Addressing Mixed Populations in Flood Frequency Analyses: A Case Study in Eastern Pennsylvania

Avital Breverman, Hydraulic Engineer, U.S. Army Corps of Engineers Hydrologic Engineering Center, Davis, CA, avital.l.breverman@usace.army.mil;
Graduate Student, Civil and Environmental Engineering, Colorado State University, Fort Collins, CO
Michael Bartles, Senior Hydraulic Engineer, U.S. Army Corps of Engineers Hydrologic Engineering Center, Davis, CA, michael.d.bartles@usace.army.mil
Gregory Karlovits, Senior Hydraulic Engineer, U.S. Army Corps of Engineers Hydrologic Engineering Center, Davis, CA, gregory.s.karlovits@usace.army.mil
Mazdak Arabi, Professor and Borland Chair of Water Resources, Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, mazdak.arabi@colostate.edu

Abstract

In many parts of the United States, floods at a single site are caused by multiple mechanisms. Example flood mechanisms include tropical cyclones and rain-on-snow. These non-homogeneous flood series are typically referred to as mixed populations. While the two latest revisions of federal flood frequency guidelines, published in 1982 and 2019, identified the treatment of mixed populations as an area of future research, no quantitative guidance exists on the classification of flood events or the incorporation of flood types into frequency analyses. Furthermore, the Mid-Atlantic region of the United States is characterized by complex meteorology and numerous flood causal mechanisms but has rarely been studied in mixed population and flood typing literature. The study presents a flood frequency analysis with consideration of flood type using annualized partial duration series and incorporation of flood samples caused by rare meteorologic events using empirical probability distributions. The analysis is performed in the Lehigh River watershed in eastern Pennsylvania. Floods along the Lehigh River are caused by tropical cyclones, rain-on-snow, and rainfall events. An automated flood classification procedure was developed using gridded meteorologic products. The automated classification procedure was validated manually using historic storm publications. Flood frequency curves were generated through two methods: using Bulletin 17C procedures adapted for mixed flood series and by combining annualized partial duration series from distinct flood causal mechanisms. While the flood frequency analysis results varied across the watershed, separation of flood series by causal mechanism generally resulted in higher flood quantiles.

Introduction

Flood frequency analysis (FFA) relates the magnitude of flood events to their annual probability of exceedance. FFA is used in the design of hydraulic structures, flood insurance, and floodplain zoning. In the United States, flood frequency guidelines have been published since 1967 (England et al., 2019). The most recent versions of these documents, *Guidelines for Determining Flood Flow Frequency Bulletin 17B* and *Guidelines for Determining Flood Flow Frequency Bulletin 17C,* were published by the Advisory Committee on Water Information Hydrology Subcommittee in 1982 and 2019, respectively. The Hydrology Subcommittee recognized the “identification and treatment of mixed distributions, including those based on hydrometeorological or hydrological conditions” in the Future Studies section of both documents (Hydrology Subcommittee Interagency Advisory Committee on Water Data, 1982;

England et al., 2019). Yet, no quantitative guidance exists on the classification of flood events or the incorporation of flood types into frequency analyses.

In Bulletin 17C, the Subcommittee on Hydrology acknowledged that flooding arises from multiple flood causal mechanisms and therefore, annual peak flows for a single gage site do not necessarily arise from a single distribution representing just one population. Consideration of mixed populations in flood series has important implications for design flood estimates, regionalization of statistical flood indices, and for improving our understanding of the underlying processes that cause flooding (Hirschboeck, 1987; Fischer & Schumann, 2021). Most importantly, separation of flood events by causal mechanism improves the likelihood that the random sample of flood events used in FFA are independent and identically distributed (IID).

While the possibility of mixed populations in a flood series has been widely recognized for decades (Beard, 1962; Hirschboeck, 1987; U.S. Army Corps of Engineers, 1993), no single definition of flood causal mechanisms exists (Tarasova et al., 2019). Annual peak discharges recorded at U.S. Geological Survey (USGS) gages do not typically include information about the underlying flood causal mechanisms, except for a single streamflow qualification code used to describe flows caused by snowmelt, hurricane, and debris dam breakup (U.S. Geological Survey, 2019).

