estimation (a median KGE of 0.48, only slightly below that of the calibrated models at 0.52); however, this model displays similar error metrics for 7Q10 estimation compared to the uncalibrated models, neither of which provide improved error metrics compared to the statistical estimates. Results suggest that the process-based model calibrated to measured streamflow data provides the best 7Q10 estimation in terms of all error metrics except median relative bias, but for all models applicable to ungaged locations, the statistical estimates from StreamStats display the lowest error metrics in every category.

3.1 Background

This study's objective is to test whether a regionally calibrated, process-based hydrology model can provide better estimates of 7Q10 flows than common statistical methods. Future updates to the National Water Model (NWM) and National Hydrologic Model (NHM) will make it possible to quickly create uncalibrated rainfall-runoff models at virtually any location on a stream in the U.S. Application of these models may prove to be very attractive to individuals seeking process-based models that incorporate rainfall-runoff modeling for low flow estimation. However, there is a paucity of research on how these models perform for specific use-cases like low-flow and/or 7Q10 estimation. Some studies have demonstrated success in using statistically simulated at-site streamflow to calibrate rainfall-runoff models for daily streamflow estimation, but the authors note that their initial results suggest these models "may not reproduce both low and high streamflow magnitudes" (Farmer et al., 2019). In Golian et al. (2021), the authors note that machine learning-based regionalization was least satisfactory for low flows when

compared to both average and high flows (Golian et al., 2021). In this paper, we study this further by analyzing the ability of the physical models to estimate both daily streamflows and 7Q10s. For this analysis, we extract 94 process-based hydrology models from the USGS' National Hydrologic Modeling network at unimpaired, gaged locations in the Northeast and Mid-Atlantic United States, and test how the 7Q10 estimates from these models compare to regression-based 7Q10 estimates for the same locations. We then calibrate these models to the measured streamflow at each location using the USGS' auto-calibration software LUCA (Hay and Umemoto, 2007). To extend calibration to ungaged locations without measured streamflow data, we use the adaptive machine learning algorithm Fuzzy C-Means clustering (Dunn, 1973) for parameter regionalization to re-calibrate the models at each location in lieu of any known measured streamflow. We then evaluate each model's ability to estimate daily streamflow and 7Q10. This process is summarized in Figure 19.

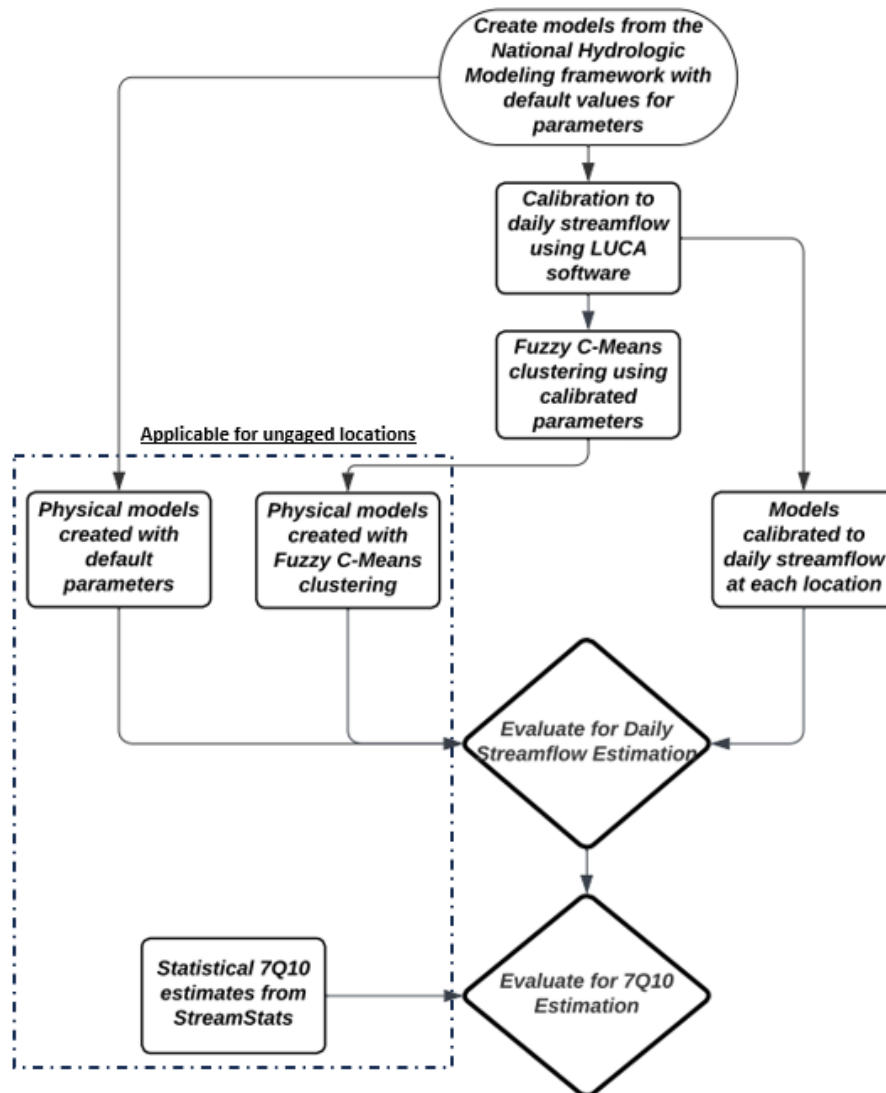


Figure 19: Summary of this study's experimental design.

The results answer the following questions:

- 1) Do uncalibrated hydrology models, created using extractions from the USGS' National Hydrologic Modeling framework, perform comparably to regression-based 7Q10 estimation?
- 2) Can models calibrated using Fuzzy C-Means clustering provide improved 7Q10 estimation compared to a simple regression model?

3.2 Data and Study Area

The following sections describe the study area and data used in this research.

3.2.1 Study Area

The study area for this analysis is the northeast United States, including the states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, Pennsylvania, New Jersey, Delaware, Maryland, Virginia, and West Virginia. This area is roughly 260,000 square miles and is not homogenous, as it covers two distinct Hydrologic Unit Code (HUC) regions of the U.S. (Seaber, 1987). Basins selected for this study have been defined as “unimpaired” in USGS’s Hydro-Climatic Data Network, HCDN-2009 (Lins et al., 2012). This set of stream gages includes 94 watersheds of varying size and physical attributes. These basins range from 2.1 mi² to 1419 mi² (Figure 20).

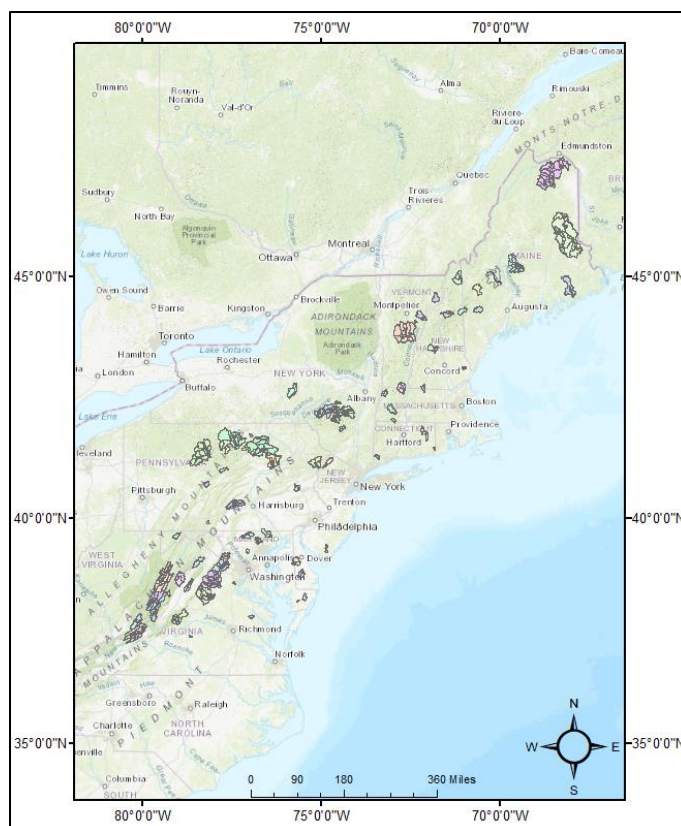


Figure 20: 94 unimpaired gaged basins in the Northeast United States.

3.2.2 Data

In this subsection, we provide all data used in this research.

3.2.2.1 Streamflow and 7Q10 Data

Streamflow data from these 94 gages were downloaded from the USGS's Current Water Data for the Nation (<https://waterdata.usgs.gov/nwis/rt>). For this experiment, the full record of streamflow was used for calculation of the 7Q10 at each site. The “fasstr” software package (<https://cran.r-project.org/web/packages/fasstr/index.html>) was used to calculate the 7Q10 directly from the daily streamflow data. This package applies a quantile distribution to daily streamflow data allowing for the efficient calculation of low flow frequency analysis metrics, including the 7Q10. These 7Q10 values were identical

to the 7Q10 values calculated by the USGS at each site, presented on the USGS's StreamStats Data-Collection Station reports (<https://streamstatsags.cr.usgs.gov/>).

3.2.2.2 Physical Hydrology Models

Physical hydrology models for each of the 94 gaged locations were extracted from the USGS's National Hydrologic Model version of the Precipitation Runoff Modeling System (NHM-PRMS). The USGS National Hydrologic Model (NHM) infrastructure was developed to support the efficient creation of local, regional, and national-scale hydrologic models for the United States (Regan et al., 2019). These models incorporate data stored in the NHM, including the basin and subbasin landcover values and area-weighted average climate forcings required to run PRMS. Selecting a location or gage that is a point-of-interest in the NHM generates a ready-to-run physical model at that location, with all necessary variables being extracted for the basin of interest, including land data (area, elevation, landcover) and the corresponding climate data. Climate forcing dataset choices include Daymet (1980-2016) (Thornton et al., 2016), Maurer (1949-2010) (Maurer et al., 2002), and Livneh (1915-2015) (Livneh et al. 2019) in the form of basin area-weighted precipitation and temperature timeseries. For this analysis, we selected the Daymet climate dataset because it offers the finest resolution of the three (1km, as opposed to 6km for Livneh and ~12km for Maurer) and allows for trend analysis, as there are no temporal discontinuities.

Additionally, the NHM-PRMS makes several major assumptions to model hydrologic processes. This includes:

1. Dividing basins into subbasins using pre-determined Hydrologic Response Units (HRUs) from the USGS' Geospatial Fabric (Bock et al., 2020).

2. Using a daily time-step, while some models utilize finer timesteps.
3. Calculating evapotranspiration using the Jensen-Haise (JH) formulation (Jensen and Haise, 1963).

The assumptions above are specific to the NHM-PRMS but should not imply that the results from this study will be specific to this hydrologic modeling software. The HRUs from the NHM-PRMS are based on pre-determined areas of homogeneity. Though some other models utilize gridded landcover data, all of the gridded values that would fall within an HRU should be similar to the value used for that HRU from the NHM-PRMS. Using a daily time-step may provide inaccurate high-flow estimates, as things like the 100-year-flood are calculated using gage data on a 15-minute scale when it is available (England et al., 2019), but the 7Q10 is always calculated from daily average streamflows. Calculating evapotranspiration using the JH formulation may have an impact on results, but the appropriate steps have been taken to minimize the impact. JH calculates evapotranspiration based on temperature for each HRU. In this study, we have used the Daymet climate dataset, which provides the finest resolution of the available climate datasets with no discontinuities. Additionally, we attempt to minimize the impact of the JH formulation by extensive model calibration. JH is calibrated during the standard calibration procedure, which will be further described in a later section.

3.2.2.3 StreamStats 7Q10 Estimates

To compare 7Q10 estimates from the physical hydrology models to current statistical estimates, we use the USGS's statistical estimation program StreamStats. This program uses multiple linear regression equations, derived in log-space, to directly estimate flow statistics (without estimating daily streamflow) (Ries et al., 2000). StreamStats is widely

utilized by resource managers in the study area and provides direct comparisons. Though the input variables vary by state, the typical process is as follows:

1. Calculate the historic 7Q10 at various gaged locations in a homogenous hydrologic area.
2. Collect the physical characteristics (watershed area, elevation, slope, etc.) for each of the watersheds attributed to the gages used above.
3. Fit a multiple linear regression, in log-space, to relate the input variables (watershed area, elevation, slope, etc.) to the corresponding 7Q10 value.
4. Delineate the watershed that is attributed to the ungaged location of interest.
5. Calculate the physical characteristics of the delineated watershed.
6. Apply the physical characteristics from the ungaged, delineated watershed to the regression equation developed in step 3 to calculate the 7Q10.

StreamStats uses varying regression equations and explanatory variables for Massachusetts (Ries et al., 2000), Rhode Island (Bent et al., 2014), New Hampshire (Flynn et al., 2002), Maine (Dudley et al., 2004), Pennsylvania (Stuckey et al., 2006), Virginia (Austin et al., 2011), and West Virginia (Wiley, 2008). StreamStats 7Q10 has not been developed in Connecticut, Delaware, Maryland, New Jersey, New York, and Vermont, which partially limits the comparisons.

3.3 Methodology

In the following section, we describe in detail all methods used in this research. This includes calibration of the NHM-PRMS models, Fuzzy C-Means clustering for regionalization, Silhouette Analysis for evaluating the optimal number of clusters, and the evaluation metrics used for both the daily streamflow models and 7Q10 estimates.

3.3.1 Calibration of the NHM-PRMS Models

Models of the 94 basins were calibrated using the procedure recommended in the PRMS IV Manual (Markstrom et al., 2015) by executing the USGS’s automated calibration software LUCA (Hay et al., 2007). This procedure applies a multi-objective, multi-step process of continuously revising sub-basin parameters. To achieve calibration, parameters are varied individually for each of the 94 locations, with the objective of minimizing the difference between the simulated daily streamflow and the measured daily streamflow at each gage. The parameters recommended for calibration in PRMS are summarized in Table 10, along with their default values and recommended calibration bounds (Markstrom et al., 2015).

Table 10: Parameters calibrated in the physical hydrology model

Variable	Description	Units	Default Value	Lower Bound	Upper Bound
dday_intcp	Monthly (January to December) intercept in degree-day equation	dday	-40	-70	10
dday_slope	Monthly (January to December) slope in degree-day equation	dday/temp units	0.4	0.2	1.1
tmax_index	Monthly (January to December) index temperature used to determine precipitation adjustments to solar radiation	temperature units	50	50	90
jh_coef	Monthly (January to December) air temperature coefficient used in Jensen Haise potential ET computations	Per degree temperature	0.014	0.005	0.10
emis_noppt	Average emissivity of air on days without precipitation.	decimal fraction	0.757	0.757	1
fastcoef_lin	Linear coefficient in equation to route preferential-flow storage down slope.	fraction/day	0.1	0.0001	1
fastcoef_sq	Non-linear coefficient in equation to route preferential-flow storage down slope.	none	0.8	0.00001	1
freeh2o_cap	Free-water holding capacity of snowpack expressed as a decimal fraction of the frozen water content of the snowpack.	inches	0.08	0.01	0.2
gwflow_coef	Linear coefficient in the equation to compute groundwater discharge for each GWR.	fraction/day	0.03	0.0005	0.10

gwstor_init	Storage in each GWR at the beginning of a simulation.	inches	1	0.01	20.0
potet_sublim	Fraction of potential ET that is sublimated from snow in the canopy and snowpack.	decimal fraction	0.5	0.1	0.75
smidx_coef	Coefficient in non-linear contributing area algorithm.	decimal fraction	0.001	0.0001	1
smidx_exp	Exponent in non-linear contributing area algorithm.	1/inch	1	0.2	1.8
soil_moist_max	Maximum available water holding capacity of capillary reservoir from land surface to rooting depth of the major vegetation type.	inches	5	0	20
soil_rechr_max_frac	Maximum storage for soil recharge zone (upper portion of capillary reservoir where losses occur as both evaporation and transpiration).	decimal fraction	0.5	0	1
soil2gw_max	Maximum amount of the capillary reservoir excess that is routed directly to the groundwater recharge	inches	0.1	0	0.5
rain_cbh_adj	Monthly (January to December) adjustment factor to measured precipitation to account for deficiencies in gage catch.	decimal fraction	1	0.01	2
snow_cbh_adj	Monthly (January to December) adjustment factor to measured precipitation to account for deficiencies in gage catch.	decimal fraction	1	0.01	2
adjmix_rain	Monthly (January to December) factor to adjust rain proportion in a mixed rain/snow event by month.	decimal fraction	1	0.01	1.4
cecn_coef	Monthly (January to December) convection condensation energy coefficient.	Calories/deg Celsius	5	0.01	20
tmax_allrain_offset	Monthly (January to December) maximum air temperature when precipitation is assumed to be rain; if HRU air temperature is greater than or equal to tmax_allsnow plus this value, precipitation is rain.	temperature units	5	0	10
tmax_allsnow	Monthly (January to December) maximum air temperature when precipitation is assumed to be snow; if HRU air temperature is less than or equal to this value, precipitation is snow.	temperature units	30	20	40

Each parameter that is to be calibrated begins with the default value and is continually refined during the calibration process. During calibration, the parameters are constrained

to lie within the process-driven limits given above. In preparation for clustering, the updated value of the parameters are normalized using standard min-max normalization:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (8)$$

Figure 21 displays the range of parameters values after calibration and normalization. Normalization causes all the parameters to share the same range of 0 to 1. These values can easily be returned to their actual values by using the minimum and maximum values given below each boxplot in Figure 21 and reversing the equation above to solve for x .

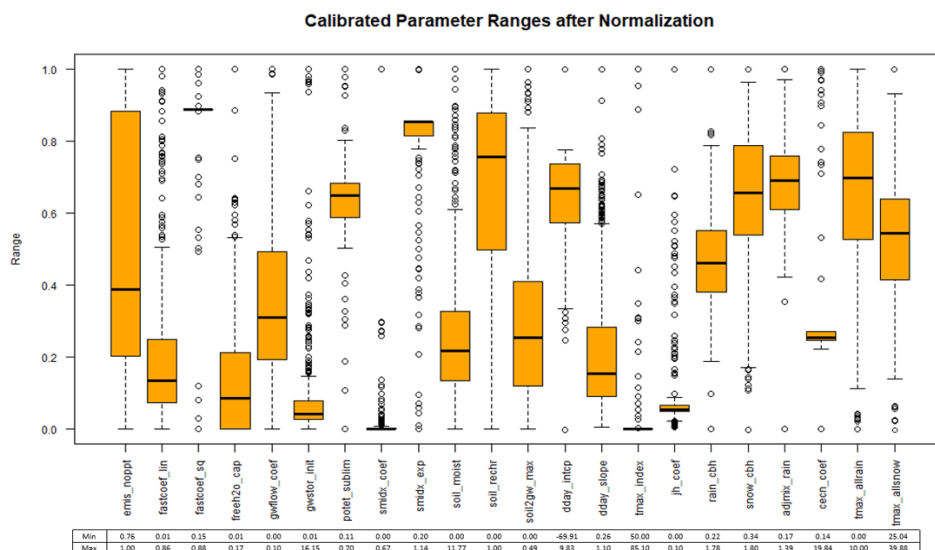


Figure 21: Parameter ranges after calibration and min-max normalization.

Initially, we tested calibration of the physical models to minimize the error for low-flow estimation rather than for daily streamflows, deviating from the recommended calibration procedure in the PRMS IV manual (Markstrom et al., 2015), as the goal of this experiment is to test hydrologic models for low flow estimation. Specifically, the LUCA software allows for calibration to the lowest annual daily streamflows. Initial results suggested that the physical models calibrated to low flows were able to estimate the magnitude of 7Q10s well, but a more thorough analysis of the hydrograph suggested that

the models were not properly maintaining streamflows throughout the rest of the year.

This discrepancy is highlighted in Figure 22, which highlights gage 01552000 in Pennsylvania on Loyalsock Creek as an example.

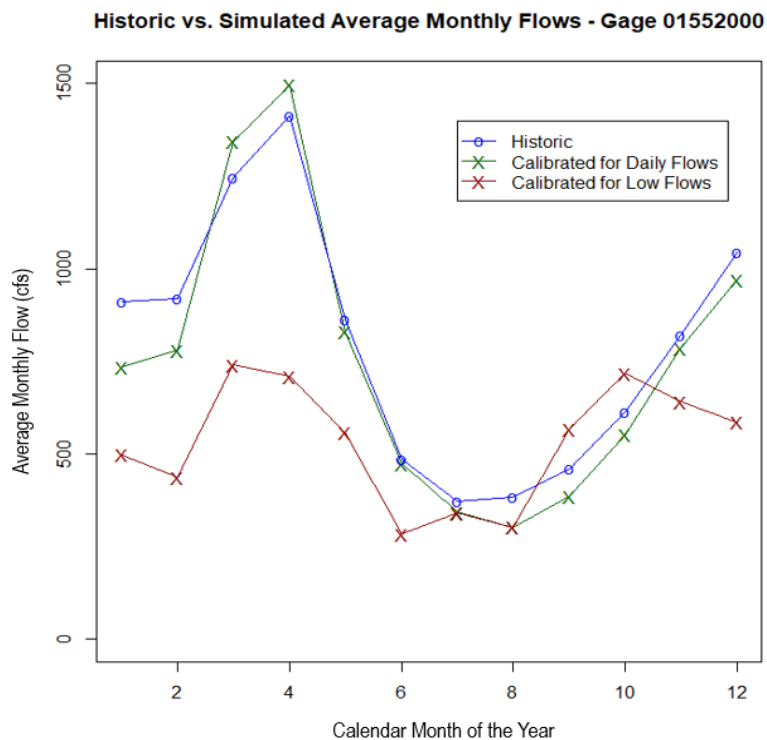


Figure 22: Example hydrograph for gage 01552000 that demonstrates the issue with calibrating daily rainfall-runoff models for low-flows. Monthly averages are calculated for 1980-2016, the full length of model runs.

Figure 22 demonstrates that the model calibrated for low flows is able to estimate flows during the two months with the lowest average historic flows (months 7 and 8, being July and August) just as well as the model calibrated for daily flows, but for all other months, the low-flow calibrated model provides streamflow estimates that do not mimic the historic hydrograph. Further analysis demonstrated that calibration parameters with individual monthly values (dday_intcp, dday_slope, tmax_index, jh_coef, rain_cbh_adj, snow_cbh_adj, adjmix_rain, cecn_coef, tmax_allrain_offset, and tmax_allsnow) were only changing for months where the lowest flows occur. For example, the rain adjustment

factor (rain_cbh_adj) finished calibration with the default parameters for all months except months 7 and 8, causing all other months to remain unchanged. The calibration process is used to continually revise parameters while maintaining the water budget for daily, monthly, and yearly timesteps, and the models calibrated to low flows did not attempt to maintain yearly water budget totals or daily flows in months where low flows do not occur.