Many recent studies have attempted to establish frameworks for classifying flood events by their generation mechanisms. Frameworks for classifying flood events include season-based, hydrograph-based, hydrological, and hydroclimatic (Tarasova et al., 2019). Hirschboeck (1987) presented one of the first hydroclimatic analyses of mixed populations. The study used synoptic weather patterns to categorize floods in the Gila River basin in Arizona into eight categories. More recently, flood studies using hydrometeorology to classify floods were performed in the Allegheny River basin in Pennsylvania and in the Big Thompson River basin in Colorado (Grote, 2017; Yu et al., 2021). Multiple studies have developed storm and flood type classifications using station and gridded meteorologic data to describe meteorologic and land surface processes (Martin, et al., 2018; Grote, 2017; Barth et al., 2019; O'Grady et al., 2022).

In Bulletin 17C, the Hydrology Subcommittee explicitly discouraged separation by season or calendar period (England et al., 2019) because the date of the maximum annual flow is variable in many parts of the United States, particularly in the Northeast (Berghuijs et al., 2016). Despite the meteorologic complexity of the Mid-Atlantic region, the area has not been included in local or regional mixed population and flood typing studies. This paper presents a case study of a FFA, with consideration of flood type, in the Lehigh River watershed in eastern Pennsylvania. The study approach consists of the following procedures: (1) streamflow data collection and preparation, (2) selection of flood events, (3) flood type classification, (4) daily-to-instantaneous flow conversion, and 5) fitting statistical models to flow data.

Data and Methods

Study Watershed: Lehigh River

The Lehigh River is one of the three main tributaries to the Delaware River and includes approximately 1,300 square miles of land area in nine counties. Elevations in the Lehigh River basin range from approximately 2,000 feet in its northern headwaters to 200 feet near the confluence with the Delaware River. The Delaware River Basin includes approximately 12,800 square miles of land area in four states (New York, New Jersey, Pennsylvania, and Delaware). The Delaware River basin is home to 8.3 million people (Delaware River Basin Commission, 2023). Furthermore, approximately 13.3 million people (4% of the United States' population) rely on the river for drinking, agricultural, and industrial use (Delaware River Basin

Commission, 2023). Flows along the Lehigh are regulated by two U.S. Army Corps of Engineers (USACE) dams, F. E. Walter Dam and Beltzville Dam. F. E. Walter Dam was constructed in 1961 and was originally authorized for flood risk management, but recreation became a Congressionally authorized purpose in 1988 (Historical Research Associates, Inc., 2012). Beltzville Dam was constructed in 1972 and authorized primarily for flood risk management, water supply, and low flow augmentation and secondarily for water quality and recreation (Historical Research Associates, Inc., 2012). The Lehigh River supports recreational activities including whitewater rafting, boating, and fishing. Two non-Federal dams located upstream of Beltzville Dam are operated by the Bethlehem Water Authority. The dams have capacities of 12,000 and 18,400 ac-ft and each regulate a drainage area of approximately 20 square miles (Bethlehem Authority, 2023). Storage effects from these reservoirs are typically negligible during flood events.

Flood events that affect the Lehigh River Basin include summer thunderstorms, heavy rains associated with the passage of tropical storm remnants, and combinations of snow and rainfall (U.S. Army Corps of Engineers, 1985). Winter snow cover in the Mid-Atlantic typically lasts days to weeks. Mean annual maximum snow water equivalent (SWE) values ranged from 2.5 to 5 inches during the 2003-2017 period (Cho & Jacobs, 2020). Events can be either localized in scale and primarily affect the Lehigh River Basin or regional in scale and affect the larger Delaware River watershed.

The USGS gages used in this analysis and their locations in the Lehigh River Basin are shown in Table 1 and Figure 1, respectively. Inflow hydrographs for F. E. Walter and Beltzville Dams were computed using measured outflow and water surface elevations, and static elevation-storage relationships. A 5-hour center moving average was used to smooth to the computed inflow hydrograph.

Table 1. Select USGS gages along Lehigh River and Pohopoco Creek

Gage ID	Short Name	Drainage Area (sq. mi.)	Availability of Mean Daily Flow (WY)	Availability of Annual Peak Flows (WY)
USGS Gage 01447500 – Lehigh River at Stoddartsville, PA	Stoddartsville	92	1944-2022	1942-2022
USGS Gage 01447800 – Lehigh River below F. E. Water Reservoir near White Haven, PA	White Haven	290	1958-2022	1955-2022
USGS Gage 01448000 – Lehigh River at Tannery, PA	Tannery	322	1917-1959	1915-1959
USGS Gage 01449000 – Lehigh River at Lehighton, PA	Lehighton	591	1983-2022	1982-2022
USGS Gage 01451000 – Lehigh River at Walnutport, PA	Walnutport	889	1947-2022	1942-2022
USGS Gage 01453000 - Lehigh River at Bethlehem, PA	Bethlehem	1,279	1903-1904, 1910-2022	1902-2022
USGS 01449800 – Pohopoco Creek below Beltzville Dam near Parryville, PA	BeltzvilleDS	96	1968-2022	1969-2022