Figure 22 also demonstrates that many models calibrated to low flows displayed relatively low flows occurring in February, where the historic measured flows were substantially higher than both summer (JJA) and early fall (SO) flows. An analysis of the measured streamflow data at each site showed that winter (DJF) low flows do typically occur in February, but at much higher magnitudes than JJA and SO flows. Further analysis supported this, showing that annual 7-day low flows in the study area typically range from July-October, with none occurring in February. In addition, the 10th percentile 7-day annual low flow (the 7Q10 calculated without fitting a distribution) has historically occurred in September for all but two gages. This is displayed in Figure 23a and b.

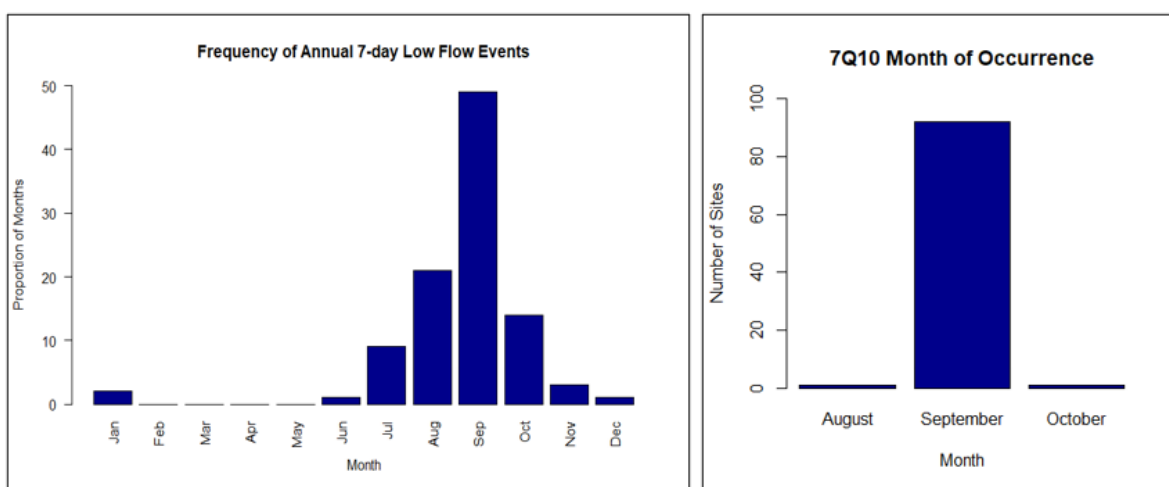


Figure 23 (a-b): Month of Historic Annual 7-Day Low Flows (a, left) and Corresponding 7-Day-10-Year Low Flow Months of Occurrence (b, right)

3.3.2 Fuzzy-C Means Clustering

Traditional calibration of a rainfall-runoff model is only possible at gaged locations. Furthermore, at these locations, 7Q10s can be directly calculated from the measured streamflow data. However, to extend calibration to ungaged locations, hydrologic regionalization can be used. Fuzzy C-Means (FCM) clustering is a clustering algorithm that utilizes soft assignments of data points to clusters (Dunn, 1973). Unlike traditional clustering algorithms, like K-means clustering (Macqueen, 1967) that create hard assignments for each data point to a single cluster, FCM assigns membership values to indicate the degree of “belongingness” of data points to each cluster. The objective function seeks to find the optimal cluster centers and membership values that minimize the overall fuzziness or uncertainty of the clustering result. The FCM algorithm provides greater flexibility in clustering tasks, as it can handle scenarios where data points may partially belong to multiple clusters or where cluster boundaries are ambiguous. This is advantageous in this application, as this methodology is tested for a large geographical area that includes two pre-determined HUC regions of the United States. By assigning membership values, FCM provides a more nuanced representation of the clustering structure and allows for capturing overlapping clusters or gradual transitions between clusters. We use these membership values to create a weighted average of the parameters to use in the physical models. This is illustrated in Figure 24.

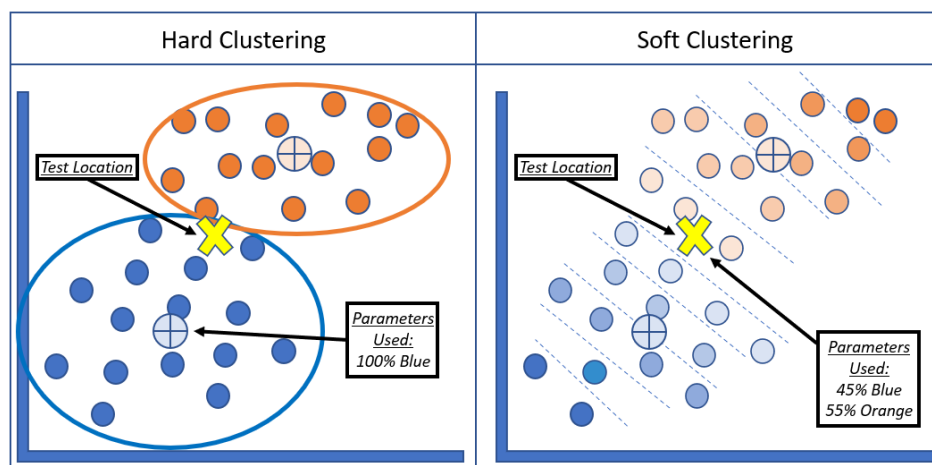


Figure 24: Difference between hard clustering algorithms (e.g. K-Means) and soft clustering algorithms (e.g. Fuzzy-C Means).

For implementation, a value “m” is used to set the “fuzzification” factor, dictating how much overlap to allow between clusters. As m is increased, the allowed overlap between clusters is increased. We execute the FCM algorithm for the total number of clusters possible with a specified range of m-values. The range of clusters possible for FCM is integers between $[2, N/2]$ (for this case between 2 and 262), as there are 525 individual sub-basins with their own set of parameters. The range of m values possible is $[1.5, \infty)$. We will use m values of $[1.5, 5]$ with increasing steps of 0.10. This will test 36 different m values for each cluster, leading to a total number of 9,396 possible cluster and fuzziness combinations for this experiment. Note that when $m = 1$, there is no overlap allowed between clusters and FCM reduces to K-Means clustering.

Clustering algorithms are typically used descriptively to highlight patterns in a dataset, but they can be used prescriptively given a set of predictor variables and response variables. For this study, the response variables for the FCM clustering are the calibrated parameters given in Table 10, as that is what must be predicted for calibrating the models at unengaged locations. The predictor variables, which will be used to predict which set of

calibrated parameters to use, are the publicly available physical parameters for each sub-basin from the NHM, all designated as numeric values. These are highlighted in Table 11.

Table 11: Predictor variables used in the Fuzzy C-Means analysis from the 94 test basins

Model Parameter	Description
Area	Area of the basin
Elevation	Mean elevation of the basin
Percent Impervious	Percent of the basin considered to be impervious
Slope	Average slope of the basin
Land Type	Primary land designation (land, lake, swale)
Soil Type	Soil type of the basin
Vegetation Type	Type of vegetation primarily covering the area

Given the above parameters, the process for creating the new models using fuzzy c-means is summarized below.

1. Calibration creates response variables for each location.
2. Predictor variables are extracted from each location.
3. All parameters are normalized using min-max normalization.
4. Clustering is used to create groups of similar basin parameters.
5. Fuzzy c-means clustering is used to evaluate each location's membership to each group.
6. New calibrated parameters are created using a weighted average of each location's membership to each cluster and the cluster's corresponding centroid.

3.3.3 Silhouette Analysis

Silhouette analysis (Rousseeuw, 1987) calculates the optimal number of clusters. This methodology evaluates the quality of clustering by assessing the separation and cohesion of clusters, as well as the fit of data points within their assigned clusters (Rousseeuw, 1987). First, a silhouette coefficient is calculated for each data point that measures how

well it fits within its cluster compared to neighboring clusters. Next, the average silhouette coefficient is computed across all data points for each value of the number of clusters. Finally, the optimal number of clusters is determined by selecting the value that maximizes the average silhouette coefficient, indicating the presence of well-separated and compact clusters.

Models created from the parameter clusters are executed with the four highest average silhouette coefficients. Because parameters are regionalized for use in a physical hydrology model, there may be some slight variations between the cluster/m-value combination with the best average silhouette coefficient and the physical hydrology model with the optimal daily streamflow. By using multiple models, the link between the optimal silhouette coefficients and optimal physical model performances is verified, as well as ensuring that the single model with the best daily streamflow and 7Q10 estimation is identified. If the silhouette coefficients suggest that there are distinct clusters, which would occur if the highest average silhouette coefficients occur when $m = 1.5$ and decrease as m is incrementally increased, we will also consider using models with $m = 1$, which would reduce the FCM to K-Means clustering.

3.3.4 Evaluation Metrics for Daily Streamflow and 7Q10 Estimates

In this experiment, the Kling Gupta Efficiency (KGE) (Gupta et al., 2009) is used to evaluate the daily streamflow models. KGE is widely used for hydrologic applications (Formetta et al. 2011; Beck et al. 2016) because it incorporates three components into its definition: the Pearson's correlation coefficient (r), the bias (β), and the error variability (α). KGE is calculated using the following formula:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (9)$$

The goal of this experiment is to evaluate the models for 7Q10 estimation, so we will use four error metrics to evaluate the errors for 7Q10 estimation. These are Relative Bias,

$$Relative\ Bias = \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (10)$$

Root Mean Square Error (RMSE),

$$RMSE = \sqrt{\frac{1}{n} * \sum_1^n (y_i - \hat{y}_i)^2} \quad (11)$$

Relative Root Mean Square Error (R-RMSE),

$$R - RMSE = \frac{\sqrt{\frac{1}{n} * \sum_1^n (y_i - \hat{y}_i)^2}}{\bar{y}} \quad (12)$$

and Unit-Area RMSE (UA-RMSE),

$$UA - RMSE = \sqrt{\frac{1}{n} * \sum_1^n \left(\frac{y_i}{DA} - \frac{\hat{y}_i}{DA} \right)^2} \quad (13)$$

where y_i is the observed 7Q10, \hat{y} is the model predicted 7Q10, n is the total number of sites, DA is the drainage area, and \bar{y} is the associated mean value. Similar studies have also chosen RMSE over MAE for its sensitivity to outliers (Ferreira et al., 2021; Mekanik et al., 2016).

3.4 Results and Discussion

The following sections discuss the uncalibrated model vs calibrated model performance for daily streamflow estimation and 7Q10 estimation, the results from the silhouette analysis, applying the optimal silhouettes with fuzzy C-Means regionalization for daily streamflow estimation, and the results of using the models calibrated using clustering for 7Q10 estimation.

3.4.1 Uncalibrated Model Performance vs Calibrated Model Performance

In Figure 25, we present the general results for the uncalibrated models from the NHM-PRMS compared to the calibrated models.

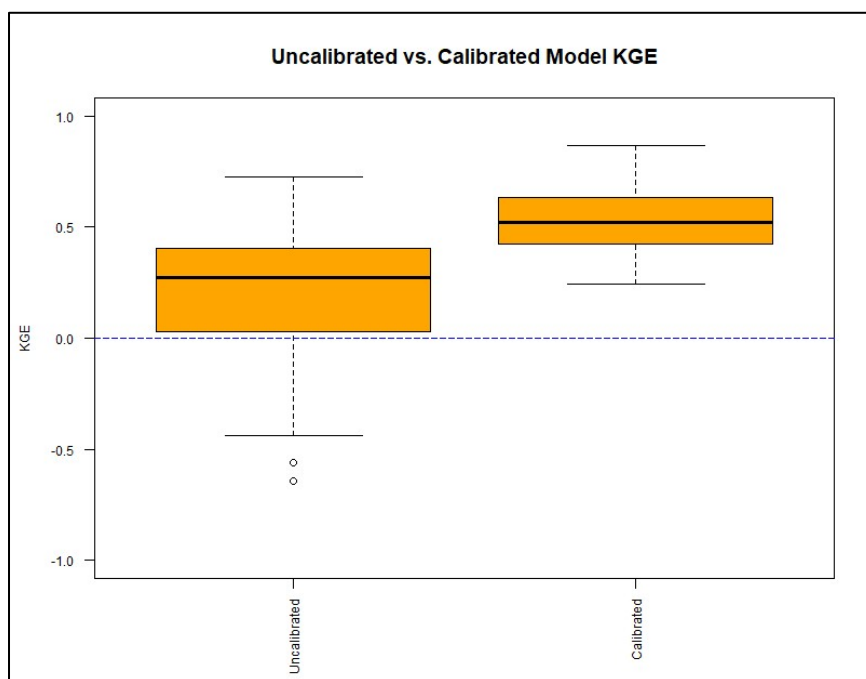


Figure 25: Results from calibration using KGE to evaluate daily streamflow estimation. As expected, calibration improved daily streamflow estimation for every basin. Before calibration, the median KGE was 0.30 with some locations having negative KGE values. Calibration improved the median daily streamflow to 0.52, with the lowest KGE value being 0.25. We will also further analyze the models by spatially disaggregating the basins into small (<100 mi²) and large basins (>100 mi²). This threshold is chosen because many regression equations used for 7Q10 estimation, including some of the StreamStats' equations in the study area, are only recommended for basins up to 100-150 mi² (e.g. Ries et al., 2000). Figure 26a and b display the physical model KGEs, but this time split by small basins (a) and large basins (b).

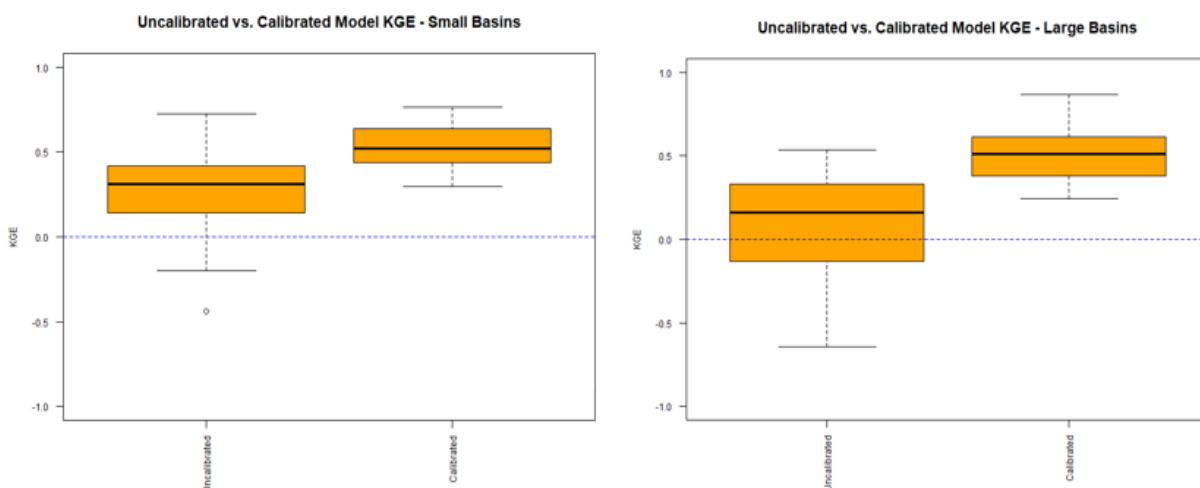


Figure 26a-b: Results from calibration for basins smaller than 100 mi² (a, left) and basins larger than 100 mi² (b, right) using KGE to evaluate daily streamflow estimation.

There are minimal differences between the calibrated models for small and large basins, but the uncalibrated models display noticeable differences. The uncalibrated models for larger basins display an inferior median KGE, more KGE values that are negative, and a much wider interquartile range with a lower 25th percentile that is negative. The uncalibrated models for the larger basins perform significantly worse than the uncalibrated models for the smaller basins in terms of KGE. Given that the calibrated models for the large basins perform similarly to the calibrated models for the small basins, this suggests that the default parameter values are not as appropriate for larger basins, requiring calibration more than that of the models for small basins.

3.4.2 Uncalibrated vs. Calibrated Models for Gaged 7Q10 Estimation

This experiment's primary goal is to test the hypothesis that models can be used for 7Q10 estimation in ungaged locations. Of the 94 basins in the study, an estimate of the 7Q10 is not available through StreamStats for 28 of the basins (StreamStats 7Q10 estimation has not been developed by the USGS for certain states, as discussed previously). These locations were removed from the analysis for the analyses presented for 7Q10 estimation.

Table 12 summarizes the Median Relative Bias, RMSE, RRMSE, and UA-RMSE of the uncalibrated and calibrated models for 7Q10 estimation compared to StreamStats.

Table 12: Error metrics for 7Q10 estimation (only StreamStats locations)

7Q10 Estimation Technique	Median Relative Bias (cfs/cfs)	RMSE (cfs)	RRMSE (cfs/cfs)	UA-RMSE (cfs/mi²)
Uncalibrated Hydrology Models	0.81	13.99	116.48	0.08
Calibrated Hydrology Models	0.61	12.33	102.63	0.07
StreamStats	0.42	13.75	114.5	0.05

For the 66 basins where StreamStats 7Q10 estimation is available, the results suggest that StreamStats provides significantly lower median relative bias and UA-RMSE to the uncalibrated models, but similar RMSE and RRMSE. The calibrated models perform best in terms of RMSE and RRMSE but provide significantly larger median relative bias and UA-RMSE than StreamStats. RMSE is heavily influenced by larger values, while UA-RMSE attempts to weigh smaller and larger values equally by scaling the larger values down based on their larger watershed areas. Because the calibrated models perform best for RMSE but not for UA-RMSE (which is influenced significantly less than RMSE by larger values) suggests that the calibrated models perform well for larger basins. Table 13 confirms this by displaying the same metrics characterized by small and large basins.

Table 13: Error metrics for 7Q10 estimation (only StreamStats locations)

7Q10 Estimation Technique	Median Relative Bias (cfs/cfs)	RMSE (cfs)	RRMSE (cfs/cfs)	UA-RMSE (cfs/mi²)
Basins smaller than 100mi ² - 35 Locations				
Uncalibrated Hydrology Models	0.90	7.26	188.82	0.09
Calibrated Hydrology Models	0.90	5.47	142.24	0.08

StreamStats	0.42	2.23	58.00	0.04
Basins larger than 100mi ² - 31 Locations				
Uncalibrated Hydrology Models	0.65	19.57	92.17	0.07
Calibrated Hydrology Models	0.42	16.25	76.53	0.06
StreamStats	0.32	19.92	93.86	0.06

Table 13 suggests that StreamStats outperforms both the uncalibrated and calibrated physical models for 7Q10 estimation in small basins. For the small basins, StreamStats' median relative bias, RMSE, RRMSE, and UA-RMSE are all half that (or more) of the calibrated models. For the large basins, StreamStats still performs best in terms of median relative bias, but the calibrated models display the same UA-RMSE and better RMSE and RRMSE than StreamStats. Given the results presented here, the calibrated physical models at gaged locations over 100 square miles can provide similar errors for 7Q10 estimation compared to StreamStats in terms of the error metrics presented, but StreamStats is primarily for ungaged basin estimation, and this calibration methodology cannot be used for ungaged locations. In the next section, we provide a methodology for calibrating models in ungaged locations using Fuzzy C-Means clustering for hydrologic regionalization.

3.4.3 Results from Silhouette Analysis with Fuzzy C-Means Clustering

We will use silhouettes to find the optimal FCM parameters. We will apply the full range of clusters possible combined with m-values ranging from 1.5 to 5 using steps of 0.1. Figure 27 displays the average silhouette widths for each combination of clusters and m-values.

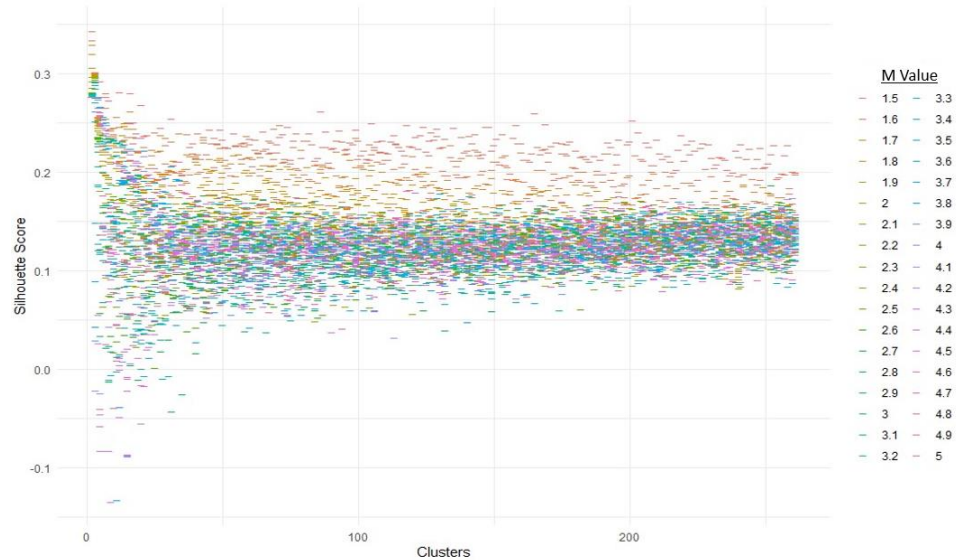


Figure 27: Silhouette Analysis for the Fuzzy C-Means clustering.

In addition, we include the top 10 average silhouette coefficients arranged from largest to smallest in Table 14.

Table 14: Highest 10 average silhouette coefficients

Number of Clusters (c)	M-Value (m)	Average Silhouette Coefficient
2	1.5	0.344265
2	1.6	0.334988
2	1.7	0.33009
2	1.8	0.32043
2	1.9	0.306479
3	1.5	0.302499
3	1.6	0.301588
3	1.8	0.301469
3	1.7	0.301405
3	1.9	0.300441
3	2	0.29949

Results suggest that the cluster and m-value combination that optimizes the average silhouette coefficient is the minimum value for both clusters and m. The best average silhouette coefficient was found to occur when c=2 clusters with an m-value of 1.5. The next 4 best silhouette coefficients remain when c=2 but decrease slightly as m is

increased by 0.10 each step. The next optimal number of clusters was found to be $c=3$ for the 5th highest average silhouette coefficient, with the minimum m -value of 1.5 once again being optimal. When the number of clusters equals 3, the silhouette coefficients decrease slightly as m is increased by 0.10 each step once again. The results suggest that there may be distinct clusters that need to be considered. Exemplified by the results in the table but can also be seen in the figure, the silhouette coefficients decrease as the m -value is increased, suggesting less optimal solutions as the fuzziness between clusters is increased. For both clusters $c=2$ and $c=3$, the smallest m -value displayed the best average silhouette coefficient.

Based on the results that suggest possible distinct clusters, rather than test the four parameter combinations which display the best average silhouette coefficient (which would be when $c=2$ and $m=1.5, 1.6, 1.7$ and 1.8), we will proceed with the four following models:

1. FCM when $c=2$ and $m=1$ (reduces to K-Means clustering where $k=2$)
2. FCM when $c=2$ and $m=1.5$ (FCM for $c=2$ with the minimum fuzziness factor $m=1.5$)
3. FCM when $c=3$ and $m=1$ (reduces to K-Means clustering where $k=3$)
4. FCM when $c=3$ and $m=1.5$ (FCM for $c=3$ with the minimum fuzziness factor $m=1.5$)

This will allow us to: 1) test the optimal result from the silhouette analysis ($c=2$ and $m=1.5$), 2) test multiple clusters ($c=2$ and $c=3$), and 3) see whether applying distinct clusters leads to better calibration.

3.4.4 Results from Clustering for Daily Streamflow Modeling

New physical models were created using clustering for the four scenarios discussed in the previous section. Figure 28 summarizes how each model created from clustering performs for daily streamflow estimation.

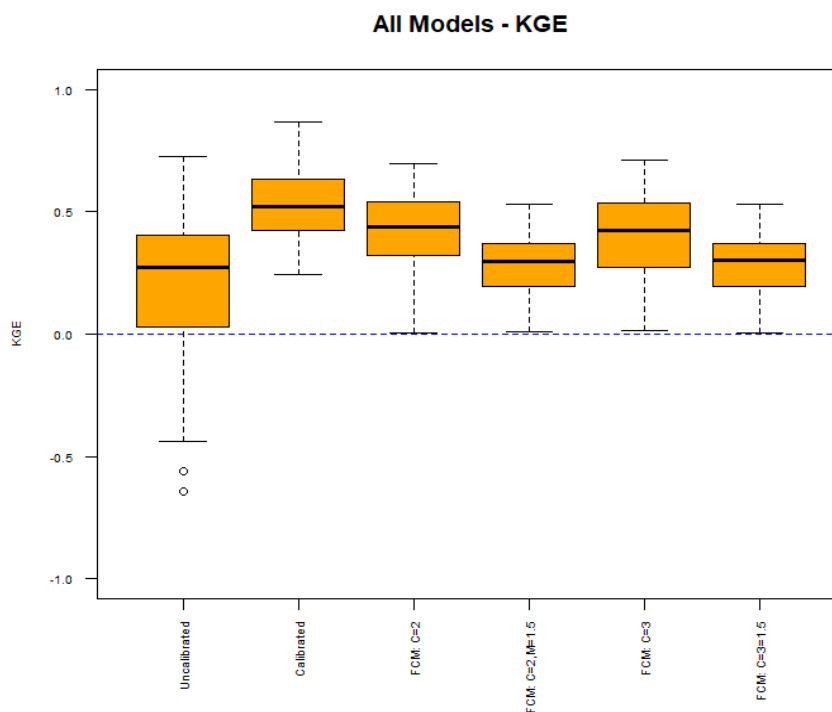


Figure 28: Daily streamflow KGE using FCM for calibration.

Figure 28 shows that some models calibrated using FCM display improved KGEs compared to the uncalibrated models. Both models with no fuzzification factor (FCM: C=2 and FCM: C=3) display median KGEs slightly below 0.50, while the models with a fuzzification factor (FCM: C=2, M=1.5, and FCM: C=3, M=1.5) display median KGEs around 0.30. Even with this improvement however, there are some locations that have a KGE close to 0 for all four models calibrated with FCM. All models calibrated using FCM display minimum KGEs close to 0, but this is still an improvement compared to the uncalibrated models where the 25th percentile KGE is just above 0. The models created

with a fuzzification factor of $m=1.5$ display much poorer median KGEs than the models created with distinct clusters. The results from silhouette analysis suggested that distinct clusters may provide more optimal solutions, and the results from Figure 10 support that using distinct clusters improves KGE significantly more on average than the models created with fuzzification factors.

In Figure 29a and b, we once again separately analyse the models created for small basins and large basins.

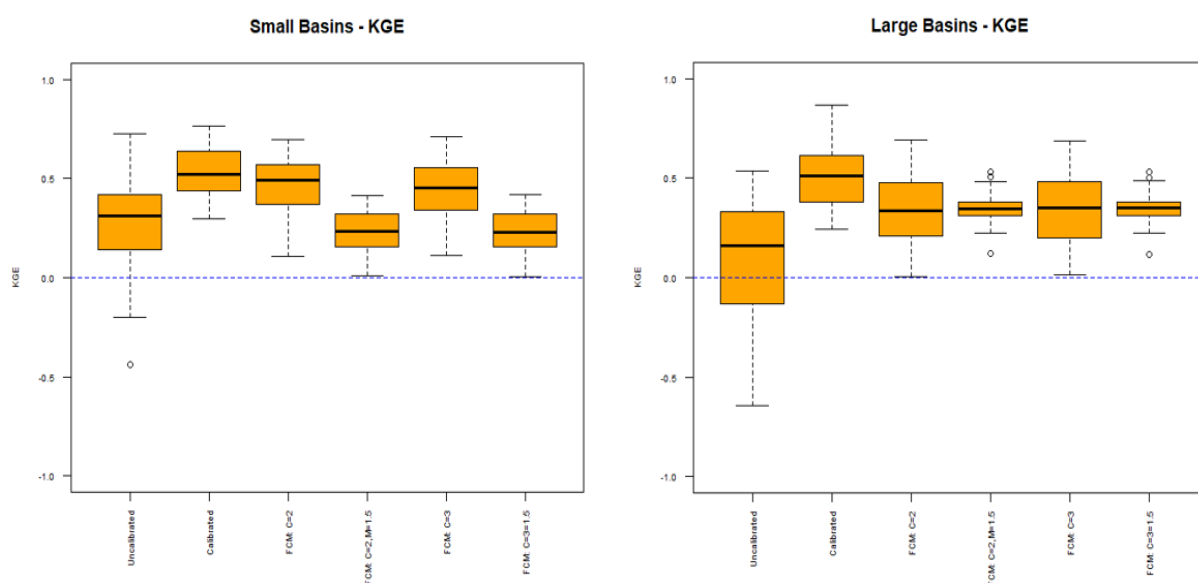


Figure 29: Results from calibration for basins smaller than 100 mi² (a, left) and basins larger than 100 mi² (b, right) using KGE to evaluate daily streamflow estimation.

The results provide further explanation for the general results given in Figure 28. Figure 29a shows that results for the smaller basins are very similar to the overall results given in Figure 28, that models calibrated with distinct clusters once again provide only slightly poorer median KGEs than the model calibrated to daily streamflow at each location. It also supports that using distinct clusters provides significant improvement compared to using a fuzzification factor. However, Figure 29b shows that for the larger basins, all models created using clustering (with or without a fuzzification factor) display almost the

same median KGE of about 0.30. The models created using a fuzzification factor even display slightly higher median KGEs, with smaller interquartile ranges and better minimum KGEs. Overall, results from Figures 29a and 29b suggest that distinct clusters should be used when creating physical models for smaller basins ($<100 \text{ mi}^2$), but for larger basins ($>100 \text{ mi}^2$), adding a fuzzification factor may provide similar results on average but less variability.

Next, we test how these models perform specifically for 7Q10 estimation.

3.4.5 Procedure for Optimal Cluster Selection for Overall Model Performance

We provide an analysis of all models for 7Q10 estimation in Table 15.

Table 15: Error metrics for 7Q10 estimation (only StreamStats locations)

7Q10 Estimation Technique	Median Relative Bias (cfs/cfs)	RMSE (cfs)	RRMSE (cfs/cfs)	UA-RMSE (cfs/mi²)
Uncalibrated Hydrology Models	0.81	13.99	116.48	0.08
Calibrated Hydrology Models	0.61	12.33	102.63	0.07
FCM (C = 2)	0.84	14.31	119.16	0.08
FCM (C = 2, M = 1.5)	0.97	19.34	161.10	0.08
FCM (C = 3)	0.83	14.47	120.48	0.08
FCM (C = 3, M = 1.5)	0.97	19.35	161.08	0.08
StreamStats	0.42	13.75	114.5	0.05

The results from Table 15 suggest that even with the improvement for daily streamflow estimation shown previously, none of the models provide better median relative bias, RMSE, RRMSE, or UA-RMSE compared to StreamStats. The models created using distinct clusters provide slightly higher RMSE and RRMSE than StreamStats but provide substantially higher relative bias and UA-RMSE. We further analyse this in Table 16, which separates small and large basins for analysis.

Table 16: Error metrics for 7Q10 estimation (only StreamStats locations)

7Q10 Estimation Technique	Median Relative Bias (cfs/cfs)	RMSE (cfs)	RRMSE (cfs/cfs)	UA-RMSE (cfs/mi²)
Basins smaller than 100mi ² - 35 Locations				
Uncalibrated Hydrology Models	0.90	7.26	188.82	0.09
Calibrated Hydrology Models	0.90	5.47	142.24	0.08
FCM (C = 2)	0.81	5.63	146.53	0.08
FCM (C = 2, M = 1.5)	0.97	6.06	157.72	0.08
FCM (C = 3)	0.82	5.55	144.47	0.08
FCM (C = 3, M = 1.5)	0.98	6.08	158.08	0.08
StreamStats	0.42	2.23	58.00	0.04
Basins larger than 100mi ² - 31 Locations				
Uncalibrated Hydrology Models	0.65	19.57	92.17	0.07
Calibrated Hydrology Models	0.42	16.25	76.53	0.06
FCM (C = 2)	0.88	20.00	94.23	0.08
FCM (C = 2, M = 1.5)	0.97	27.48	129.44	0.08
FCM (C = 3)	0.88	20.27	95.49	0.07
FCM (C = 3, M = 1.5)	0.97	27.47	129.45	0.08
StreamStats	0.32	19.92	93.86	0.06

Table 16 shows that for small basins, StreamStats provides the best 7Q10 estimation by far for all metrics employed. In terms of median relative bias, RMSE, RRMSE, and UA-RMSE, StreamStats' estimates provide roughly half of the error compared to all other methods. For larger basins however, results are much more mixed. StreamStats does provide the best median relative bias by far, but for all other metrics, the calibrated models and models created with no fuzzification factor perform comparably to StreamStats for 7Q10 estimation. The calibrated models provide better RMSE and RRMSE, as well as an identical UA-RMSE. The models created using FCM with no fuzzification factor also display comparable but slightly larger RMSE, RRMSE, and UA-RMSE than StreamStats.

In Figure 30a, b, c, and d, we provide a plot of the 7Q10 bias for each model. Bias is calculated by [estimated – actual], meaning points above the 0 line suffer from overestimation and points below suffer from underestimation. The corresponding line through the points follows a loess smoothing curve (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/loess>) with the corresponding standard error highlighted in gray. This provides those interested in using these models for 7Q10 estimation with three valuable insights:

1. For any of the models, the average bias that can be expected with a basin area of X.
2. The corresponding confidence in that estimate, given by the highlighted area.
3. If any of the models are consistently overestimating or underestimating 7Q10s for a range of basin sizes.

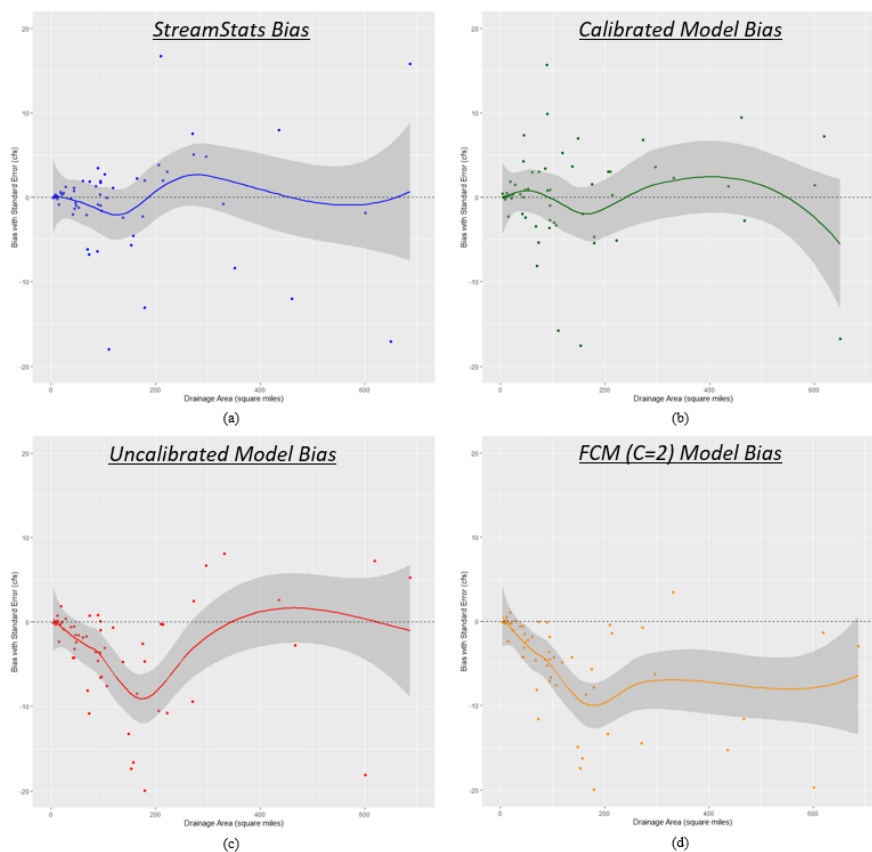


Figure 30: Bias created from each model used, arranged by drainage area.

For both StreamStats and the calibrated models, the average bias remains around 0 for all range of basin sizes. There seems to be some slight oscillation that suggests slight overestimation or underestimation depending on the basin size, but 0 remains in the highlighted confidence interval for the full range of basin sizes. For the uncalibrated models and models created using FCM ($C = 2$), there are clear patterns that demonstrate weaknesses in these modeling approaches. The uncalibrated models are heavily underestimating 7Q10s for basin sizes that range from 50-250mi², and the models calibrated with FCM ($C=2$) are consistently underestimating 7Q10s for basins larger than 50 mi². Given the results, the only physical models that provide sufficient 7Q10 estimation are the individually calibrated hydrology models, which cannot be used in ungaged basins, while StreamStats can.

3.5 Conclusions

Based on the results from this experiment, we conclude that resource managers who require 7Q10 estimates in gaged basins can use calibrated models in basins larger than 100mi² and expect similar errors to current statistical estimates. For basins smaller than that, statistical estimates still provide smaller median relative bias, RMSE, RRMSE, and UA-RMSE for 7Q10 estimation. 7Q10s in these smaller basins can be extremely small values, however, and it should be considered how accurate estimates need to be to be sufficient for their exact design application. For example, the 7Q10 is frequently used in wastewater treatment plant design as a mixing flow. Though it is difficult to attribute exact costs to value changes, an estimated mixing flow of 1.00 cfs could cause a significantly different design than a mixing flow of 0.10 cfs, only having a difference of 0.90 cfs. However, in the case of a future climate change study where the accuracy of the value itself is less important than model-predicted future changes in the value, the process-based estimates readily allow for the incorporation of altered climate and/or landcover projections. Given few alternatives, the process-based models also provide the ability for resource managers to estimate 7Q10s in states where StreamStats 7Q10 estimation is not yet developed. Future work related to this study could include testing this procedure with other process-based models, determining other ways to calibrate physical models in ungaged locations for both daily streamflow and low flow estimation, and further analyzing the performance of physical models for gaged and ungaged high flow estimation (e.g. 100-year-flood estimation) as well.

CHAPTER 4

A Stakeholder-Driven Decision Support System Web Tool for Calculating Historic and Future Projected 7-Day, 10-Year Low Flows in the Northeast United States

Resource managers require estimates of infrequent, low streamflow events (such as the 7-day, 10-year low flow – 7Q10) for many reasons, including wastewater permitting, municipal water planning and management, and protecting aquatic species. This paper highlights the development of a Decision Support System (DSS) web tool designed to assist resource managers in the northeast U.S. in calculating the 7Q10, for both gaged and ungaged streams. Interviews were conducted initially with resource managers throughout the U.S. to identify techniques used to calculate the 7Q10 in practice. The interviews indicated that techniques were inconsistent, and few methods attempted to incorporate future climate change into calculations. A second round of interviews with only resource managers from the northeast U.S. identified areas in which improvements to current techniques could enhance both the efficiency and accuracy of calculations. These interviews led to a stakeholder user-group that consistently provided insight into the development of the new webtool for calculating the historic 7Q10 at/near gages and estimating the 7Q10 at ungaged locations, also allowing users to project future climate-driven changes in 7Q10s at ungaged locations. In addition to standard methods for calculating the 7Q10 at gages and estimating the 7Q10 at ungaged locations, the completed tool provides users with the ability to: 1) calculate the 7Q10 using a user-selected subset of the historic streamflow record at a gage, 2) apply the Mann-Kendall test for detecting trends in annual 7-day low flows, and 3) apply a procedure from the Transportation Research Board's NCHRP 15-61 for incorporating climate change projections into ungaged extreme flow estimation.

4.1 Introduction

Hydrologic drought management poses a formidable challenge both internationally and domestically due to a combination of complex and interconnected factors. Water resources are intricately linked to various sectors such as agriculture, industry, and municipal supply, making it challenging to balance competing demands during hydrologic drought events (Smakhtin, 2001). Though short-term human activity (such as withdrawals, releases, and hydropower management) can contribute to hydrologic droughts, short and medium-term weather are the underlying natural physical factor in determining the severity of a drought (Smakhtin, 2001). Long-term changes in Green House gases are also generating further fluctuations in the global climate and water resources (Milly et al., 2008). The inherent uncertainty surrounding the severity of hydrologic droughts exacerbates the difficulty in designing adaptive policies and infrastructure to combat droughts (Mount et al., 2015). As a result, effective long-term drought management requires diverse stakeholder collaboration to ensure that management decisions consider all factors, including climate science, hydrology, economics, and social dynamics (Cardwell et al, 2008).

In addition, effective hydrologic drought management requires an extensive understanding of the hydrologic cycle paired with ample historic data for a complete scientific analysis. Though weather is the natural physical factor in determining the severity of a hydrologic drought, topography dictates the shape of a watershed (a process known as delineation), and both landcover and topography also affect the runoff into the river. The depth of the water table can also dictate the minimum flow, as the lowest possible flow in a river may be limited to the height of the water table in that location.

These factors vary geographically and in every watershed, making it difficult to generalize calculations for predicting the severity of drought flows in basins that do not have ample historic data (referred to as ungaged basins). The necessity to analyze watersheds in ungaged basins, whether for daily streamflow or extreme flows (e.g., hydrologic droughts), provides the opportunity to create computer programs that merge the ability to delineate watersheds, calculate the landcover and topography for the delineated watershed, and evaluate historic data in other gaged basins to guide predictions for the delineated watershed. In the United States, publicly available computer programs, such as the USGS' StreamStats (Ries et al., 2008) and the EPA's Basins (US EPA, 2019) include modules that efficiently delineate watersheds, calculate the landcover and/or topographic characteristics of the delineated watershed, and apply historic drought data from other locations for estimating the severity of hydrologic droughts in the delineated watershed (e.g., Ries et al., 2000). These programs conform to generally accepted procedures for calculating extreme flow values on purely historic data, but recent uncertainties in the global climate have caused some to rethink traditional methods for calculating current and future extreme flows (Milly et al., 2008).

Our research goal is to create a computer program, typically referred to as a Decision Support System (DSS), for hydrologic droughts that incorporates recent developments in extreme flow calculation. DSSs are traditionally defined as applications that can analyze large sets of data for organizational or business purposes that enable improved decision-making. DSSs allow decision-makers to efficiently analyze the perceived consequences of different potential decisions, and as a result, they have become valuable tools in the field of water resources planning and management (Serrat-

Capdevila, 2011). Dr. Peter Loucks is credited as being one of the first researchers to discuss the use of DSSs specifically in water resources planning and management in literature, publishing a paper in 1991 presenting a brief overview of some of (at the time of publication) the current developments in decision support system technology and its application to water resources problems (Loucks, 1991). Four years later, he published a more thorough analysis critiquing and discussing the challenges associated with developing and implementing DSSs, which has been extensively cited in literature (Loucks, 1995). DSSs have continued to be prominent in the field of water resources planning and management, and some more recent applications of DSSs for water related issues include creating an integrated DSS for the urban food-water-energy nexus (Ghodsvali et al., 2023), a System Dynamic-DSS model of sustainable groundwater resources management using the water-food-energy security Nexus in Alborz Province (Rahmani et al., 2022), and an irrigation supporting system for water use efficiency improvement in precision agriculture (Bonfante et al., 2019). Exemplified by StreamStats and Basins, estimating the severity of hydrologic droughts is a strong candidate for DSSs because it requires a standard framework that pulls data from many sources and relies on parameters that cannot be determined until a location is selected and the corresponding watershed is delineated.

Literature highlights the importance of involving the ultimate users of a DSS (e.g., decision makers) in the DSS's development to ensure that the tool is actionable as intended. In 2008, the U.S. Army Corps of Engineers noted that Shared Vision Planning- a structured process for involving stakeholders in the development of water management solutions- leads to more sustainable solutions that address conflicting interests and

incorporate diverse perspectives (Cardwell et al, 2008). This paper also provides recommendations for the development of an interactive computer model (e.g., a DSS) for supporting water resources planning and management, noting that interactive computer models for decision support should involve extensive stakeholder engagement because it 1) ensures that the model will integrate all issues relevant to the decision making process, 2) becomes fully transparent with respect to input data, assumptions, and the calculation/production of outputs, and 3) results in a user-friendly interface, encouraging the decision makers (who may not be experts in the science) to continue using the program after development (Cardwell et al, 2008). Another review paper discussing DSSs in water resources planning and management documents that a collaborative and participatory framework that brings together stakeholders from various sectors, including government agencies, communities, industries, and environmental organizations is necessary to ensure actionable results (Serrat-Capdevila, 2011). This concept is also emphasized extensively in recent environmental literature, including a case study of Lake Como (Bertoli et al., 2022), a review of successful decision support tools and their characteristics (Wong-Parodi et al., 2020), and the development of a holistic framework for the successful management of environmental flows (Mussehl et al., 2022).