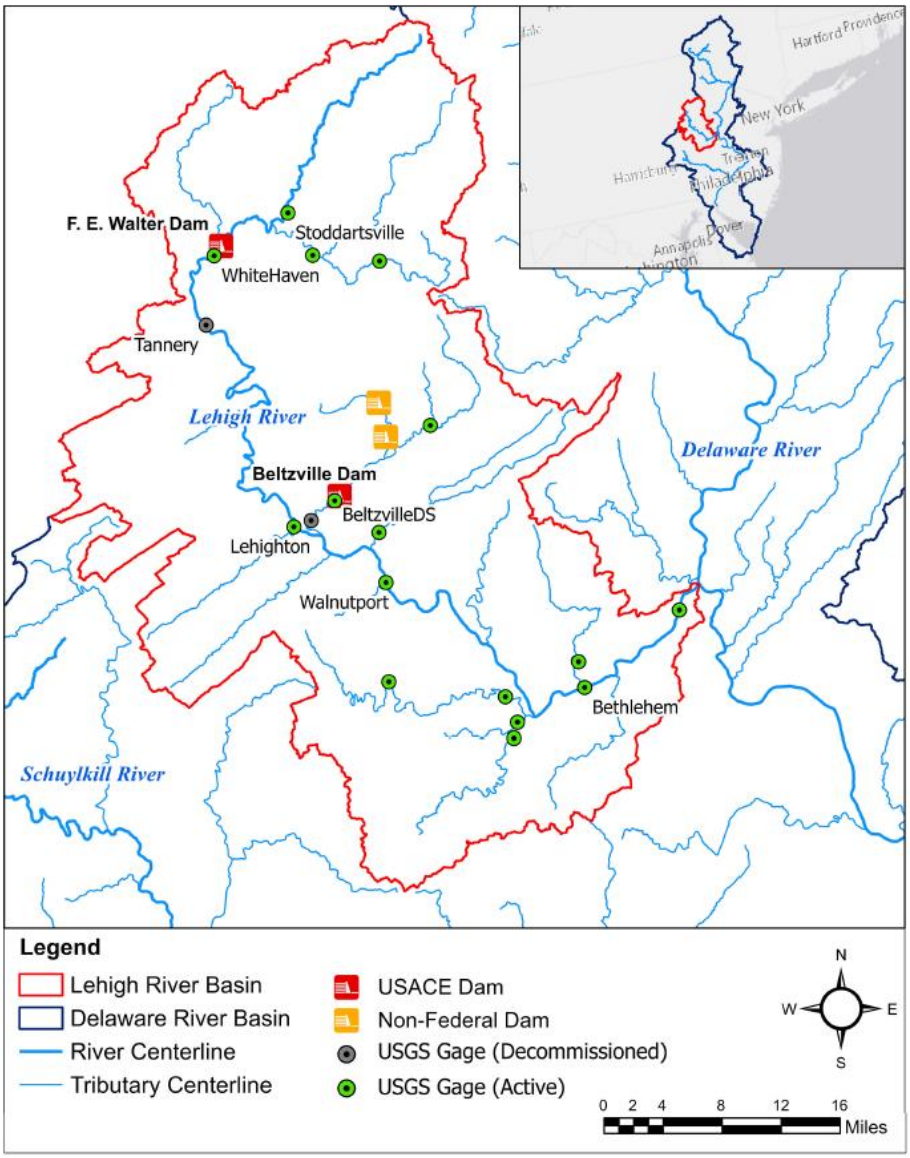

Figure 1. Lehigh River basin

Data Collection and Preparation

FFA procedures, such as those documented in Bulletin 17C, involve fitting analytical probability distributions to streamflow records that constitute "a representative time sample of random, homogeneous events" (England et al., 2019). Gage record lengths are often significantly shorter than the return period of design events. In addition, Bulletin 17C procedures are designed for streamflow records that aren't appreciably altered by upstream regulation or other anthropogenic impacts. In Bulletin 17C, the development of "national guidance on methods for estimating flood flow frequency curves at stream locations affected by varying degrees of regulation" was identified as an area of future work. Pre-dam streamflow records are not typically long enough to use in FFA. Therefore, regulated streamflow records often need to be transformed to unregulated streamflow for FFA.

This study used hydrologic routing to transform regulated systematic streamflow to unregulated streamflow. Flow routing is a mathematical routine used to describe the change in magnitude,

speed, and shape of a flood wave as it travels through a waterway (Maidment, 1993). Routing describes the attenuation and delay in hydrographs. Flow routing is classified as lumped/hydrologic or distributed/hydraulic (Maidment, 1993). A hydrologic routing model, developed in the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) version 4.10, was used to deregulate the post-dam systematic record. Previously developed storage-outflow relationships (U.S. Army Corps of Engineers Philadelphia District, 2017) were used in the Modified Puls routing method.

The unregulated streamflow records were computed using procedures detailed by the U.S. Army Corps of Engineers (1998) and Pearson (2021). Since the Tannery gage was discontinued in 1959, the unregulated streamflow records at the White Haven and Tannery gages were combined using a drainage area scaling adjustment. This combined record is referred to as White Haven/Tannery.

Selection of Flood Events

Two approaches to modeling extreme events are the block maxima and peak over threshold methods (Coles, 2001). Bulletin 17C assumes the use of block maxima, where the block is defined as a water year (WY) spanning October 1 – September 30. This method results in one peak flow for each WY, referred to as the annual maximum series (AMS).

The peaks over threshold approach extracts values above a certain threshold (Coles, 2001; Naghettini, 2017). The selected values behave as extreme events and are referred to as the partial duration series (PDS). Two diagnostic graphs are typically used to select the flow threshold: (1) mean excess and (2) Generalized Pareto (GP) distribution shape parameter, computed using L-moments (Hosking & Wallis, 1997). The flood threshold was selected through evaluation of shape parameter stability and nonlinearity in the mean excess plot's slope (Naghettini, 2017).

To ensure independence between selected floods events, the number of days required between events was defined as (Hydrology Subcommittee Interagency Advisory Committee on Water Data, 1982):

$$d = \lceil 5 + ln(A) \rceil \qquad (1)$$

where $\lceil x \rceil$ denotes the ceiling function and A is the drainage area in square miles. The minimum separation between independent events of 6 days.

The annual maximum series (AMS) and partial duration series (PDS) were developed from the mean daily unregulated streamflow record.

Flood Type Classification

Flood events were categorized using both automated and manual approaches. Manual classification of flood events was accomplished using historic storm publications (National Oceanic and Atmospheric Administration, 2022). The manual flood classifications were used to validate the automated classification procedure. The automated classification scheme used SWE and tropical cyclone and tropical storm remnant (TC/TSR) datasets to classify flood events as caused by (1) rain-on-snow, (2) TC/TSR, or (3) rainfall. These flood types were developed based on a priori knowledge of the region. The automated classification procedure used an antecedent time period of d days, computed using equation (1), to relate the relevant antecedent time period to the site drainage area.

Tropical cyclones are synoptic-scale (hundreds to thousands of kilometers) low-pressure storm systems (Hirschboeck, 1991). Tropical cyclones that impact the United States originate in the

western North Atlantic Ocean, the Gulf of Mexico and the Caribbean Sea, and the eastern North Pacific Ocean (Hirschboeck, 1991). Tropical cyclones are responsible for some of the highest rainfall accumulations in the United States, including Hurricanes Harvey and Florence. The International Best Track Archive for Climate Stewardship (IBTrACS) provides the most comprehensive catalog of TC/TSR tracks (Knapp et al., 2010; Knapp et al., 2018). IBTrACS version 4 data for the North Atlantic are available from 1851 to near present. Smith et al. (2011) defined tropical cyclone flood peaks if a tropical cyclone passed within 500 km (310 mi) of the stream gage within a 2-week window, centered on the day of the peak. In this study, flood events were attributed to TC/TSR if they met the following criteria: (1) the TC/TSR track occurred within d days of the flood event and (2) the TC/TSR track passed within 250 km (155 mi) of the gage location.

Snow-related gridded products typically have shorter temporal coverage than other meteorologic gridded products. As a result, multiple snow datasets were used in this analysis to classify flood events as rain-on-snow driven. The University of Arizona (UA) (Broxton et al., 2019) and 20th Century Reanalysis Project version 3 (20CRV3) (Compo et al., 2011) datasets were used for estimates of snow water equivalent (SWE) data. The UA dataset provides daily estimates of SWE and snow depths over the conterminous United States for October 1981 – October 2021, with a horizontal spatial scale of 4 km x 4 km. The dataset was developed by assimilating SWE and snow depth data from the Snow Telemetry (SNOTEL) network stations, snow depth from the Cooperative Observer Program (COOP) stations, and Parameter-elevation Regression on Independent Slopes Model (PRISM) daily 4 km x 4 km precipitation and temperature datasets (Broxton et al., 2019). The 20CR dataset is an ensemble of four-dimensional weather and atmospheric condition maps (Slivinski et al., 2019; Compo et al., 2011). Version 3 of the dataset spans 1836 – 2015 and has a spatial resolution of 1 degree x 1 degree (approximately 60 miles x 45 miles in the Lehigh River basin).