As a result, we interviewed resource managers throughout the U.S. that utilize extreme low flows in their occupation to determine ways in which current programs used for extreme flow estimation (e.g., StreamStats and Basins) could be improved. These interviews provide important context on the current state of hydrologic drought calculation, also referred to as low flow calculation and/or estimation in the U.S. These interviews confirmed that resource managers desire improvements to current

methodologies. Many managers stated that they seek a simple and appropriate method to incorporate future climate change into extreme flow calculations, but they have yet to find an efficient way to do so. From those interviews, we developed a smaller stakeholder user-group, comprised of resource managers from the northeast U.S. to provide recommendations and/or collaborate in the development of this DSS. We followed generally accepted procedures for developing a stakeholder driven DSS (Figure 31) and each step is discussed in the following sections of this paper.

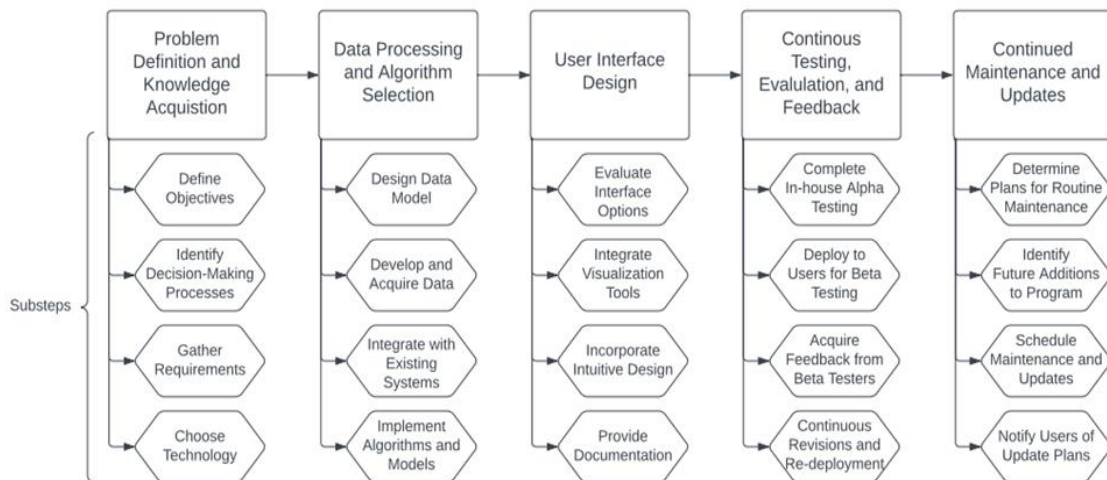


Figure 31: Generalized process for creating an effective decision support system. Rectangular boxes highlight the major steps and hexagonal boxes represent substeps for each major step.

4.2 Problem Definition and Knowledge Acquisition

In this section, we highlight the process for defining the research problem and the steps taken to acquire the required knowledge to create an effective program.

4.2.1 Define Objectives

In this research, we began with a broad research objective that was refined through consistent stakeholder engagement. The broad research objective was to incorporate new

methods for extreme flow estimation to assist resource managers in low flow calculations and estimation. Next, we identified four broad subgoals, summarized below:

1. Identify current ways in which low flows are being commonly calculated and applied in practice.
2. Determine the data and decision support needs of resource managers, specifically pertaining to low flow calculations.
3. Work collaboratively with resource managers in the region to develop a DSS that they will be encouraged to use after development.
4. Devise an efficient strategy, trusted by the resource managers, to incorporate future climate projections into low flow calculations.

4.2.2 Identify Decision-Making Processes and Gather Requirements

To identify the decision-making processes involved in hydrologic drought management, we completed a literature review and interviewed those who we determined utilize extreme flows in an operational capacity. The following subsections highlight the literature review and major results from the interviews.

4.2.2.1 Literature Review

In the United States, one extremely common hydrologic drought planning metric is the 7-day, 10-year low flow (7Q10). The 7Q10 is the lowest 7-day average flow that occurs (on average) once every 10 years (EPA Office of Water, 2018). The 7Q10 is used in a variety of design applications, including wastewater treatment plant design, aquatic species protection, and hydropower management (Blum et al., 2019). Traditionally, the 7Q10 is calculated at gaging sites using the full record of historic daily streamflow available. Estimating the 7Q10 at other locations on a gaged stream can be performed by relating

the gaged basin area to the corresponding up or downstream ungaged watershed area (e.g., Asquith and Thompson, 2008). For stream locations without gages nearby, methods exist that relate the physical characteristics of a gage's corresponding watershed to 7Q10 values, typically using multiple linear regression in log space (Ries et al., 2000). These equations are trained on publicly available predictor variables (such as watershed area, elevation, slope, among other physical characteristics) which allow for an estimation of 7Q10s in ungaged basins. StreamStats and Basins both include modules for the calculation of a regression-based 7Q10 estimate. These estimated 7Q10 values at gaged and ungaged locations conform to the generally accepted procedures for calculating 7Q10 values highlighted earlier, but these methods have significant drawbacks.

Because the 7Q10 values calculated at gages and the regression equations trained on these values only utilize historic data, they inherently rely on "stationarity," the assumption that the statistical properties of streams do not change over time. Recent studies suggest that hydrologic processes may no longer satisfy that assumption, due to global and regional climate changes, exposing a weakness in the stationary modelling approach (Milly et al., 2008; Bayazit, 2015; Salas et al., 2018; Hesarkazzazi et al., 2021). For instance, in 2022 it was estimated that the southwestern U.S. was experiencing its driest 22-yr period since 800 CE and approximately 20% of this deviation from previous extremes can be attributed to recent anthropogenic changes (Williams et al., 2022). In the northeast U.S., studies have found both average baseflows and 7-day summer baseflows are increasing with statistical significance (Hodgkins and Dudley, 2011; Ayers et al., 2018). In the Mid-Atlantic, Blum et al. (2019) found increasing 7Q10s in the northern part of the Mid-Atlantic (New York, Pennsylvania) and decreasing 7Q10s in lower Mid-

Atlantic (Virginia, Maryland) (Blum et al., 2019). Additionally, Blum et al. (2019) found that if there is a statistically significant trend in the annual low flows at gage (suggesting that the annual low flows are increasing or decreasing) that only using the most recent 30 years of historic streamflow data to calculate the 7Q10 reduces error and bias in 7Q10 estimators compared to use of the full record (Blum et al., 2019).

7Q10 is an estimate of a stream's current, recurring low flow. When used in design, it represents an estimate of the lowest 7-day flow expected in any 10-year period. Blum et al. (2019) reevaluates the meaning of "current" in the context of 7Q10. When global climate was assumed stationary, using the entire historic data (e.g., the full record of streamflow data at a gage) to calculate extreme flows provided the "best estimate" of the 7Q10. Because the global climate is now assumed to be non-stationary, other techniques to estimate low flows like the 7Q10 must be considered. Blum et al. (2019) presented a framework for determining when to use a shorter, subset historic streamflow record to calculate the 7Q10, providing physical evidence that a shorter, recent streamflow record in some situations is more representative of the current low flow conditions on a stream than using the entire record (Blum et al., 2019). This method, although more comprehensive than only using the full record of streamflow all the time, still relies purely on historic data and does not consider the impacts of further climate changes. Other methods must be explored to incorporate the projected impacts of future climate change on streamflow.

Two primary approaches have emerged in literature for incorporating future climate into streamflow estimation: statistically based methods that relate similar hydrologic data and physically based methods that model the hydrologic process.

Physically based approaches traditionally estimate streamflow by incorporating the most important features of the rainfall/runoff process and meteorological data in computer models to generate daily streamflow time-series data (e.g., Siddique et al., 2020). When used for extreme flow analysis, the required extreme flow(s) (e.g., the 7Q10) are calculated from the computer-generated daily streamflow. To simulate future climate change, future climate projections (instead of historic values) are used as meteorological inputs into the physical model and the extreme flows are once again calculated from the computer-generated daily flows (e.g., Siddique and Palmer, 2021). In contrast, statistical methods (like the regression equations developed for low flow estimation in ungaged basins) can also be altered to incorporate future climate projections. Dimitriadis and Koutsoyiannis (2018) present a stochastic method for incorporating non-stationary processes into equations that rely on stationarity. Another procedure (summarized in the Transportation Research Board's National Cooperative Highway Research Program (NCHRP) Project 15-61) suggests that when the regression equations utilize climate variables as inputs, future projected climate values can simply be used as new inputs into the regression equation. This method is deemed most appropriate if the use of the regression equation does not require extensive extrapolation when used with the future climate data (Kilgore et al., 2019).

4.2.2.2 Interviews for Knowledge Acquisition

To gain a comprehensive understanding of the state of low flow calculations, in addition to a formal literature review, we interviewed resource managers throughout the U.S. involved in the calculation and/or utilization of low flows in practice. We began by researching governmental organizations and individuals from those organizations that we

determined may be involved with extreme flow calculation. Potential interviewees included, but were not limited to, managers of aquatic flows, permit writers, biologists engaged in establishing instream flow targets, and transportation engineers. Via email, we contacted these organizations/individuals introducing ourselves, describing the project, and seeking their participation in future interviews. If the individual contacted was not directly involved with extreme flow estimation, alternative interviewees were requested. In many cases, the interviewee engaged other members of their organization in their interviews. In total, sixteen virtual interviews were conducted with those that we determined were indeed utilizing extreme flows in an operational capacity. General information about each individual or organization interviewed is summarized in Appendix B.

For each interview, the representatives were asked a standardized list of eleven questions. These questions, as well as the most common responses to each question, are included in Table 17.

Table 17: Questions for Initial Interviews

Question	Reasoning	Most Common Response(s)
What is your current job title, how long have you been at this organization, and what is your primary role in the organization?	Determine background information and expertise of interviewee.	Job title, # of years worked, etc.
Are there decisions or management choices upon which low flows are important in your work? If so, can you give some examples?	Determine the exact use(s) of extreme flows in the interviewee's profession.	National Pollutant Discharge Elimination System (NPDES) permits require 7Q10 values.
How do you currently calculate those values in (or near) streamflow gages?	Verify that extreme flows are being calculated as documented in the introduction section.	Use the full historic record or the last 30 years of streamflow data to calculate the value and then use a watershed area/area ratio to scale the flow if the required location is not directly at the gage.
How are these values estimated on ungauged streams?	Verify that extreme flows are being calculated as documented in the introduction section.	Using the USGS' StreamStats program, if available.
What tools have proven most valuable to you in estimating low flows? Why do you use these specific tools?	Determine current tools that the interviewee frequently uses for extreme flow estimation and why they are beneficial.	The USGS' StreamStats, SWToolbox, and/or DFlow programs. These programs are accurate and very efficient.
Are there limitations to your current approaches, or is there additional information that would aid you in decision making?	Determine ways in which we can improve current methodologies/tools during the development of our tool.	StreamStats' low flow estimation has not been developed in all states, regression equations are old or always utilize the full record of streamflow at every training location.
If you use ungauged estimates of low flows in your work, do you consider the uncertainty in calculating and/or presenting those values?	As climate change has large uncertainties, determine if resource managers are aware of the uncertainty in their current estimation techniques.	Yes, but a specific value is required for design, so only the estimate itself is moved forward with.
If you answered yes to the previous question, what (if anything) would you change about the calculation of the values themselves and/or the presentation of the results?	Determine ways in which we can improve current methodologies/tools during the development of our tool.	More recent regression equations, potentially trained on the last 30 years of data at each training location.

Do you consider the impacts of climate change when calculating low flows?	Determine if resource managers are incorporating climate change in extreme flow estimates.	There is interest but there are few trusted methods in which we could justify applying projected changes to the public.
Would you be willing to test a new Decision Support System for calculating/estimating low flows in your state/area?	Determine if they would be interested in using the tool after development.	Yes/No
Are there any additional suggestions you would like to make about the development of our Decision Support System?	Provide a place where they can provide their own opinion(s) on the current state of extreme flow estimation.	Responses varied

The initial interviews identified two immediate functions that the DSS could improve for decision makers. The first is that many resource managers have either chosen to, or in some cases, are required to use only the last 30 years of streamflow data to calculate the 7Q10 at streamgages. In general, this has led many resource managers to limit their use of currently available DSSs, though many noted that StreamStats is still widely used for the delineation of ungaged watersheds and the retrieval of the physical characteristics of said watershed. As an example, it has been the policy of the Environmental Protection Agency Region 1 (covering the northeast U.S.) to *always* use the last 30 years of streamflow data to calculate the 7Q10 (EPA Office of Water, 2018). This change occurred before Blum et al. (2019) recommended using the last 30 years of streamflow data to calculate the 7Q10 in the case where a statistically significant trend is detected in the annual low flows.

Secondly, resource managers are aware that climate change will impact future low flows but have no generally accepted procedure to incorporate climate change into their calculations. It was already a broad subgoal of the DSS to include a methodology for incorporating climate change into calculations, but the interviews confirmed that resource

managers are aware of this shortcoming to current methodologies. Most mentioned that they would welcome a methodology for doing this if it is efficient and trustworthy. Some resource managers demonstrated a thorough understanding of methods for incorporating climate change into extreme flow calculations, some of which are summarized in the literature review. During these interviews, we were able to have deeper discussions about which methods might meet the needs of the resource managers if incorporated into a DSS. The process of incorporating a physically based rainfall-runoff model to estimate climate change's impact on extreme flows was often believed to require more resources than were available. They highlighted the complexity of setting up rainfall-runoff models and the large uncertainties attributed to using daily streamflow estimation solely for the purpose of calculating infrequent, extreme flows. Other approaches that incorporated a simple statistical methodology were considered superior because it could be more easily justified to public stakeholders.

Though a stochastic framework for incorporating non-stationary processes into equations that rely on stationarity (like the one presented in Dimitriadis and Koutsoyiannis, (2018)) could be applied to incorporate future climate projections into statistical extreme flow calculations, resource managers noted that the procedure from the Transportation Research Board's (TRB) National Cooperative Highway Research Program's (NCHRP) Project 15-61: *Applying Climate Change Information to Hydrologic and Hydraulic Design of Transportation Infrastructure* (Kilgore et al., 2019) is the most likely method to be accepted in practice. They reiterated that most resource managers are already familiar with, or frequently use, regression equations for estimating flows in ungaged basins, and that applying these same equations with altered climate variables

will only cause slight revisions to their current methodologies. Although relatively easy to implement, resource managers indicated that this methodology has not yet been incorporated into standard practices for two reasons:

1. Many previously developed regression equations for extreme flow estimation do not utilize any climate variables, making it impossible.
2. It would require extensive additional work, including:
 - a. Downloading multiple downscaled future climate projections and verifying that the future climate values are appropriate for the study area.
 - b. Extracting the exact, required predictor climate variables (average annual precipitation, seasonal totals, etc.) for the ungaged watershed of interest.
 - c. Verifying that the results seem appropriate, especially if the new climate variables surpass the original training bounds of the equation (extrapolation).

In previous research, machine learning algorithms have been evaluated compared to classical, statistical methodologies to create regional regression equations for estimating 7Q10s, applicable for the entire northeast U.S. (DelSanto et al., 2023). Input variables included watershed area, topographical variables (average watershed elevation and slope), landcover variables (percentage of the watershed considered forest and wetland), and climate variables (low cumulative precipitation and corresponding high temperatures). Because the regression equations are applicable for the entire northeast U.S. and include statistically significant climate variables, the method presented in NCHRP 15-61 could be combined with these equations to incorporate future climate

projections directly into ungaged, regression-based 7Q10 estimation, addressing the first issue raised above.

To address the second issue previously noted, we envisioned as DSS that provides suggested values for resource managers to use based on the results of the analysis. The following steps have been taken to retrieve the data and ensure that the projections are appropriate:

1. Nine Downscaled Global Circulation Models (GCMs) were downloaded for the study area. The simulated historic estimates provided by the GCMs are compared to observed historic values to ensure that the selected GCMs are appropriate.
2. The localized future climate projections are examined to ensure that they align with anticipated climate changes both globally and regionally. Specifically, future yearly and summer (JJA) climate projections are analyzed to provide general future climate projections made by the GCMs.
3. The exact climate variables required for the regression equations are extracted from the GCMs and the values are examined to determine if extrapolation will be necessary.
4. The average projected climate change is applied to the equation and the resulting percent change in 7Q10 is presented. The use of these values can be supported by comparing the results to historic, observed changes in 7Q10s or the results from a physical model.

We provide this full analysis in Appendix C, with the main results summarized below:

1. Extrapolation will only be needed for the temperature variable. In the short-term (2020-2060), only the top 25% percentile of predicted temperature changes will

require extrapolation. In the long-term (2060-2100), roughly 75% of all predicted temperatures will require extrapolation.

2. Future temperature changes will likely outweigh the impacts of future precipitation changes. Low flow events such as the 7Q10 occur during extended periods of little to no precipitation, which can only reach up to 0 cumulative precipitation, while evapotranspiration values increase as a function of temperature and will continue to impact low flows.
3. In the worst-case scenario (RCP8.5 for the years 2060-2100), where temperatures and precipitation are expected to change 10°F and 15% respectively, the model is predicting changes up to around -55% from the historic 7Q10 values.

4.2.3 Choose Technology

Resource managers requested that the DSS be deployed as a web-application, similar to StreamStats, that does not require the installation of any local programs or packages. We determined that utilizing Shiny (<https://shiny.posit.co/>) would be appropriate for the requirements of the DSS. Shiny is a web application framework for R, with recently added python functionality, that allows the developer to deploy interactive web applications directly from a local integrated development environment without requiring knowledge of traditional web-based programming languages like HTML, CSS, or JavaScript (<https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/index.html>).

4.3 Data Processing and Algorithm Selection

In this section, we provide background on data processing and the utilization of algorithms.

4.3.1 Design Data Model

Based on the responses during the initial interviews, a prototype DSS was developed to present to stakeholders for evaluation. This first prototype was created to demonstrate the potential of the DSS and provide stakeholders with an early vision for what future iterations could look like or include. This version did not include any modules pertaining to future climate projections and was made to target smaller issues raised by the stakeholders that would further future collaboration efforts. This prototype only included the following:

1. A module for automatically calculating the 7Q10 at a streamgage, utilizing both the full record of streamflow and the last 30 years of streamflow data.
2. A module for estimating the 7Q10 at ungaged locations using the regression equations cited earlier, trained on the full record of streamflow (3 regression equations from DelSanto et al., 2023) and the last 30 years of streamflow at every training location (3 additional regressions equations using the same methodologies).

Figure 32 highlights these modules and the main required steps for each.

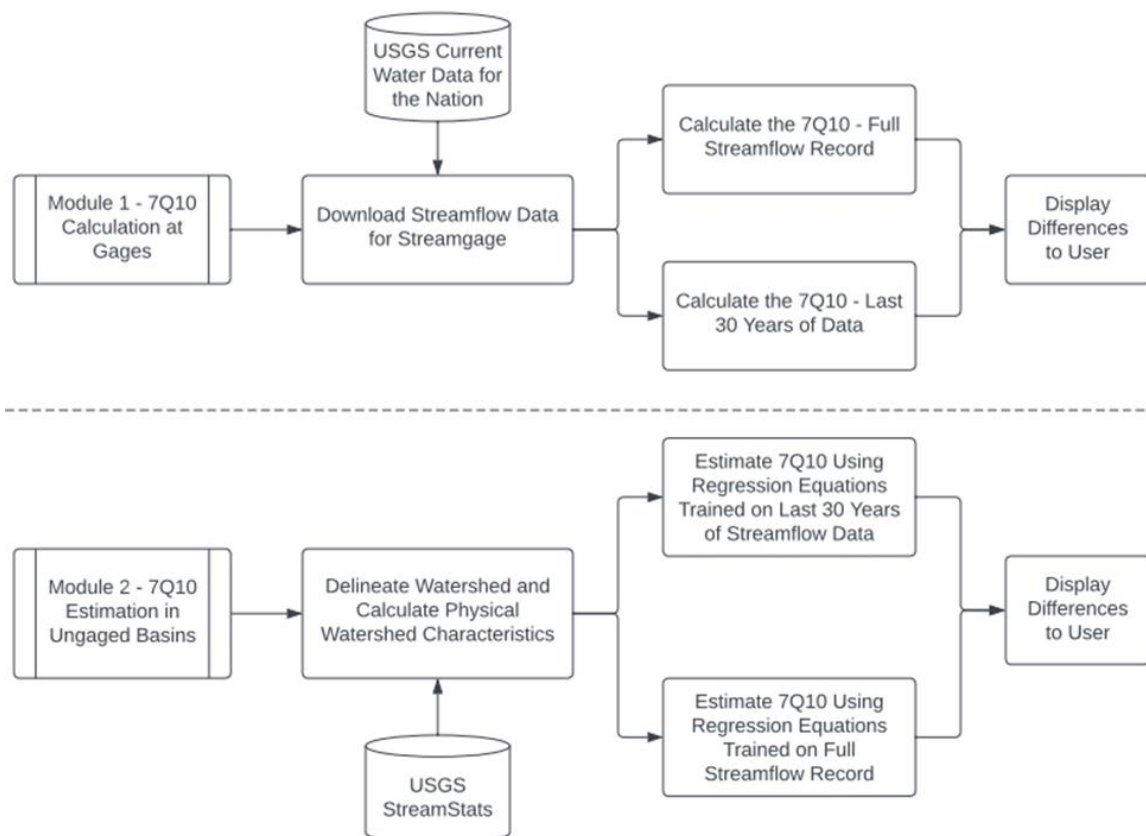


Figure 32: Data model for the first prototype of the decision support system. The first module utilizes daily streamflow data from the USGS’ Current Water Data for the Nation (U.S. Geological Survey, 2016) and the second utilizes the watershed delineation and corresponding physical watershed characteristics from the USGS’ StreamStats program (Ries et al., 2008).

4.3.2 Acquire Data, Integrate with Existing Systems, and Apply Algorithms

This prototype required data from the USGS’ StreamStats (Ries et al., 2008) and USGS’ Current Water Data for the Nation (U.S. Geological Survey, 2016). From StreamStats, the program calls the StreamStats API for delineating the watershed given a selected latitude and longitude, and then extracts the physical characteristics of the delineated watershed. The user can then input these values into the regression equations to estimate the 7Q10. From the USGS’ Current Water Data for the Nation, the maximum amount of daily streamflow data is retrieved and then the data is used to calculate the full record 7Q10. After, the streamflow data is truncated to only the last 30 years of data, and the 7Q10 is

recalculated from the recent data. To calculate the 7Q10 from the daily streamflow data, we utilized the `fasstr` package (<https://cran.r-project.org/web/packages/fasstr/index.html>). To estimate the 7Q10 for the ungaged basins, we applied previous research (DelSanto et al., 2023) discussed earlier.

4.4 User Interface Design

In this section, we highlight the design of the user interface. This includes some additions that were incorporated into the current iteration of the program.