Yu et al. (2021) used a 5 mm snowmelt threshold to classify flood events as snowmelt-driven (potentially including rain-on-snow events) in the Colorado Front Range The threshold was refined by Yu et al. (2022) to separately classify snowmelt-driven and rain-on-snow events, using a minimum volume input of 10 mm, from either precipitation and/or snowmelt.

In the Lehigh River, flood events were classified as rain-on-snow driven if the maximum decrease in SWE in the d days prior to the event exceeded 10 mm (0.4 in). For the period where both UA and 20CRV3 estimates were available, a maximum decrease in SWE of 10 mm from either source was used. Snowmelt-driven events were included within the rain-on-snow classification because snowmelt events are relatively infrequent in the Mid-Atlantic region (Welty & Zeng, 2021). The remaining flood events were classified as rainfall-driven. The decision tree approach used to classify flood events is shown in Figure 2.

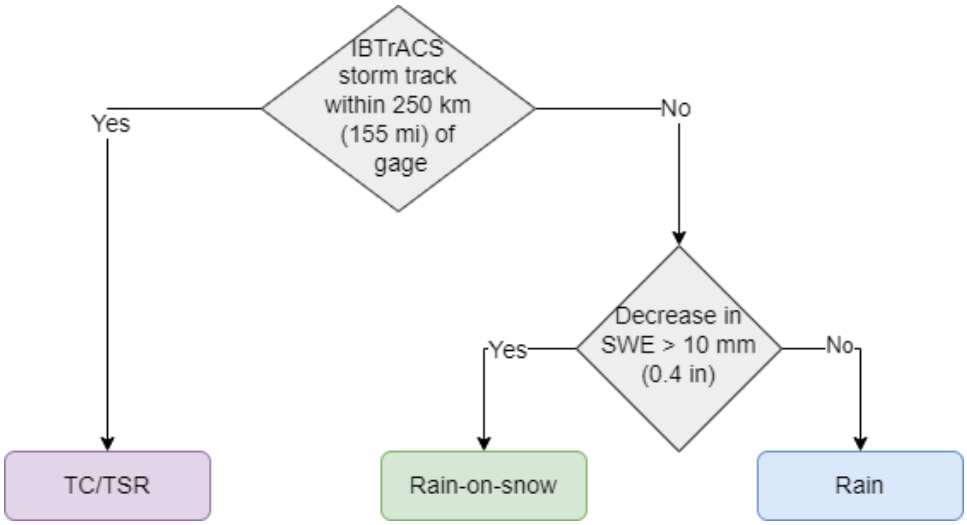

Figure 2. Flood typing decision tree

The automated and manual flood type classifications were compared by computing the number of PDS flood events in each category during calendar years 1958-2018, the period when National Oceanic and Atmospheric Administration historic storm publications were available. The automated and manual flood type classifications showed agreement (Figure 3). All tropical cyclone-induced floods were correctly identified using the automated classification. The discrepancies in the number of rainfall and rain-on-snow events were attributed to the coarseness of the 20CRV3 dataset. In addition, the monthly historic storm publications were missing information on multiple flood events at each site (classified as "No Data" in Figure 3).

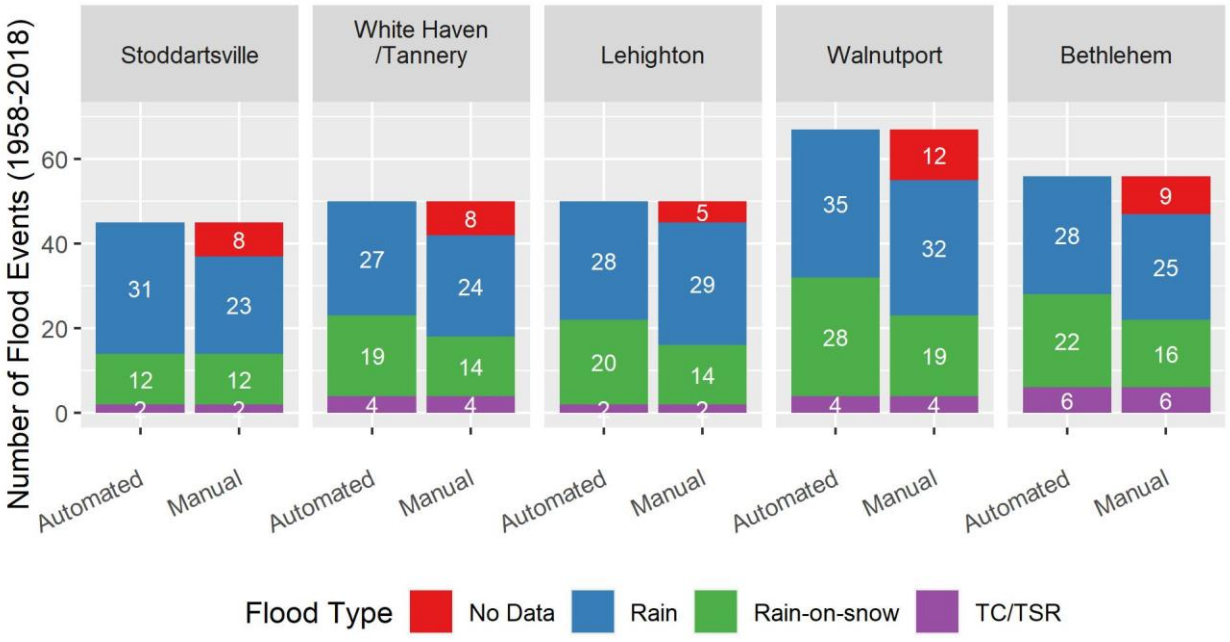

Figure 3. 1958-2018 PDS flood event types using automated and manual classification methods

Daily-to-instantaneous Streamflow Conversion

Quantile mapping was used to develop a daily-to-instantaneous flow relationship. Quantile mapping is a bias correction method applied to modeled data such that its distribution matches that of the observed data (Piani et al., 2010). A transfer function, TF, is derived such that the modeled variables' distribution function, F, is equivalent to that of the observed variable (Piani et al., 2010):

$$F_{model}(x_{model}) = F_{observed}(x_{observed}) \quad (2)$$

$$x_{corrected} = TF(x_{model}) \quad (3)$$

Quantile mapping is advantageous as it is independent of statistical distribution. The R library "qmap" (Gudmundsson, 2016) was used to develop parametric relationships for the mean daily and instantaneous discharge pairs.

Pairs of mean daily and instantaneous (or hourly) flows, corresponding to the same event, were developed from pre-regulation or deregulated periods. All flow pairs were used for quantile mapping model development. Quantile mapping model performance was evaluated using the root mean square error (RMSE)-observed standard deviation ratio (RSR). The RSR is computed as (Moriasi et al., 2007):

$$\text{RSR} = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_i (y_{obs,i} - y_{sim,i})^2}}{\sum_i (y_{obs,i} - \bar{y})^2} \quad (4)$$

The results of the quantile mapping model development are shown in Table 2.

Table 2. Daily-to-instantaneous flow quantile mapping model results

Gage Short Name	Total Number of Flood Pairs	Model RSR
Stoddartsville	63	0.09
White Haven/Tannery	100	0.07
Lehighton	23	0.18
Walnutport	37	0.13
Bethlehem	66	0.09

The quantile mapping power function performance was evaluated by randomly selecting one-third of the flow pairs at each gage in 100 iterations. For each iteration, the RSR of the instantaneous flow from the quantile mapping was computed. The RSR values from the quantile mapping validation are shown in Figure 4. The model RSR values and mean and median validation RSR values were less than 0.20 for all gages, indicating excellent performance. The RSR values were lower (i.e. model performance was better) for gages with higher numbers of flood pairs and longer records.