4.4.1 Evaluate Options, Integrate Visualization Tools, & Intuitive Design

To create an intuitive program, we mimic methods already applied by those involved in low flow calculations. This makes the program more intuitive and utilizes trusted methods, as we are not incorporating any entirely new techniques that may require extensive explanation. The first module of the program follows the typical procedure for calculating the 7Q10 on daily streamflow data from a streamgage, with slight adjustments requested by stakeholders themselves. The second module follows the same standard procedures for estimating the 7Q10 in ungaged basins, adding adjustments requested by stakeholders and an additional tab for applying future climate projections to the regression equations.

For interface themes, RShiny includes a variety of standardized themes. We chose the cerulean theme (<https://bootswatch.com/cerulean/>) because it includes a light blue navbar and utilizes easily distinguishable tabs for separating distinct steps. We deemed a light blue theme appropriate, as it both mimics the color of water and is traditionally considered a welcoming color. For visualization tools, we have utilized the package

ggplot2 (<https://ggplot2.tidyverse.org/>) for all graphics. We utilize the leaflet package (<https://leafletjs.com/>) for all maps. These are highlighted in Figure 33.

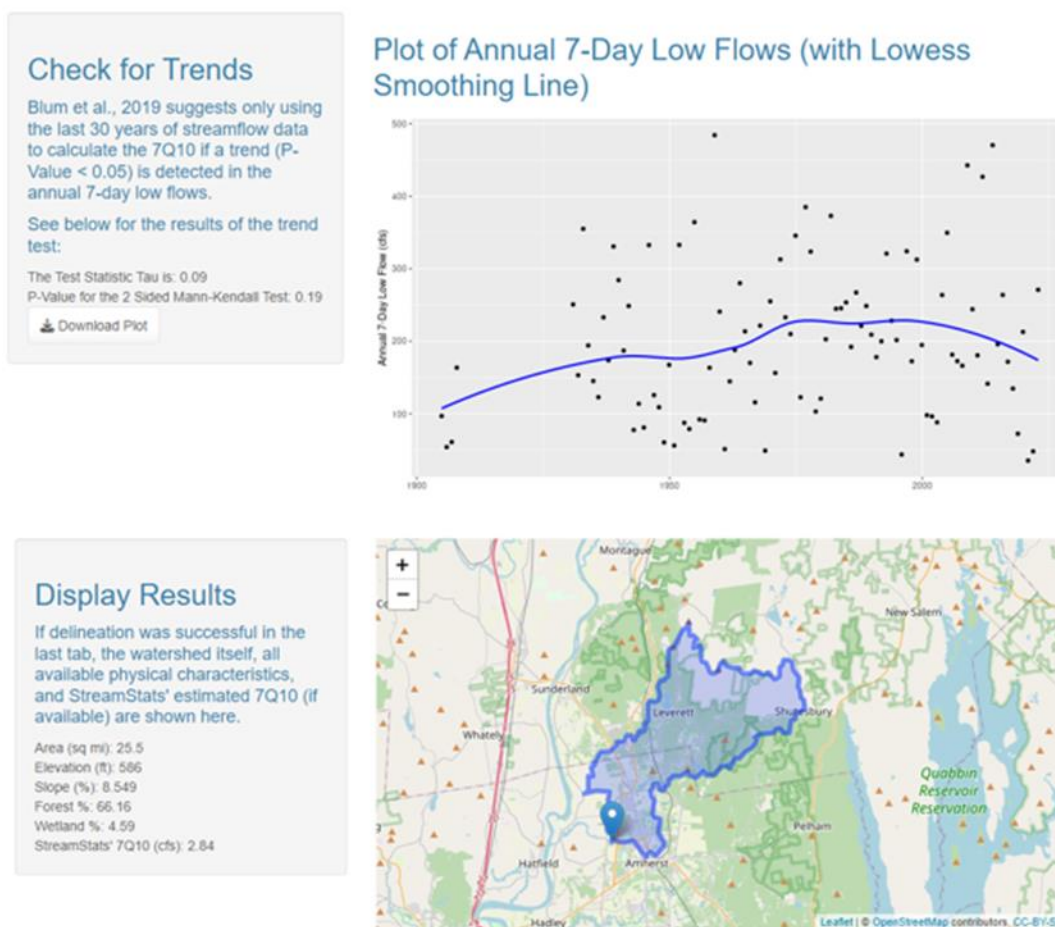


Figure 33: Examples of graphics (created with ggplot2) and maps (created with leaflet) used in the decision support system webtool.

4.4.2 Provide Documentation

In the final tab of the program, we provide a document illustrating the calculation process for all steps in the program to the user and the methods/algorithms utilized in each step (Figure 34).

DSS For Low Flows- Documentation


Link: <https://andrewdelsanto.shinyapps.io/gagedlowflowanalysisstool/>

Description: This module is designed for assisting resource managers in calculating the 7-day-10-year low flow (7Q10) at, or near, a USGS streamgauge. This tool is recommended for gages and locations slightly down or upstream of a gage where the unaged watershed area is within 50% of the gaged watershed area:

$$\text{At a gage or where } \left[0.50 < \frac{\text{Unaged Watershed Area}}{\text{Gaged Watershed Area}} < 1.50 \right]$$

Summary of Steps:

1. Download Streamflow Data
 - a. Enter an 8-digit NWIS gage number and press "Download Streamflow Data".
 - b. All available streamflow data is downloaded and stored.
 - c. The first and last date of available streamflow data is displayed.
2. Display Streamflow Data
 - a. Display summary statistics (quantile flows) on left.
 - b. Plot unrefined streamflow data with quantiles displayed as dashed lines.




Step by Step Walkthrough and Notes


Step 1: Download Streamflow Data

The program begins by allowing you to download the streamflow data from an NWIS streamgauge. This assumes that you know the 8-digit gage number for a streamgauge that is active and has continuous daily data. Note, this will not work with partial record stations.

If you would like to see a map of current NWIS streamgages, you can use the new National Water Dashboard, which has an interactive map showing gages and their most recent discharge data (<https://dashboard.waterdata.usgs.gov/app/nwd/en/?region=lower48&aci=default>).



Opening the program should bring you to a page that looks like this:



Open the first dropdown box

Retrieve Streamflow

Enter an 8-digit NWIS gage number with data from this:

Enter the date of the first and last data to retrieve:

Enter the date of the first and last data to retrieve:

Enter the date of the first and last data to retrieve:

If quantity retrieved, the first and last dates of daily streamflow data available will be displayed below. Once the data is successfully retrieved, you can download or view the data with the data and daily streamflow information.

Download

Figure 34: Documentation provided for the decision support system webtool.

4.5 Continuous Testing, Evaluation, and Feedback

In this section, we discuss the process for continuous testing, standardized evaluation, and feedback from the stakeholder user-group.

4.5.1 In-House Alpha Testing, External Beta Testing, and Acquiring Feedback

We extensively tested the program to identify potential calculation errors and interface errors. In addition, we gave an initial demonstration to the stakeholder user-group to present the first prototype. Following the presentation, we deployed the first iteration of the program online for them to test. A standardized questionnaire was sent out to all participants to guide further development. These questions and example answers are included in Appendix D.

Suggestions from the user group generated during the presentation and from the questionnaire resulted in numerous changes to the DSS after prototype demonstrations (Appendix E). Though many of the additions requested by stakeholders were minor, three were considered major methodological additions, which are summarized below:

1. Ensure the DSS can calculate the 7Q2 and 7Q30 from the streamflow data as well. These are used in other similar permits.

2. Rather than automatically calculate the 7Q10 using the last 30-years of streamflow data, allow the user to subset the streamflow data to their liking and then also calculate the 7Q2, 7Q10, and 7Q30 on the user-selected subset. This will satisfy those who want (or are required) to use the last 30 years of streamflow data to calculate the 7Q10 while also giving others the opportunity to manually subset flows.
3. Include the framework presented in Blum et al. (2019) to determine if there is a statistically significant trend in the annual low flows. This allows a user to justify whether or not to limit the data analysis to last 30 years of streamflow data when calculating low flow frequency.

4.5.2 Continuous Revisions and Re-Deployment

After receiving feedback from stakeholders on the prototype, we developed and deployed two more iterations of the program, and requested additional feedback from the user group. Changes during these iterations did not modify the basic framework of the system and included incorporating the ability to download all data and figures, documenting all methodologies, and modifying the graphical user interface. The current version of the program follows the same two modules, but has additional steps based on revisions and stakeholder feedback, highlighted in Table 18 and Table 19.

Table 18: Module Steps for Calculating 7Q10 at Gages

Steps	Description & Documentation
Step 1: Download Streamflow Data	The user can download the current continuous streamflow data for any gage from the USGS's Current Water Data for the Nation (https://waterdata.usgs.gov/nwis/rt). This uses the USGS' data Retrieval package (https://www.usgs.gov/software/dataretrieval) to retrieve the full record of streamflow for a chosen gage, beginning with the earliest available date of flow to the most recent date available. For gages that are still operational, the most recent date of streamflow available is the previous day's data (average daily flow). For

	gages that have been discontinued, the last date available is the date on which the gage was discontinued.
Step 2: Display Streamflow	As requested by stakeholders, this tab displays the raw, unrefined streamflow data for the gage. This plot uses the R Package ggplot2 (https://ggplot2.tidyverse.org/) and allows the user to change the x and y scales of the graph using the sliderInput function in RShiny to analyze certain time periods and/or flow ranges. (https://shiny.posit.co/r/reference/shiny/0.11/sliderinput)
Step 3: Display Streamflow (log scale)	Similar to step 2, this tab displays the raw, unrefined streamflow data for the gage, but with the y axis (streamflow, in cfs) on a log scale. As requested by stakeholders, it can be difficult to display both high and low flows on the same plot, so this plot allows the user to better visualize both high and low flows in the same figure. This tab uses the same functions as step 2.
Step 4: Plot Low Flows	Though we display the full streamflow in steps 2 and 3, this tool is primarily designed for low flow analysis. In step 4, we plot the lowest 10th percentile flows, along with the 7Q2, 7Q10, and 7Q30 calculated on the full record of streamflow data for the gage. The “fasstr” software package (https://cran.r-project.org/web/packages/fasstr/index.html) is used to calculate these values.
Step 5: Mann-Kendall Test	In this tab, we follow the procedure in Blum et al. (2019) and test the annual 7-day low flows for trends using the Mann-Kendall test (Mann, 1945; Kendall 1975; Davison and Hinkley, 1997; Hipel and McLeod, 2005) using the MannKendall function from the Kendall R Package (https://search.r-project.org/CRAN/refmans/Kendall/html/MannKendall.html). Additionally, a plot of the annual 7-day low flows with a lowess smoothing curve from the R Package gplots (https://www.rdocumentation.org/packages/gplots/versions/3.1.3/topics/lowess) is fit to the annual 7-day low flows with the default parameters.
Step 6: Subset Streamflow	In the first prototype, the last 30 years of daily streamflow data was automatically subset from the full record of streamflow data and the 7Q30 was recalculated using the last 30 years of streamflow data. Based on stakeholder input, the final DSS allows the user to select a beginning and end date to subset the streamflow data between. This gives the user the flexibility to choose any subset of streamflow data that they require, which also includes the last 30 years of streamflow data if they deem it necessary to use. This tab does not use any additional R packages.
Step 7: Plot Low Flows for Subset Streamflow	In this tab, the low flow graphic, similar to the one in step 4, is displayed for the user-selected subset. Additionally, stakeholders requested that the full-record 7Q2, 7Q10, and 7Q30 be displayed on this graph as partially transparent lines so that the new values, calculated on the user-selected subset, could be compared to the full-record values.
Step 8: Flow Scaling	Stakeholders identified that for ungaged locations near a gage, particularly for those up or downstream of a gage with watershed areas that are within a 50% tolerance of the gaged watershed, they sometimes use a simple area/area scaling ratio to calculate the up or downstream 7Q10 value, similar to the methodology presented in Asquith and Thompson (2008). In this tab, a simple interface is provided to scale all flows, where the user only needs to enter the up or downstream watershed area and a new set of values will be calculated. This assumes that runoff is homogenous, and that daily and low flows operate on a strictly area/area scale. Users should verify that this method is appropriate for their watershed before use.

Table 19: Module Steps for Ungaged Basin 7Q10 Estimation

Steps	Description & Documentation
Step 1: Background Information	In this tab, an explanation of how this tool works and other ways to retrieve the physical characteristics of the watershed needed for regression is included. This tab is the result of stakeholder input, as demonstrations of the tool supported that a full summary of the methodology was useful for current and future users of the tool. This tab does not use any data or methods.
Step 2: Contact StreamStats	In this tab, the user can attempt to contact StreamStats to retrieve the physical characteristics of the watershed using the StreamStats API (https://streamstats.usgs.gov/docs/streamstatsservices/#/). This process delineates the watershed that corresponds to a user-selected latitude and longitude, which is shown on a map, created using the Leaflet package (https://leafletjs.com/). Based on stakeholder input, a pop-up box is included to let the user know the program is not frozen while it attempts to contact StreamStats. This utilizes the “withProgress” function in Shiny (https://shiny.posit.co/r/reference/shiny/0.11/withprogress). After retrieval, the watershed area is presented to the user below the latitude and longitude.
Step 3: Display Watershed and Physical Characteristics	In this tab, the user is presented with a map of the delineated watershed and the available physical characteristic retrieved directly from StreamStats. This map once again utilizes the leaflet package, and no additional work is required from the user.
Step 4: Regression-Based 7Q10 Estimation	In this tab, the user can use regional regression equations to estimate the 7Q10. This tool is built on previous research that tested multiple machine learning methods to develop a regional methodology for statistical 7Q10 estimation in the Northeast United States (DeSanto et al., 2023). Three statistical regression methods (multiple linear regression, logarithmic regression, and the random forest machine learning algorithm) are used for estimation. These equations are trained on 1) the full-record of streamflow available at 165 locations throughout the northeast and mid-Atlantic United States, and 2) the last 30-years of streamflow data at the same locations, for a total of 6 estimates.
Step 5: Climate-Altered 7Q10	In this tab, the user can estimate how a changing climate, using altered precipitation and temperature values, will impact their 7Q10 value. Recommended by multiple stakeholders, this follows the procedure from the Transportation Research Board’s (TRB) National Cooperative Highway Research Program’s (NCHRP) <i>Project 15-61: Applying Climate Change Information to Hydrologic and Hydraulic Design of Transportation Infrastructure</i> (Kilgore et al., 2019). This applies a regression equation from the previous step with altered climate variables to estimate how an altered climate will impact the 7Q10. The full-record logarithmic regression equation was selected for use in this step, as it was found to be the most theoretically appropriate for extreme flow estimation (Tasker and Stedinger, 1989). Additional information on this methodology is included in Appendix C.

4.6 Continued Maintenance and Updates

Currently, plans for maintenance and future updates are under refinement. Given the interactions thus far with the user group, there is optimism that the DSS will provide opportunities and motivation potential for integrating this software into the user group's existing protocols for estimating hydrologic low flows. Anticipated routine maintenance and methodological additions to the program is presented in Figure 35.

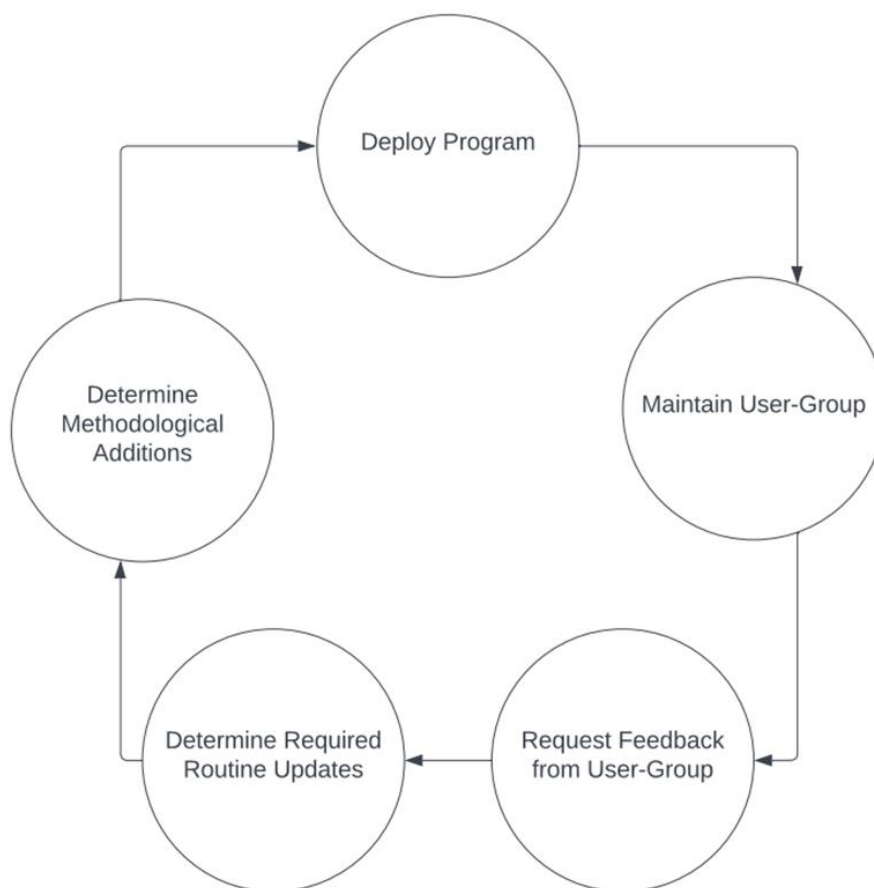


Figure 35: Stakeholder-driven plans for maintaining the program.

We plan to provide routine maintenance in the case that any data sources are discontinued or if the existing software that is utilized by the DSS is substantially modified. Short-term, anticipated modifications to the DSS that will enhance its likely implementation include:

1) Assist users in determining whether to use streamgage data and flow scaling to calculate a 7Q10 versus when to use purely ungaged 7Q10 estimation, especially when the location of interest is on a stream with a gage nearby. Resource managers reiterated that they will typically use data from a streamgage combined with flow scaling when the location of interest is near a streamgage. Similar to the findings in Asquith and Thompson (2008), members of the user group typically only use the streamgage data with flow scaling if the ratio of the ungaged watershed size to the gaged basin size is within a 50% tolerance, smaller or larger. This assumes that runoff is homogenous throughout the basin and the 7Q10 value, calculated from the gaged streamflow data, is multiplied by a ratio of the ungaged watershed area to the gaged basin area to scale up or down the value. Most mentioned that using a 50% tolerance is relatively arbitrary and they would like more definitive ways to determine if runoff should be considered homogenous or if they should utilize purely ungaged 7Q10 estimation.

2) Inserting a module for incorporating rainfall-runoff modeling directly into the program. Some stakeholders reiterated the use of rainfall-runoff modeling for incorporating future climate projections directly into the water budget, but that it would mostly be used in parallel to (or to provide further justification for) the results projected by the statistical regression model.

3) Providing state-by-state estimates of projected climate changes for users, rather than the regional projected changes provided in the program currently.

4.7 Summary

This chapter highlights the development of a DSS for calculating, estimating, and projecting future 7Q10s. At the time of publication, there are few trusted tools for

resource managers to incorporate climate change projections into low flow calculation and estimation. This DSS allows for future climate projections to be incorporated into ungaged 7Q10 estimation and offers numerous advantages that can greatly benefit researchers, water resource managers, and policymakers who require 7Q10s for planning and permitting purposes. This tool was created in collaboration with resource managers in the region who provided recommendations for the methodology and interface. This process allowed the developers to better understand the motivations of decision makers in calculating 7Q10, the ways in which these calculations influence decision making, the time and data limitations encountered by those tasked with making these calculations, and how this process could be improved with a decision support system. We suggest that engaging our user-group throughout the process will encourage those who helped with its development to implement it in their professional duties. Future work to improve the DSS is planned, including additions to the flow scaling methodology to assist users in deciding if the watershed is homogenous, the addition of physical modeling for projecting future changes in 7Q10s in parallel to the statistical projections presented here, and state-by-state estimates of projected climate changes rather than the regional suggestions currently provided.

CHAPTER 5

CONCLUSIONS

The research presented in the three previous chapters advances hydrologic research by presenting a technique to estimate the 7Q10 using machine learning (Chapter 2), illustrating the value of physically-based hydrologic models to estimate 7Q10 by simulating daily flows at ungaged sites (Chapter 3), and creating a DSS to project changes in the 7Q10 that resource managers can incorporate in hydrologic design decisions (Chapter 4). In Chapter 2, traditional statistical methods for 7Q10 estimation are demonstrated to perform well for both larger watersheds (multiple linear regression) or smaller watersheds (multiple linear regression in log-space). Three machine learning algorithms (neural networks, random forests, and generalized additive models) are evaluated compared to the classical methods for improved 7Q10 estimation. Results suggest that neural networks and generalized additive models suffer from major drawbacks, but the random forest algorithm maintains the smallest error for all range of basin sizes, performing similarly to multiple linear regression in larger watersheds and multiple linear regression in log space for smaller watersheds. The Random Forest methodology can be used as an alternative to traditional methods, especially in states where StreamStats has not been developed or is limited to a certain range of watershed sizes. Future work in this domain includes testing other advanced statistical methods and/or machine learning algorithms for better 7Q10 estimation. Additionally, the Random Forest methodology is shown to work well for a large geographical footprint, so future work may include determining the boundaries at which assuming hydrologic homogeneity is no longer satisfied.

Chapter 3 tests the efficacy and suitability of a novel machine learning algorithm, Fuzzy C-Means clustering, for the calibration of a physical rainfall-runoff model for ungaged locations. This method allows the soft clustering of variables, providing additional flexibility compared to hard clustering algorithms like K-Means. The analysis supports that K-Means provide clusters with better Silhouette scores, suggesting that hard clusters may be more appropriate than soft clusters. This was confirmed during modeling, as we found that models calibrated with K-Means clustering provided the best average KGE compared to the models calibrated with Fuzzy C-Means. In the optimal case, calibration improved daily streamflow estimation significantly, almost providing the same average KGE as the models individually calibrated to measured streamflow.

In addition, it was determined that even with the improvement for daily streamflow estimation, all the models calibrated with clustering (both K-Means and Fuzzy C-Means) were unable to provide similar errors for 7Q10 estimation compared to StreamStats' 7Q10 estimates for the same watersheds. In the case of a future climate change study where the accuracy of the value itself is less important than model-predicted future changes in the value, the physically based estimates readily allow for the incorporation of altered climate and/or landcover projections. Given few alternatives for estimating 7Q10 at ungaged locations efficiently, the physically based models also provide the ability for resource managers to estimate 7Q10s in states where StreamStats 7Q10 estimation is not yet developed. Future work related to this chapter could include testing this procedure with other physically based rainfall-runoff models, determining other ways to calibrate physical models in ungaged locations for daily streamflow and/or

low flow estimation, and further analyzing the performance of physical models for gaged and ungaged high flow estimation as well.

In Chapter 4, we incorporate methods presented in Chapters 2 and 3 into a stakeholder driven decision support system webtool. After the initial interviews described in the introduction, resource managers in the northeast U.S. were invited to participate in the development of a decision support system webtool to improve 7Q10 estimation in ungaged basins. The participating resource managers assisted in the framework development process by highlighting major and minor changes that could be made to the evolving DSS tool that would improve their ability to properly estimate 7Q10s in ungaged basins. The managers contributed to the overall structure of the program and determined what elements should be incorporated into the graphical user interface to facilitate decision making. Some of the major improvements in the DSS suggested by the managers include: 1) flexibility in defining streamflow data used to calculate the 7Q10, 2) the creation of regression equations incorporating only the last 30 years of streamflow data at every location, and 3) the ability to incorporate future climate change projections into the ungaged 7Q10 estimation. Future work is planned to improve the DSS, including additions to the flow scaling methodology to assist users in deciding if the watershed is homogenous, the addition of physical modeling for projecting future changes in 7Q10s in parallel to the statistical projections presented here, and state-by-state estimates of projected climate changes rather than the regional suggestions currently provided.

This dissertation contributes to the extensive collection of literature pertaining to low flow estimation in ungaged basins. Despite the scientific advancements highlighted in this research, there are still many areas of research needed to cope with the wide range

of impacts associated with low flows in streams. Improvements to current drought calculation/estimation methods and the forethought to promote their adoption in practice must be further studied and encouraged. Research into climate change's impacts on drought events must be conducted prior to the projected increases in earth's global average temperature and the associated impacts on the hydrologic cycle.

APPENDIX A

Station	Source	State	Station Name	Watershed Area (mi ²)	7Q10 (cfs)
01013500	HCDN	ME	Fish River near Fort Kent, Maine	869.8	79.67
01022500	HCDN	ME	Narraguagus River at Cherryfield, Maine	221.5	29.24
01030500	HCDN	ME	Mattawamkeag River near Mattawamkeag, Maine	1419.4	59.16
01031500	HCDN	ME	Piscataquis River near Dover-Foxcroft, Maine	296.9	15.54
01047000	HCDN	ME	Carrabassett River near North Anson, Maine	351	44.96
01052500	HCDN	NH	Diamond River near Wentworth Location, NH	148.2	16.95
01054200	HCDN	ME	Wild River at Gilead, Maine	69.9	9.59
01055000	HCDN	ME	Swift River near Roxbury, Maine	96.8	6.82
01057000	HCDN	ME	Little Androscoggin River near South Paris, Maine	73.7	2.39
01073000	HCDN	NH	Oyster River near Durham, NH	12.1	0.35
01073860	HCDN	MA	Small Pox Brook at Salisbury, MA	1.83	0.15
01078000	HCDN	NH	Smith River near Bristol, NH	85.9	6.04
01094340	MA Low Flow Report	MA	Whitman River near Westminster, Mass.	21.7	0.89
01094396	MA Low Flow Report	MA	Philips Brook at Fitchburg, Mass.	15.8	0.34
01094760	MA Low Flow Report	MA	Wausacum Brook near West Boylston, Mass.	7.41	0.06
01095220	MA Low Flow Report	MA	Stillwater River near Sterling, Mass.	30.4	1.06
01095380	MA Low Flow Report	MA	Trout Brook near Holden, Mass.	6.79	0.05
01095915	MA Low Flow Report	MA	Mulpus Brook near Shirley, Mass.	15.7	0.39
01095928	MA Low Flow Report	MA	Trapfall Brook near Ashby, Mass.	5.89	0.02
01096000	MA Low Flow Report	MA	Squannacook River near West Groton, MA	64.4	6.52
01096504	MA Low Flow Report	MA	Reedy Meadow Brook at East Pepperell, Mass.	1.92	0.24
01096505	MA Low Flow Report	MA	Unkety Brook near Pepperell, Mass.	6.84	0.46
01096515	MA Low Flow Report	MA	Salmon Brook at Main Street at Dunstable, Mass.	18.2	2.34
01096805	MA Low Flow Report	MA	North Brook near Berlin, Mass.	15.4	0.54
01096855	MA Low Flow Report	MA	Danforth Brook at Hudson, Mass.	6.62	0.14

01096935	MA Low Flow Report	MA	Elizabeth Brook at Wheeler Street at Stow, Mass.	17.2	0.76
01097280	MA Low Flow Report	MA	Fort Pond Brook at West Concord, Mass.	24.9	0.89
01097300	MA Low Flow Report	MA	Nashoba Brook near Acton, MA	12.9	0.12
01099400	MA Low Flow Report	MA	River Meadow Brook at Lowell, Mass.	25.6	0.98
01100608	MA Low Flow Report	MA	Meadow Brook near Tewksbury, Mass.	4.09	0.15
01100700	MA Low Flow Report	MA	East Meadow River near Haverhill, MA	5.54	0.15
01101100	MA Low Flow Report	MA	Mill River near Rowley, Mass.	7.7	0.39
01102490	MA Low Flow Report	MA	Shaker Glen Brook near Woburn, Mass.	3.05	0.17
01103015	MA Low Flow Report	MA	Mill Brook at Arlington, Mass.	5.35	0.38
01103253	MA Low Flow Report	MA	Chicken Brook near West Medway, Mass.	7.23	0.18
01103435	MA Low Flow Report	MA	Waban Brook at Wellesley, Mass.	10.2	0.13
01103440	MA Low Flow Report	MA	Fuller Brook at Wellesley, Mass.	3.91	0.11
01104960	MA Low Flow Report	MA	Germany Brook near Norwood, Mass.	2.37	0.08
01104980	MA Low Flow Report	MA	Hawes Brook at Norwood, Mass.	8.64	0.29
01105568	MA Low Flow Report	MA	Cochato River at Holbrook, Mass.	4.31	0.09
01105575	MA Low Flow Report	MA	Cranberry Brook at Braintree Highlands, Mass.	1.72	0.01
01105600	MA Low Flow Report	MA	Old Swamp River near South Weymouth, MA	4.47	0.16
01105630	MA Low Flow Report	MA	Crooked Meadow River near Hingham Center, Mass.	4.91	0.27
01105820	MA Low Flow Report	MA	Second Herring Brook at Norwell, Mass.	3.17	0.03
01105830	MA Low Flow Report	MA	First Herring Brook near Scituate Center, Mass.	1.72	0.01
01105861	MA Low Flow Report	MA	Jones River Brook near Kingston, Mass.	4.74	0.49
01105930	MA Low Flow Report	MA	Paskamanset River at Turner Pond near New Bedford,	8.09	0.32
01105937	MA Low Flow Report	MA	Shingle Island River near North Dartmouth, Mass.	8.59	0.06
01105947	MA Low Flow Report	MA	Bread and Cheese Brook at Head of Westport, Mass.	9.25	0.14
01106000	MA Low Flow Report	RI	Adamsville Brook at Adamsville, RI	7.99	0.05

01106460	MA Low Flow Report	MA	Beaver Brook near East Bridgewater, Mass.	8.94	0.34
01107400	MA Low Flow Report	MA	Fall Brook near Middleboro, Mass.	9.3	1.32
01108180	MA Low Flow Report	MA	Cotley River at East Taunton, Mass.	7.48	0.47
01108600	MA Low Flow Report	MA	Hodges Brook at West Mansfield, Mass.	3.83	0.03
01109087	MA Low Flow Report	MA	Assonet River at Assonet, Mass.	20.7	0.62
01109090	MA Low Flow Report	MA	Rattlesnake Brook near Assonet, Mass.	4.22	0.11
01109225	MA Low Flow Report	MA	Rocky Run near Rehoboth, Mass.	7.21	0.07
01109460	MA Low Flow Report	MA	Dark Brook at Auburn, Mass.	11.1	0.94
01111200	MA Low Flow Report	MA	West River below West Hill Dam, near Uxbridge, MA	27.8	1.80
01111225	MA Low Flow Report	MA	Emerson Brook near Uxbridge, Mass.	7.26	0.63
01111300	MA Low Flow Report	RI	Nipmuc River near Harrisville, RI	16	0.25
01112190	MA Low Flow Report	MA	Muddy Brook at South Milford, Mass.	6.17	0.14
01117468	HCDN	RI	Beaver River near Usquepaug, RI	8.87	1.78
01118300	HCDN	CT	Pendleton Hill Brook near Clarks Falls, CT	4	0.02
01121000	HCDN	CT	Mount Hope River near Warrenville, CT	27.1	0.65
01123000	HCDN	CT	Little River near Hanover, CT	30.1	4.36
01123140	MA Low Flow Report	MA	Mill Brook at Brimfield, Mass.	13.8	1.29
01123200	MA Low Flow Report	MA	Stevens Brook at Holland, Mass.	4.39	0.09
01124390	MA Low Flow Report	MA	Little River at Richardson Corners, Mass.	8.58	0.20
01134500	HCDN	VT	Moose River at Victory, VT	75.3	5.97
01137500	HCDN	NH	Ammonoosuc River at Bethlehem Junction, NH	88.2	27.36
01139000	HCDN	VT	Wells River at Wells River, VT	95.1	14.12
01139800	HCDN	VT	East Orange Branch at East Orange, VT	8.8	0.71
01142500	HCDN	VT	Ayers Brook at Randolph, VT	31.7	2.11
01144000	HCDN	VT	White River at West Hartford, VT	691.2	90.51
01150900	HCDN	VT	Ottauquechee River near West Bridgewater, VT	23.3	3.31
01162500	HCDN	MA	Priest Brook near Winchendon, MA	19.2	0.47
01162900	MA Low Flow Report	MA	Otter River at Gardner, Mass.	19.2	2.57

01164300	MA Low Flow Report	MA	Lawrence Brook at Royalston, Mass.	15.6	0.32
01167200	MA Low Flow Report	MA	Fall River at Bernardston, Mass.	22.3	1.46
01168300	MA Low Flow Report	MA	Cold River near Zoar, Mass.	29.6	1.69
01168400	MA Low Flow Report	MA	Chickley River near Charlemont, Mass.	27.1	3.24
01169000	HCDN	MA	North River at Shattuckville, MA	89.1	8.82
01170100	HCDN	MA	Green River near Colrain, MA	41.3	4.73
01181000	HCDN	MA	West Branch Westfield River at Huntington, MA	94	6.03
01187300	HCDN	MA	Hubbard River near West Hartland, CT	20.8	0.45
01195100	HCDN	CT	Indian River near Clinton, CT	5.68	0.02
01208990	HCDN	CT	Saugatuck River near Redding, CT	20.8	0.31
01333000	HCDN	MA	Green River at Williamstown, MA	43.3	4.80
01350000	HCDN	NY	Schoharie Creek at Prattsville NY	236.5	9.47
01350080	HCDN	NY	Manor Kill at West Conesville near Gilboa, NY	32.4	1.67
01350140	HCDN	NY	Mine Kill near North Blenheim, NY	16.9	0.22
01362200	HCDN	NY	Esopus Creek at Allaben, NY	63.7	5.48
01365000	HCDN	NY	Rondout Creek near Lowes Corners, NY	38.4	6.26
01411300	HCDN	NJ	Tuckahoe River at Head of River, NJ	30.6	6.70
01413500	HCDN	NY	East Brook Delaware River at Margaretville, NY	163.7	11.17
01414500	HCDN	NY	Mill Brook near Dunraven, NY	24.9	2.26
01415000	HCDN	NY	Tremper Kill near Andes, NY	33.1	1.59
01423000	HCDN	NY	West Branch Delaware River at Walton, NY	331.9	23.49
01434025	HCDN	NY	Biscuit Brook above Pigeon Brook at Frost Valley, NY	3.72	0.45
01435000	HCDN	NY	Neversink River near Claryville, NY	66.6	13.99
01439500	HCDN	PA	Bush Kill at Shoemakers, PA	118.1	7.75
01440000	HCDN	NJ	Flat Brook near Flatbrookville, NJ	64.8	7.53
01440400	HCDN	PA	Brodhead Creek near Analomink, PA	67.6	7.60
01451800	HCDN	PA	Jordan Creek near Schnecksville, PA	52.4	2.78
01466500	HCDN	NJ	McDonalds Branch in Lebanon State Forest, NJ	2.1	0.82
01484100	HCDN	DE	Beaverdam Branch at Houston, DE	3.5	0.13
01485500	HCDN	MD	Nassawango Creek near Snow Hill, MD	54.6	1.17
01486000	HCDN	MD	Manokin Branch near Princess Anne, MD	4.3	0.04
01487000	HCDN	DE	Nanticoke River near Bridgeville, DE	72.4	15.99
01491000	HCDN	MD	Choptank River near Greensboro, MD	112.8	4.09
01510000	HCDN	NY	Otselic River at Cincinnatus, NY	147.9	9.42
01516500	HCDN	PA	Corey Creek near Mainesburg, PA	12.2	0.01
01518862	HCDN	PA	Cowanesque River at Westfield, PA	90.6	1.45

01532000	HCDN	PA	Towanda Creek near Monroeton, PA	213.9	2.96
01539000	HCDN	PA	Fishing Creek near Bloomsburg, PA	271	17.83
01542810	HCDN	PA	Waldy Run near Emporium, PA	5.3	0.09
01543000	HCDN	PA	Driftwood Br Sinnemahoning Cr at Sterling Run, PA	272.4	4.51
01543500	HCDN	PA	Sinnemahoning Creek at Sinnemahoning, PA	686.6	15.50
01544500	HCDN	PA	Kettle Creek at Cross Fork, PA	137.1	5.07
01545600	HCDN	PA	Young Womans Creek near Renovo, PA	46.3	1.60
01547700	HCDN	PA	Marsh Creek at Blanchard, PA	43.8	0.66
01548500	HCDN	PA	Pine Creek at Cedar Run, PA	601.2	24.84
01549500	HCDN	PA	Blockhouse Creek near English Center, PA	37.6	0.79
01550000	HCDN	PA	Lycoming Creek near Trout Run, PA	174.8	7.84
01552000	HCDN	PA	Loyalsock Creek at Loyalsockville, PA	436.1	23.63
01552500	HCDN	PA	Muncy Creek near Sonestown, PA	23.4	1.19
01557500	HCDN	PA	Bald Eagle Creek at Tyrone, PA	44.5	3.19
01564500	HCDN	PA	Aughwick Creek near Three Springs, PA	205	4.41
01567500	HCDN	PA	Bixler Run near Loysville, PA	15	2.32
01568000	HCDN	PA	Sherman Creek at Shermans Dale, PA	206.3	15.95
01580000	HCDN	MD	Deer Creek at Rocks, MD	94.4	23.91
01583500	HCDN	MD	Western Run at Western Run, MD	60.2	11.57
01586610	HCDN	MD	Morgan Run near Louisville, MD	26	3.63
01591400	HCDN	MD	Cattail Creek near Glenwood, MD	22.8	1.71
01594950	HCDN	MD	McMillan Fort near Fort Pendleton, MD	2.3	0.00
01596500	HCDN	MD	Savage River near Barton, MD	48.1	1.03
01605500	HCDN	WV	South Branch Potomac River at Franklin, WV	179.1	20.64
01606500	HCDN	WV	South Branch Potomac River near Petersburg, WV	650.4	54.55
01613050	HCDN	PA	Tonoloway Creek near Needmore, PA	10.8	0.00
01620500	HCDN	VA	North River near Stokesville, VA	17.3	0.23
01632000	HCDN	VA	N F Shenandoah River at Cootes Store, VA	209.8	0.84
01632900	HCDN	VA	Smith Creek near New Market, VA	94.6	7.64
01634500	HCDN	VA	Cedar Creek near Winchester, VA	101.9	4.67
01638480	HCDN	VA	Catoctin Creek at Taylorstown, VA	89.6	0.66
01639500	HCDN	MD	Big Pipe Creek at Bruceville, MD	103.2	8.10
01644000	HCDN	VA	Goose Creek near Leesburg, VA	331.7	2.01
01658500	HCDN	VA	S F Quantico Creek near Independent Hill, VA	7.5	0.00
01664000	HCDN	VA	Rappahannock River at Remington, VA	619.7	10.76
01666500	HCDN	VA	Robinson River near Locust Dale, VA	178.8	9.18
01667500	HCDN	VA	Rapidan River near Culpeper, VA	467.1	17.08

01669000	HCDN	VA	Piscataway Creek near Tappahannock, VA	27.7	0.38
01669520	HCDN	VA	Dragon Swamp at Mascot, VA	109	0.01
02011400	HCDN	VA	Jackson River near Bacova, VA	157.4	17.06
02011460	HCDN	VA	Back Creek near Sunrise, VA	60.4	2.01
02013000	HCDN	VA	Dunlap Creek near Covington, VA	164	10.68
02014000	HCDN	VA	Potts Creek near Covington, VA	153.2	17.64
02015700	HCDN	VA	Bullpasture River at Williamsville, VA	110.2	26.07
02016000	HCDN	VA	Cowpasture River near Clifton Forge, VA	461.2	57.01
02017500	HCDN	VA	Johns Creek at New Castle, VA	106.6	7.72
02018000	HCDN	VA	Craig Creek at Parr, VA	329.1	30.75
02027000	HCDN	VA	Tye River near Lovingston, VA	93	4.03
02027500	HCDN	VA	Piney River at Piney River, VA	47.6	2.60
02028500	HCDN	VA	Rockfish River near Greenfield, VA	94.8	2.50
02038850	HCDN	VA	Holiday Creek near Andersonville, VA	8.5	0.33

APPENDIX B

Organization	Job Title(s)	Initial Interview Date
Maine Department of Transportation	Hydrologist/Hydraulic Engineer	5/11/2023
VT Agency of Natural Resources	Hydrologist	3/23/2023
VT Agency of Transportation	Hydraulics Project Engineer	5/12/2023
NH Department of Environmental Conservation	NPDES Permitting, Instream Flow Specialists, Permits Engineer	7/9/2021
NH Department of Transportation	DOT Group Leader	5/12/2023
MA Wildlife	Climate Change Specialist	8/5/2021
MA Department of Transportation	Manager of Sustainability	5/11/2023
CT Department of Energy & Environmental Protection	Supervising Fisheries Biologist, Environmental Analyst	3/20/2023

CT Department of Transportation	Transportation Principal Engineer	5/9/2023
NY Fish and Wildlife	Biologist 1, Ecology	3/14/2023
EPA Region 1	NPDES Permitting	1/11/2022
Maryland Department of Natural Resources	Manager of Fisheries	7/21/2021
Georgia Department of Natural Resources	Instream Flow Protection	7/9/2021
Idaho Water Board	Board Member & Lawyer	7/21/2021
Montana Fish, Wildlife, and Parks	Water Resources Specialist	7/15/2021
Washington Department of Ecology	Program Manager	7/9/2021

APPENDIX C

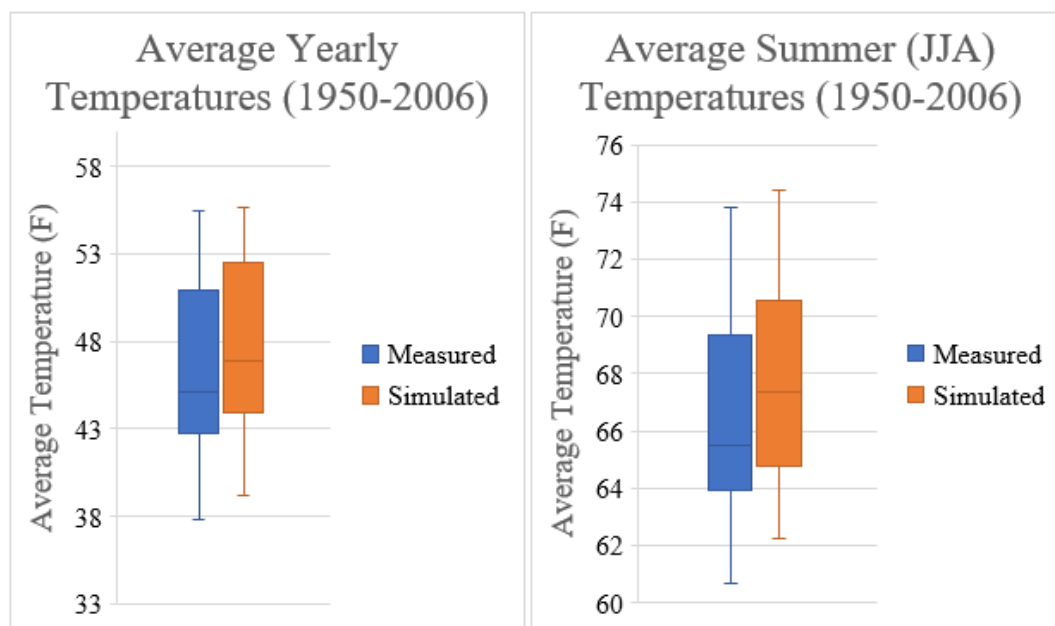
Appendix C.1 – Verification of GCM Data

Future climate projections, in the form of nine downscaled Global Circulation Models (GCMs) for RCP scenarios 4.5 and 8.5, were retrieved from <https://climate.northwestknowledge.net/MACA/index.php>. The MACAv2-METDATA dataset (Abatzoglou, 2011), which utilizes Multivariate Adaptive Constructive Analogs (MACA) (Abatzoglou and Brown, 2012) downscaling was selected for use, as it has been shown to outperform the Localized Constructive Analogs (LOCA) (Pierce et al., 2014) downscaling method for capturing extreme events (Wang et al., 2020). Climate forcings in the MACAv2-METDATA were drawn from a statistical downscaling of GCM data from the Coupled Model Intercomparison Project 5 (CMIP5, Taylor et al. 2010) utilizing a modification of the Multivariate Adaptive Constructed Analogs (MACA, Abatzoglou

and Brown, 2012) method with the METDATA (Abatzoglou, 2011) observational dataset as training data. In the table below, the nine GCMs selected for use are listed, as they were found to be most appropriate for the study area (Karmalkar et al., 2019).