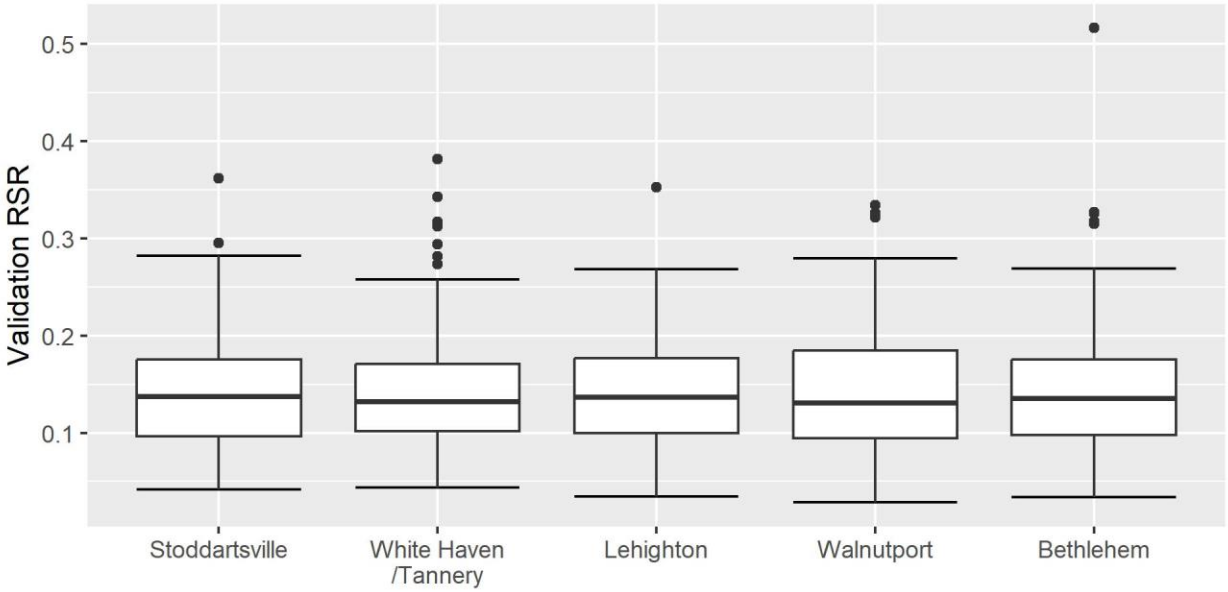

Figure 4. Results of daily-to-instantaneous flow quantile mapping validation

Statistical Models for Flood Frequency Analysis

Two statistical models were used to perform flood frequency analyses: (1) Log-Pearson Type III distribution fit using the Expected Moments Algorithm (EMA), as described in Bulletin 17C and (2) conversion of the PDS/GP model to the corresponding annual maximum distribution, modeled using the Generalized Extreme Value (GEV) distribution and fit using L-moments.

The Hydrologic Engineering Center's Statistical Software Program (HEC-SSP) version 2.3 was used to perform flow frequency analyses in accordance with Bulletin 17C. The Bulletin 17C EMA framework enables FFA to be computed using flood series combined from systematic, historical, and paleoflood records, including censored values. Systematic records that are useful for estimated flood frequency include peak flows, daily flows, reservoir inflows and pool elevations, and streamflow measurements (England et al., 2019). Historical data is the result of anecdotal evidence, such as high-water marks on bridge abutments, rather than from instrumentation. Paleoflood data were not available in the Lehigh River basin. Finally, periods of missing data are described using perception thresholds. Perception thresholds describe the range of measurable or perceivable discharges, as shown in Figure 5. The missing values are within a range, but the exact value is unknown. Historical data for the Lehigh River gages was available from a previous basin study (U.S. Army Corps of Engineers, 1985). Specifically, the August 1955 flood event was the largest flow since at least 1786 for the Bethlehem gage and 1902 for the remaining gages. Perception thresholds were used to describe periods of missing data during the systematic record. A perception threshold of 32,000 cfs was selected for the Bethlehem gage, shown in Figure 3, based on the smallest historical event in the record (i.e. a historical event with a magnitude 32,000 cfs was recorded indicating that discharges up to that magnitude were perceivable). The perception threshold at each gage was assumed to be constant over the historical period.