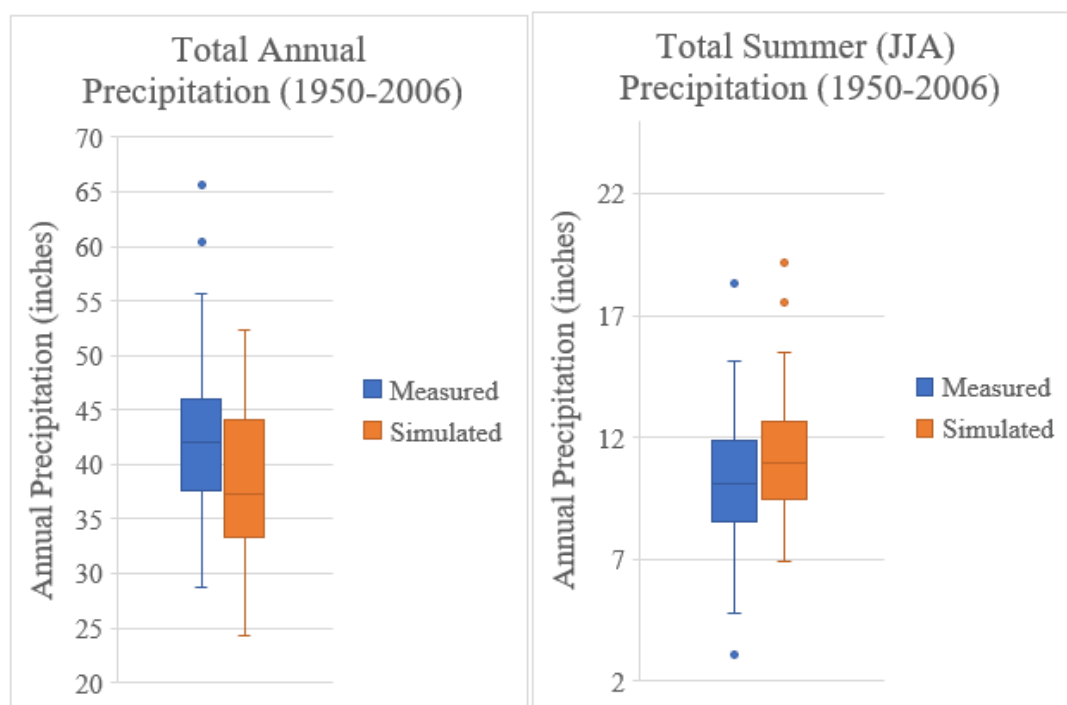
Model Name	Model Agency
<u>bcc-csm1-1</u>	Beijing Climate Center, China Meteorological Administration
<u>bcc-csm1-1-m</u>	Beijing Climate Center, China Meteorological Administration
<u>CanESM2</u>	Canadian Centre for Climate Modeling and Analysis
<u>CNRM-CM5</u>	National Centre of Meteorological Research, France
<u>CSIRO-Mk3-6-0</u>	Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence, Australia
<u>GFDL-ESM2M</u>	NOAA Geophysical Fluid Dynamics Laboratory, USA
<u>HadGEM2-ES</u>	Met Office Hadley Center, UK
<u>inmcm4</u>	Institute for Numerical Mathematics, Russia
<u>IPSL-CM5A-LR</u>	Institut Pierre Simon Laplace, France

After retrieval, the GCMs were evaluated to ensure that their simulated historic estimates provide similar results to the observed, historic data at each gage, providing more confidence in the future climate projections. In Figures 1a and 1b, the GCM predictions at each gage used in the training of the regression equations were averaged for both yearly temperatures and summer (June, July, August, denoted JJA) temperatures for the entire length of the historic, simulated dataset (1950-2006). These values were compared to the observed temperatures at the same locations for the same period from the Livneh (2015) dataset.



Figures 1a (left) and 1b (right): Ensemble average temperatures predicted by the nine GCMs at each gage for 1950-2006 compared to the historic, measured temperatures for the same locations.

The median simulated temperatures are slightly higher in both Figure 1a and 1b than the observed temperatures for both the yearly and summer averages, but the full range of values and Interquartile Range (IQR) remain similar. In Figures 2a and 2b, the same plots are displayed for precipitation totals at each gage.

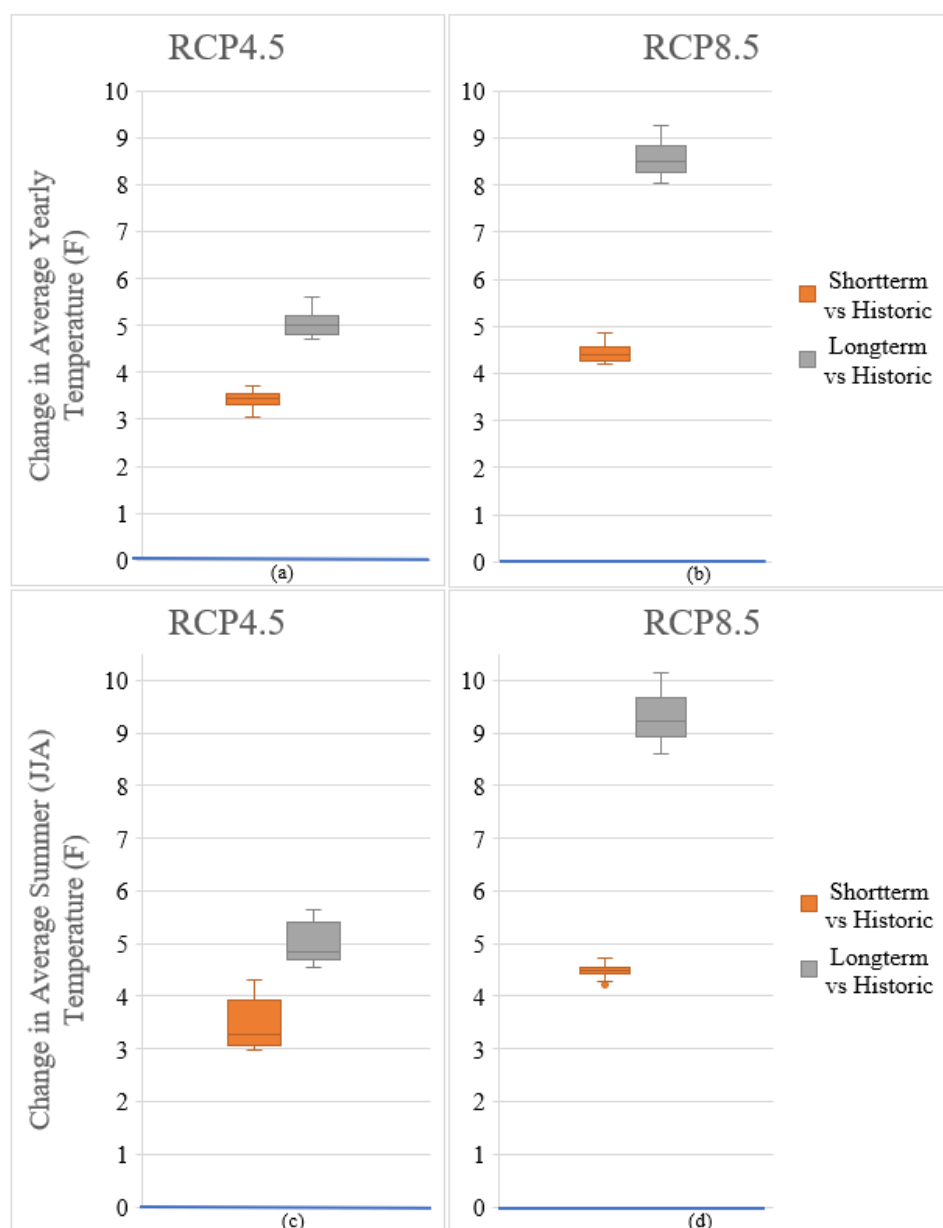


Figures 2a (left) and 2b (right): Ensemble average precipitation totals predicted by the nine GCMs at each gage for 1950-2006 compared to the historic, measured precipitation totals for the same locations.

The annual precipitation totals in Figure 2a show that the GCMs are underestimating the total amount of yearly precipitation, while the summer totals shown in Figure 2b demonstrate that the GCMs are slightly overestimating the JJA precipitation totals. Regardless, Figures 1a-b and 2a-b provide support that the climate variables from the GCMs provide similar temperature averages and precipitation totals to the measured historic data from 1950-2006 for both the year and summer.

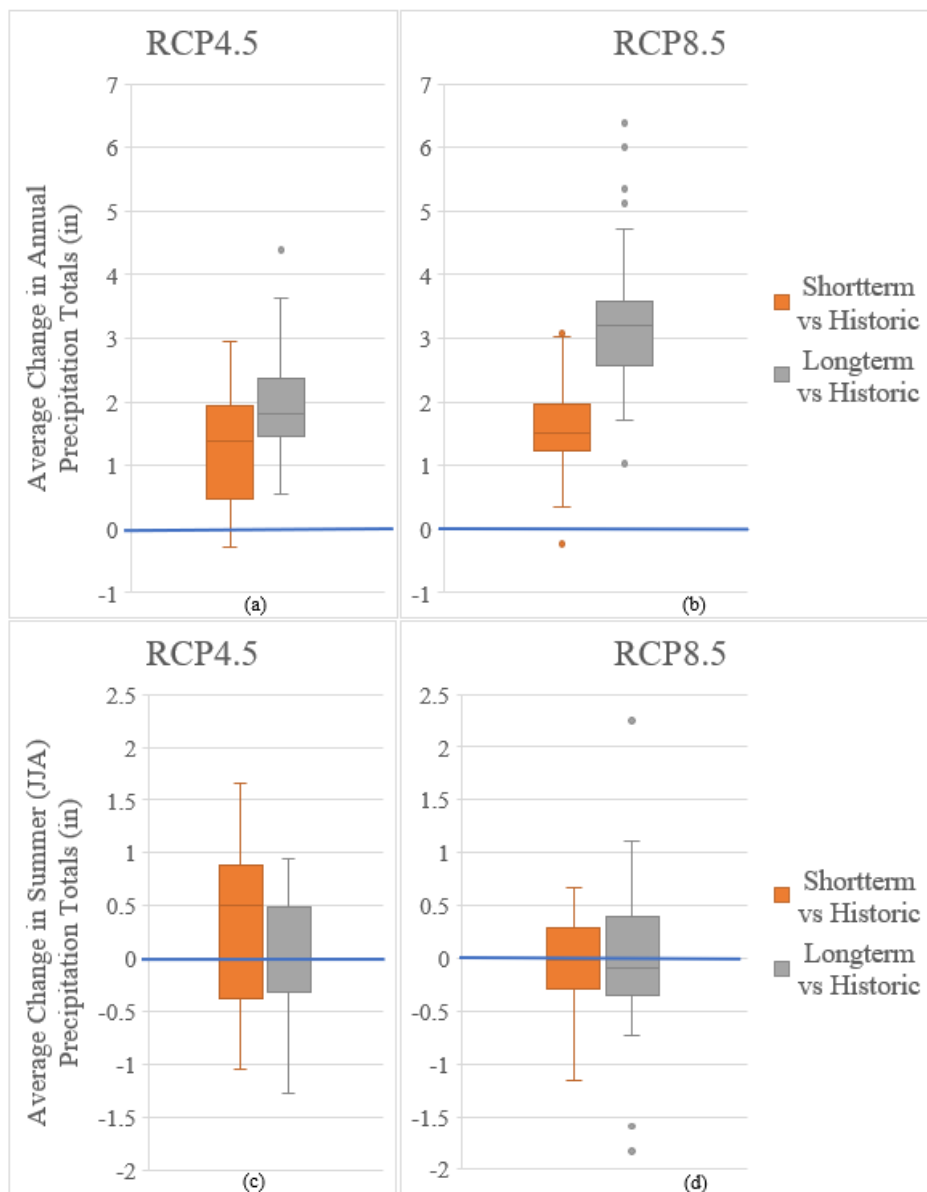
Suggested by some resource managers, the future GCM projections were split into the short-term (2020-2060) and long-term (2060-2100) futures for both RCP4.5 and RCP8.5 scenarios. They highlighted the commonality of this in literature and the large differences that may arise in the next 30-40 years versus the following 30-40 years. In

Figures 3a-d, future short-term and long-term temperature change projections are given for both RCP scenarios, compared to the historic climate values at each gage.



Figures 3a-d: Ensemble average temperature projections from the nine GCMs for both RCP4.5 and RCP8.5 scenarios, separated by short-term (2020-2060) and long-term (2060-2100) projections, at each gage.

Figures 3a-d demonstrate that the projected temperature increases greatly vary depending on the RCP scenario and time period, just as resource managers predicted. Additionally, precipitation totals for the same periods and RCP scenarios are provided in Figures 4a-d.



Figures 4a-d: Ensemble average precipitation projections from the nine GCMs for both RCP4.5 and RCP8.5 scenarios, separated by short-term (2020-2060) and long-term (2060-2100) projections, at each gage.

Figures 4a-b demonstrate that the GCMs predict increases in yearly precipitation totals throughout the study area, regardless of the RCP scenario. However, Figures 4c-d demonstrate that the GCMs are not predicting increases during the summer months besides for the short-term, RCP4.5 scenario. The other three scenarios and time periods display median precipitation totals that are either equal to the observed historic total (RCP4.5, long-term and RCP8.5 short-term) or less (RCP8.5, long-term).

Appendix C.2 - Extraction and Verification of Climate Variables

To use the GCMs for estimating future 7Q10s, the exact climate variables used in the regression equations must be extracted for use. From DelSanto et al. (2023), the climate variables that were found to be statistically significant for 7Q10 estimation are:

1. The lowest 30-day cumulative precipitation, limited to abnormally hot periods ($X > 90$ th percentile temperatures).
2. Average 30-day temperature during the corresponding period of low cumulative precipitation.

Here, we test if extrapolation will be needed to provide the future 7Q10 estimates. This will occur if the future climate projections for the two climate variables exceed the range of values used in the training of the equations. For this verification, the long-term (2060-2100) RCP4.5 and RCP8.5 scenarios will be examined, as the temperature and precipitation changes differ the most in the long-term from the observed, historic values. Given the future temperature predictions in Figures 3a-d, it is unlikely that the future temperature values will remain within the training bounds of the equation. However, given the summer JJA precipitation predictions in Figures 4a-d, the precipitation values may remain within the original training bounds of the equation. In Figure 5, the lowest

30-day cumulative precipitation, limited to abnormally hot periods, is displayed for the historic, observed period and the long-term RCP4.5 and RCP8.5 scenarios.

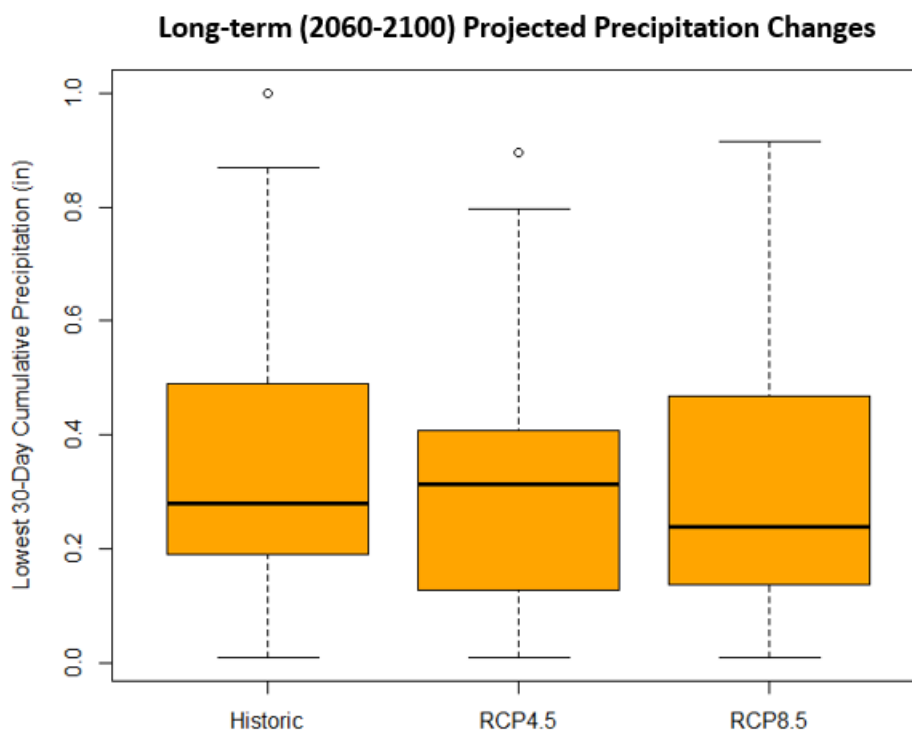


Figure 5: Historic lowest 30-day precipitation, limited to abnormally high temperatures (above the 90th percentile), compared to average projections for 2060-2100 from nine GCMs for both RCP4.5 and RCP8.5 scenarios.

Figure 5 supports that the GCMs are predicting 30-day cumulative precipitation values in the future that remain within the training bounds of the regression equations, as they were trained on the historic values. Additionally, in Figure 6, the corresponding 30-day average temperatures for the long-term RCP4.5 and RCP8.5 scenarios are displayed.

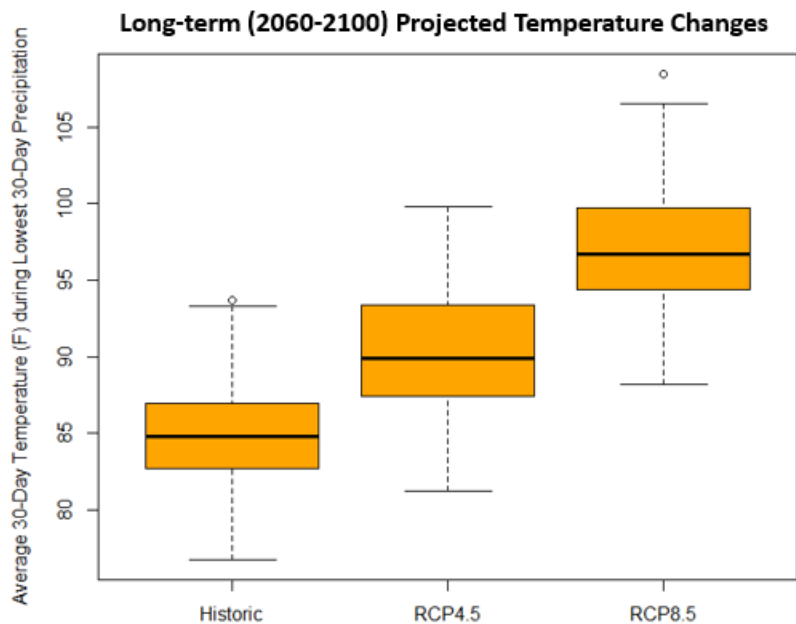


Figure 6: Historic 30-day average temperatures during the period of lowest precipitation compared to average projections for 2060-2100 from nine GCMs for both RCP4.5 and RCP8.5 scenarios.

As expected, Figure 6 displays temperature ranges in the long-term that exceed any values that the regression equations have been trained on. For the RCP4.5 scenario, the equation will only be extrapolating for the top 25th percentile of temperatures. For the RCP8.5 scenario, the equation will be extrapolating for the top 75th percentile of temperatures.

Appendix C.3 - Suggested Climate Changes and Initial Tests

In this section, we analyze the changes in 7Q10s projected by the regression equation, given the future projected climate values analyzed in the previous subsection. Using the results from the Figures 5 and 6, coupled with a similar analysis of the short-term projected changes, the following future climate projections are suggested to the user for use:

Projected Temperature Change	Projected Precipitation Change
Short-term (2020-2060)	
+ 3°F	+ 3%
+ 4°F	- 1%
Long-term (2060-2100)	
+ 5°F	- 8%
+ 10°F	- 15%

The largest changes are predicted in the long-term RCP 8.5 scenario, where the projected changes are +10°F for temperature and -15% for precipitation. In the following table, three arbitrary values (1, 10, and 100cfs) are used as tests to analyze the resulting 7Q10 projections given the maximum projected changes (+10°F for temperature and -15% for precipitation).

Table 4: Future Projected Changes in 7Q10s Given Largest Projected Climate Changes

Historic 7Q10 (cfs)	Projected 7Q10 (cfs)	% Change
1.00	0.45	-55.00%
10.00	4.52	-54.80%
100.00	45.22	-54.78%

The results in this table suggest that the regression equation predicts 7Q10 changes of up to -55% given the maximum projected climate changes from the GCMs. To further support if these projected changes are appropriate, future work could include:

1. Analyzing average historic changes in 7Q10s per year.

2. Analyzing the historic changes in 7Q10s in relation to recent observed temperature and precipitation changes.
3. Comparisons to a calibrated physical model where simulation tests can be completed to identify the relationship between climate changes and 7Q10 changes.

APPENDIX D

Question	Sample Response
Did the program open and work properly when you attempted to use it?	Yes
Does the interface seem intuitive enough for someone who did not attend the demonstration to use without further explanation?	No- multiple changes would be required, including clearer methodologies and documentation.
For the module pertaining to calculating the 7Q10 at a gage, please answer the following questions:	
1. Was it sufficiently easy to retrieve the streamflow data?	Yes
2. Did this module run quickly enough?	Yes, but including some type of “calculating” pop up box may be helpful to let the user know the program is not frozen.
3. Were the outputs/graphics easy to understand, and are they useful to your work?	Yes

4. Is the watershed area/area scaling methodology sufficient?	Yes, but a bit simplistic. Consider including a way to determine if this methodology is appropriate.
5. What other features in this part of the program would be valuable to you?	Ensure that you differentiate between water years and calendar years when calculating the 7Q10.
For the module pertaining to unengaged 7Q10 estimation, please answer the following questions:	N/A
1. Did the program work properly as it attempted to retrieve the physical characteristics of the watershed, required for the regression equations?	No, the program sometimes failed to retrieve the required physical characteristics.
2. Did this module run quickly enough?	Including some type of “calculating” pop up box may be helpful to let the user know the program is not frozen.
3. Were the outputs from this module easy to understand, and are they useful to your work?	Yes, but more things should be included, like the uncertainty in the 7Q10 estimates.
4. Are the graphics associated with this module useful and intuitive?	A map of the watershed could be included for better visualization.
5. Are there any additional features that we could include in this module which would be valuable to your work?	Better documentation, uncertainties, etc. could be included to increase trust in the results.
In its current state, does this program seem like it will be useful to your work? Please identify any changes or additions that would assist you or encourage you to use a future iteration of the program.	Maybe, but we could see this program being very useful if the changes above are addressed.

APPENDIX E

Addition	Reasoning
<u>Gaged 7Q10 Calculation</u>	
Plot the unrefined streamflow data, both in log space and standard space.	Plotting the unrefined streamflow allows the user to determine if there are any major gaps or problems with the data. Plotting the data in log space allows the user to analyze both low and high flows at once, especially for larger watersheds that have very large high flows.
Allow the user to move the x and y limits of the streamflow graphs so they can better examine specific time periods (x axis) and/or flow ranges (y axis).	Allowing the user to look at specific time periods is useful for examining specific years of flows (e.g., only the last 30 years). Allowing the user to look at specific flow ranges is useful for analyzing flows above or below a certain threshold (e.g., how often does the streamflow go below 10cfs?)
In addition to the 7Q10, also calculate the 7Q2 and 7Q30 at the gage if possible.	The 7Q2 and 7Q30 are also used frequently in planning/permitting (e.g., the 7Q30 is used in many nutrient loading permits)
Include a plot of only the smallest 10th percentile of flows, with the 7Q2, 7Q10, 7Q30, and minimum recorded streamflow for the gage displayed on the graph.	Including a plot of only the lowest 10 th percentile flows, as well as the 7Q2, 7Q10, and 7Q30 displayed on it, allows the user to better examine the low flows specifically.

<p>As discussed in Blum et al. (2019), provide a module where the Mann-Kendall test (Mann, 1945; Kendall 1975) is applied to the annual 7-day low flows to see if the low flows at a site are changing with statistical significance.</p>	<p>Stakeholders can better justify using a subset of streamflow data to calculate the 7Q10 if the Mann-Kendall test suggests a statistically significant change in the annual low flows (e.g., if the low flows are decreasing over time, using the last 30 years of data to calculate the 7Q10 is supported mathematically).</p>
<p>Rather than calculate the “recent” 7Q10 using the last 30 years of streamflow data by default, allow the user to see all available streamflow for the gage of interest and manually select a subset of data for calculating the low flow metrics.</p>	<p>The user can choose to use the last 30 years of streamflow data if they deem it appropriate, but allowing them to select exact start and end dates provides more flexibility if they only want to remove some of the data, select specific water years, etc.</p>
<p>Given the user-selected subset, plot the lowest 10th percentile flows once again, displaying the subset 7Q2, 7Q10, and 7Q30 values, as well as the full-record 7Q2, 7Q10, and 7Q30 as partially transparent.</p>	<p>This allows the user to visually confirm if the low flows have been increasing or decreasing. If they determine that the 7Q10 has decreased in recent years, they may be inclined to use that value as a conservative 7Q10. If it is similar to the full-record 7Q10 or larger, they may be more inclined to use the full-record 7Q10.</p>
<p>The ability to download all data and plots from the DSS.</p>	<p>They may want to include the graphics and results in the actual permit responses, so they would need that documentation.</p>
<p><u>Ungaged 7Q10 Estimation</u></p>	
<p>Extensive documentation and links to where the information is being retrieved from.</p>	<p>If the stakeholders are to use this for permitting, they need extensive documentation on all methods and links.</p>

<p>A “loading” pop-up that notifies the user while the program is retrieving information.</p>	<p>This program contacts the StreamStats API, which can take a long time. They wanted some way to determine if the program was frozen.</p>
<p>An interactive map that displays the watershed and physical characteristics next to it.</p>	<p>This helps the user determine if the watershed was correctly delineated, as well as displaying the corresponding physical characteristics.</p>
<p>Additional notes on the applicability of each statistical method for estimating 7Q10s.</p>	<p>The users wanted suggestions on when multiple linear regression, logarithmic regression, and the random forest algorithm provide the best 7Q10 estimates.</p>

BIBLIOGRAPHY

- Abatzoglou J. T. (2011). Development of gridded surface meteorological data for ecological applications and modelling " International Journal of Climatology, doi: 10.1002/joc.3413.
- Abatzoglou J.T. and Brown T.J. (2012). A comparison of statistical downscaling methods suited for wildfire applications. International Journal of Climatology, doi: 10.1002/joc.2312.
- Arnell, N. W., & Lloyd-Hughes, B. (2014). The global-scale impacts of climate change on water resources and flooding under new climate and socio-economic scenarios. *Climatic Change*, 122(1–2), 127–140.
- Asquith, W. H., & Thompson, D. B. (2008). Alternative regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using PRESS minimization: U.S. Geological Survey Scientific Investigations Report 2008–5084, 40 p.
- Austin, S. H., Krstolic, J. L., & Wiegand, U. (2011). Low-Flow Characteristics of Virginia Streams: U.S. Geological Survey Scientific Investigations Report 2011-5143. U.S. Geological Survey: Reston, VA, USA.
- Ayers, J., Villarini, G., Jones, C., Schilling, K., & Farmer, W. (2022). The Role of Climate in Monthly Baseflow Changes across the Continental United States. *Journal of Hydrologic Engineering*, 27(4), 04022006. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002170](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002170)
- Bayazit, M. (2015). Nonstationarity of Hydrological Records and Recent Trends in Trend Analysis: A State-of-the-art Review. *Environ. Process.* 2, 527–542. <https://doi.org/10.1007/s40710-015-0081-7>.
- Beck, H. E., van Dijk AI, J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52, 3599–3622. <https://doi.org/10.1002/2015WR018247>.
- Bent, G. C., Steeves, P. A., & Waite, A. M. (2014). Equations for Estimating Selected Streamflow Statistics in Rhode Island: U.S. Geological Survey Scientific Investigations Report 2014-5010. U.S. Geological Survey: Reston, VA, USA.
- Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, 43(9), 4382–4390. <https://doi.org/10.1002/2016GL068070>

- Bertoli, L., Balzarolo, D., Todini, E. (2022). On the Benefits of Collaboration between Decision Makers and Scientists: The Case of Lake Como. *Hydrology* 2022, 9, 187. <https://doi.org/10.3390/hydrology9110187>
- Blum, A. G., Archfield, S. A., Hirsch, R. M., Vogel, R. M., Kiang, J. E., & Dudley, R. W. (2019). Updating estimates of low-streamflow statistics to account for possible trends. *Hydrological Sciences Journal*, 64(12), 1404–1414. <https://doi.org/10.1080/02626667.2019.1655148>
- Bock, A. E., Santiago, M., Wieczorek, M. E., Foks, S. S., Norton, P. A., & Lombard, M. A. (2020). Geospatial Fabric for National Hydrologic Modeling, version 1.1 (ver. 3.0, November 2021): U.S. Geological Survey data release. <https://doi.org/10.5066/P971JAGF>
- Bonfante, A., Monaco, E., Manna, P., De Mascellis, R., Basile, A., Buonanno, M., Cantilena, G., Esposito, A., Tedeschi, A., De Michele, C., Belfiore, O., Catapano, I., Ludeno, G., Salinas, K., Brook, A. (2019). LCIS DSS—An irrigation supporting system for water use efficiency improvement in precision agriculture: A maize case study. *Agricultural Systems*, 176, 102646. ISSN 0308-521X. <https://doi.org/10.1016/j.agsy.2019.102646>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Cardwell, H., Langsdale, S., and Stephenson, K. (2008). The Shared Vision Planning Primer: How to incorporate computer aided dispute resolution in water resources planning. <https://www.iwr.usace.army.mil/Portals/70/docs/iwrreports/2008-R-02.pdf>
- Dimitriadis, P., and Koutsoyiannis, D. (2018). Stochastic synthesis approximating any process dependence and distribution. *Stoch Environ Res Risk Assess* 32, 1493–1515. <https://doi.org/10.1007/s00477-018-1540-2>.
- Dudley, R. W. (2004). Estimating Monthly, Annual, and Low 7-Day, 10-Year Streamflows for Ungaged Rivers in Maine: U.S. Geological Survey Scientific Investigations Report 2004-5026. U.S. Geological Survey: Reston, VA, USA.
- Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3), 32-57. DOI: 10.1080/01969727308546046
- England, J. F., Jr., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas, W. O., Jr., Veilleux, A. G., ... & Mason, R. R., Jr. (2019). Guidelines for determining flood flow frequency — Bulletin 17C (ver. 1.1, May 2019): U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p. <https://doi.org/10.3133/tm4B5>

EPA Office of Water. (2018). Low Flow Statistics Tools: A How-To Handbook for NPDES Permit Writers. https://www.epa.gov/sites/default/files/2018-11/documents/low_flow_stats_tools_handbook.pdf

Farmer, W. H., LaFontaine, J. H., & Hay, L. E. (2019). Calibration of the US Geological Survey National Hydrologic Model in Ungauged Basins Using Statistical At-Site Streamflow Simulations. *Journal of Hydrologic Engineering*, SP 04019049, 24(11). Doi: 10.1061/(ASCE)HE.1943-5584.0001854

Ferreira, R. G., da Silva, D. D., Elesbon, A. A., Fernandes-Filho, E. I., Veloso, G. V., Fraga, M. d. S., Ferreira, L. B. (2021). Machine learning models for streamflow regionalization in a tropical watershed, *Journal of Environmental Management*, Volume 280, 2021, 111713, ISSN 0301-4797, <https://doi.org/10.1016/j.jenvman.2020.111713>.

Flynn, R. H., & Tasker, G. D. (2002). Development of Regression Equations to Estimate Flow Durations and Low-Flow-Frequency Statistics in New Hampshire Streams: U.S. Geological Survey Scientific Investigations Report 02-4298. U.S. Geological Survey: Reston, VA, USA.

Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., & Rigon, R. (2011). The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: Models of flow generation and propagation/routing. *Geoscientific Model Development*, 4, 943–955. <https://doi.org/10.5194/gmd-4-943-2011>

Ghodsvali, M., Dane, G., Vries, B. d. (2023). An integrated decision support system for the urban food-water-energy nexus: Methodology, modification, and model formulation, *Computers, Environment and Urban Systems*, Volume 100, 2023, 101940, ISSN 0198-9715, <https://doi.org/10.1016/j.compenvurbsys.2023.101940>.

Golian, S., Murphy, C., & Meresa, H. (2021). Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland. *Journal of Hydrology: Regional Studies*, 36, 100859. <https://doi.org/10.1016/j.ejrh.2021.100859>.

Guo, Y, Zhang, Y, Zhang, L, Wang, Z. (2021). Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *WIREs Water*. 2021; 8:e1487. <https://doi.org/10.1002/wat2.1487>

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *Journal of Hydrology*, 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.

Guyadeen, D., Thistlethwaite, J. & Henstra, D. Evaluating the quality of municipal climate change plans in Canada. *Climatic Change* 152, 121–143 (2019). <https://doi.org/10.1007/s10584-018-2312-1>

- Hastie, T., & Tibshirani, R. J. (1986). *Generalized Additive Models*. Chapman and Hall.
- Hay, L., and Umemoto, M. (2007). Multiple-Objective Stepwise Calibration Using Luca. <https://doi.org/10.3133/ofr20061323>
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013). Global flood risk under climate change. *Nature Clim. Change*, 3(9), 816–821.
- Hodgkins, G. A., & Dudley, R. W. (2011). Historical summer base flow and stormflow trends for New England rivers. *Water Resources Research*, 47, W07528. <https://doi.org/10.1029/2010WR009109>
- Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6, 65–70. <https://doi.org/10.2307/4615733>
- Iman, R. L., & Conover, W. J. (1983). *A Modern Approach to Statistics*. John Wiley.
- Jensen, M.E., & Haise, H.R. (1963). Estimating Evapotranspiration from Solar Radiation. *Journal of the Irrigation and Drainage Division*, 89, 15-41.
- Karmalkar, A.V., Thibeault, J.M., Bryan, A.M. et al. (2019). Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change* 154, 367–386. <https://doi.org/10.1007/s10584-019-02411-y>
- Kendall, M.G. 1975. *Rank Correlation Methods*, 4th edition, Charles Griffin, London.
- Kilgore, R., Thomas, W. O. Jr., Paul, S., Douglass, S., Webb, B., Hayhoe, K., Stoner, A., Jacobs, J. M., Thompson, D. B., Herrmann, G. R., Douglas, E., & Anderson, C. (2019). Applying Climate Change Information to Hydrologic and Coastal Design of Transportation Infrastructure Design Practices. Prepared for the National Cooperative Highway Research Program, Transportation Research Board. https://onlinepubs.trb.org/Onlinepubs/nchrp/docs/NCHRP1561DesignPracticesGuide_rev.pdf
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55, 11344–11354. <https://doi.org/10.1029/2019WR026065>
- Li, X., Khandelwal, A., Jia, X., Cutler, K., Ghosh, R., Renganathan, A., et al. (2022). Regionalization in a global hydrologic deep learning model: From physical descriptors to random vectors. *Water Resources Research*, 58, e2021WR031794. <https://doi.org/10.1029/2021WR031794>

Lins, H. F. (2012). USGS Hydro-Climatic Data Network 2009 (HCDN-2009); Fact Sheet 2012-3047. U.S. Geological Survey: Reston, VA, USA.

Livneh, B., Bohn, T. J., Pierce, D. W., Muñoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., Brekke, L. (2015). A Spatially Comprehensive, Meteorological Data Set for Mexico, the U.S., and Southern Canada (NCEI Accession 0129374). NOAA National Centers for Environmental Information.

Livneh, B., & National Center for Atmospheric Research Staff. (2019). The Climate Data Guide: Livneh Gridded Precipitation and Other Meteorological Variables for Continental US, Mexico and Southern Canada. Retrieved from <https://climatedataguide.ucar.edu/climate-data/livneh-gridded-precipitation-and-other-meteorological-variables-continental-us-mexico>.

Loucks, D. P., and da Costa., J. R. (1991). Decision Support Systems, Water Resources Planning, NATO ASI Series G: Ecological Sciences, 26 3-54053-097-5.

Loucks, D. P., Stedinger, J. R., and Haith, D. A. (1981). Water Resource Systems Planning and Analysis, 0-13945-923-5. Hall, Inc.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281–297

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, Journal of the Econometric Society, 13:163-171.

Markstrom, S.L., Regan, R.S., Hay, L.E., Viger, R.J., Webb, R.M.T., Payn, R.A., & LaFontaine, J.H. (2015). PRMS-IV, the precipitation-runoff modeling system, version 4: U.S. Geological Survey Techniques and Methods, book 6, chap. B7, 158 p., <https://dx.doi.org/10.3133/tm6B7>

Maurer, E.P., A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen. (2002). A Long-Term Hydrologically-Based Data Set of Land Surface Fluxes and States for the Conterminous United States, *J. Climate* 15, 3237-3251

McCulloch, W. S., & Pitts, W. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5, 115–133. <https://doi.org/10.1007/BF02478259>

Mekanik, F., Imteaz, M. A., & Talei, A. (2016). Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. *Climate Dynamics*, 46, 3097–3111. <https://doi.org/10.1007/s00382-015-2755-2>

- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., Stouffer, R. J. (2008). Stationarity Is Dead: Whither Water Management. *Science*, 319(5863), 573–574.
- Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics*, 21, 3301–3307. <https://doi.org/10.1093/bioinformatics/bti499>.
- Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>.
- Mount, J., Hanak, E., Chappelle, C., Gray, B., Lund, J., Moyle, P., & Thompson, B. (2015). Policy Priorities for Managing Drought. Public Policy Institute of California. https://www.ppic.org/wp-content/uploads/content/pubs/report/R_315EHR.pdf
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57, e2020WR028091. <https://doi.org/10.1029/2020WR028091>
- Pierce, D. W., D. R. Cayan, and B. L. Thrasher (2014). Statistical downscaling using Localized Constructed Analogs (LOCA). *Journal of Hydrometeorology*, volume 15, page 2558-2585.
- Rahmani, M., Jahromi, S. H. M., Darvishi, H. H. (2023). SD-DSS model of sustainable groundwater resources management using the water-food-energy security Nexus in Alborz Province, *Ain Shams Engineering Journal*, Volume 14, Issue 1, 2023, 101812, ISSN 2090-4479, <https://doi.org/10.1016/j.asej.2022.101812>.
- Regan, R.S., Juracek, K.E., Hay, L.E., Markstrom, S.L., Viger, R.J., Driscoll, J.M., LaFontaine, J.H., Norton, P.A. (2019). The U. S. Geological Survey National Hydrologic Model infrastructure: Rationale, description, and application of a watershed-scale model for the conterminous United States. *Environmental Modeling & Software*, 111, 192-203. <https://doi.org/10.1016/j.envsoft.2018.09.023>
- Ries, K. G., III. (2000). Methods for Estimating Low-Flow Statistics for Massachusetts Streams: U.S. Geological Survey Water Resources Investigations Report 00-4135. U.S. Geological Survey: Reston, VA, USA.

Ries, K. G., III, Guthrie, J. D., Rea, A. H., Steeves, P. A., & Stewart, D. W. (2008). StreamStats: A Water Resources Web Application. U.S. Geological Survey Fact Sheet 2008-3067. U.S. Geological Survey.

Rousseeuw, P. J. (1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics*, 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Rumsey, C. A., Miller, M. P., Susong, D. D., Tillman, F. D., Anning, D. W. (2015). Regional scale estimates of baseflow and factors influencing baseflow in the Upper Colorado River Basin. *Journal of Hydrology: Regional Studies*, 4, 91–107. <https://doi.org/10.1016/j.ejrh.2015.04.008>

Salas, J. D., Obeysekera, J., & Vogel, R. M. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: A review. *Hydrological Sciences Journal*, 63(3), 325–352. <https://doi.org/10.1080/02626667.2018.1426858>

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). Hydrologic unit maps. (1987)

Selvanathan, S., Sreetharan, M., Rand, K., Smirnov, D., Choi, J., & Mampara, M. (2016). Developing Peak Discharges for Future Flood Risk Studies Using IPCC's CMIP5 Climate Model Results and USGS WREG Program. *Journal of the American Water Resources Association*, 52(4), 979–992. <https://doi.org/10.1111/1752-1688.12407>.

Serrat-Capdevila, A., B., J., & V., H. (2011). Decision Support Systems in Water Resources Planning and Management: Stakeholder Participation and the Sustainable Path to Science-Based Decision Making. InTech. doi: 10.5772/16897

Shepard, D. S. (1984). Computer mapping: The SYMAP interpolation algorithm. In G. L. Gaile, C. J. Willmott, & D. Reidel (Eds.), *Spatial Statistics and Models* (pp. 133–145). Springer.

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, 20(7), 2611–2628. <https://doi.org/10.5194/hess-20-2611-2016>

Siddique, R., Karmalkar, A., Sun, F., & Palmer, R. (2020). Hydrological extremes across the Commonwealth of Massachusetts in a changing climate. *Journal of Hydrology: Regional Studies*, 32, 100733. <https://doi.org/10.1016/j.ejrh.2020.100733>.

Siddique, R. and Palmer, R. (2021). Climate Change Impacts on Local Flood Risks in the U.S. Northeast: A Case Study on the Connecticut and Merrimack River Basins. *Journal of the American Water Resources Association* 75–95. <https://doi.org/10.1111/1752-1688.12886>.

Smakhtin, V. U. (2001). Low flow hydrology: A review. *Journal of Hydrology*, 240(3–4), 147–186. [https://doi.org/10.1016/S0022-1694\(00\)00340-1](https://doi.org/10.1016/S0022-1694(00)00340-1)

Soleymani, S. A., Goudarzi, S., Anisi, M. H., Hassan, W. H., Idris, M. Y. I., Shamshirband, S., & Ahmedy, I. (2016). A novel method to water level prediction using RBF and FFA. *Water Resources Management*, 30(9), 3265–3283. <https://doi.org/10.1007/s11269-016-1347-1>

Stuckey, M. H. (2006). Low-Flow, Base-Flow, and Mean-Flow Regression Equations for Pennsylvania Streams: U.S. Geological Survey Scientific Investigations Report 2006-5130. U.S. Geological Survey: Reston, VA, USA.

Tasker, G. D., & Stedinger, J. R. (1989). An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111(1–4), 361–375. [https://doi.org/10.1016/0022-1694\(89\)90268-0](https://doi.org/10.1016/0022-1694(89)90268-0)

Taylor, K.E., R.J. Stouffer, G.A. Meehl (2012). An Overview of CMIP5 and the experiment design. MS-D-11-00094.1

Tidwell V. C., Passell H. D., Conrad S. H., and Thomas R. P. (2004). System dynamics modeling for community-based water planning: application to the Middle Rio Grande. *Aquatic Sciences* 66 1 16 DOIs 00027-004-0722-9.

Thornton, M.M., Thornton, P.E., Wei, Y., Mayer, B.W., Cook, R.B., Vose, R.S. (2016). Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, TN. <https://doi.org/10.3334/ORNLDAAC/1345>.

Tongal, H., & Martijn, J. B. (2018). Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. *Journal of Hydrology*, 564, 266–282.

US EPA (2019). BASINS 4.5 (Better Assessment Science Integrating point & Non-point Sources) Modeling Framework. National Exposure Research Laboratory, RTP, North Carolina. BASINS Core Manual. Accessed 7/24/2023.

U.S. Geological Survey, 2016, National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), accessed December 1, 2023, at URL <http://waterdata.usgs.gov/nwis/>

Vogel, R. M., & Kroll, C. N. (1990). Generalized low-flow frequency relationships for ungaged sites in Massachusetts. *Journal of the American Water Resources Association*, 26(2), 241–253. <https://doi.org/10.1111/j.1752-1688.1990.tb01367.x>

Wang, G., C. J. Kirchoff, A. Seth, J. T. Abatzoglou, B. Livneh, D. W. Pierce, L. Fomenko, and T. Ding (2020). Projected Changes of Precipitation Characteristics

Depend on Downscaling Method and Training Data: MACA versus LOCA Using the U.S. Northeast as an Example. *J. Hydrometeor.*, 21, 2739–2758, <https://doi.org/10.1175/JHM-D-19-0275.1>.

Wang, W., Zhao, Y., Tu, Y., Dong, R., Ma, Q., & Liu, C. (2023). Research on Parameter Regionalization of Distributed Hydrological Model Based on Machine Learning. *Water*, 15(3), 518. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/w15030518>.

Wardropper, C., & Brookfield, A. (2022). Decision-support systems for water management. *Journal of Hydrology*, 610, 127928. <https://doi.org/10.1016/j.jhydrol.2022.127928>.

Wiche, G.J. & Holmes, R.R. Jr. (2016). Streamflow Data. 10.1016/B978-0-12-801884-2.00013-X.

Wiley, J. B. (2008). Estimating Selected Streamflow Statistics Representative of 1930–2002 in West Virginia. U.S. Geological Survey Scientific Investigations Report 2008-5105, Version 2. U.S. Geological Survey.

Williams, A. P., Cook, B. I., & Smerdon, J. E. (2022). Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. *Nature Climate Change*, 12(3), 232–234. <https://doi.org/10.1038/s41558-022-01290-z>

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.

Yalcin Sumer, D. & Lansey, K. (2004) Evaluation of Conservation Measures in the Upper San Pedro Basin. In: *Critical Transitions in Water and Environmental Resources Management*, Pcdgs. World Water and Environmental Resources Congress 2004.

Zhang, S., Lu, L., Yu, J., & Zhou, H. (2016). Short-term water level prediction using different artificial intelligent models. In *Proceedings of the 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (pp. 1–6). IEEE. <https://doi.org/10.1109/Agro-Geoinformatics.2016.7577678>

Zhu, C., & Lettenmaier, D. P. (2007). Long-term climate and derived surface hydrology and energy flux data for Mexico: 1925–2004. *Journal of Climate*, 20, 1936–1946.