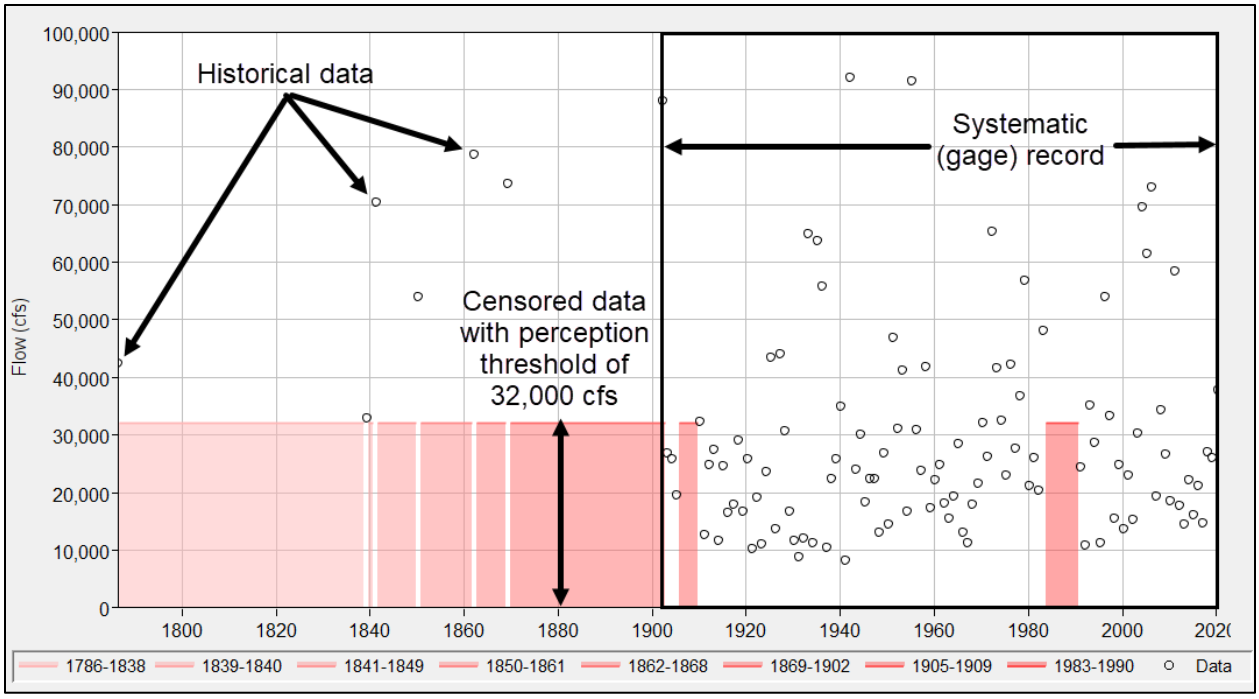

Figure 5. Treatment of Bethlehem gage systematic, historical, and censored data in a Bulletin 17C Analysis in HEC-SSP

Extreme value distributions are the limiting distributions for the minimum and maximum values of a large sample of random events. The GEV distribution family and GP distribution are used to model IID random extreme events (block maxima and threshold excesses, respectively). The GP distribution was used to model the PDS flood events at each gage. However, the GP distribution describes conditional exceedance, not annual exceedance probability (AEP). The relationship between the parameters of the GP and GEV distributions can be used to compute equivalent parameters of the GEV distribution of block maxima from the parameters of the GP distribution of threshold excesses (Coles, 2001).

The occurrence of flood peaks exceeding the selected threshold value is assumed to be described by a Poisson process (Madsen et al., 1997a):

$$\lambda = N/t \qquad (5)$$

where N is the number of observed exceedances in the period of t years and λ is the Poisson rate parameter.

The relationships between the GP distribution location q_0, scale α^*, and shape κ^* parameters and GEV distribution parameters for location ξ, scale α, and shape κ are (Madsen et al., 1997b):

$$\xi = q_0 + \frac{\alpha}{\kappa}(1 - \lambda^{-\kappa}) \qquad (6)$$

$$\alpha^* = \alpha\, \lambda^{-\kappa} \qquad (7)$$

$$\kappa = \kappa^* \qquad (8)$$

The R package "lmom" (Hosking, 2022) was used to estimate distribution parameters for each type of flood-generating mechanism using L-moments. Flood quantiles corresponding to common AEPs were computed from the resulting GEV distribution.

Bulletin 17C recommends a minimum systematic record length of 10 years to warrant statistical analysis (England et al., 2019). Except for the Bethlehem gage, the Lehigh River gages had less than 10 floods attributed to TC/TSR during their records. In addition, many of the tropical cyclone floods were produced by the same storm events. Consequently, regionalization was not pursued because the few sample flood events were correlated. As a result, an empirical distribution was used for the TC/TSR flood events.

The empirical annual exceedance probability was computed using the unbiased Weibull plotting position (Maidment, 1993):

$$PP = \frac{i}{N+1} \quad (9)$$

where i is the rank of the flood (i = 1 corresponds to the largest flood event) and N is the total number of years of record.

Floods caused by one mechanism are assumed to be independent of floods caused by the other mechanisms. The GEV distributions of the rainfall and rain-on-snow flood events were combined with an empirical frequency distribution developed from TC/TSR floods in the AMS. The combined probability of flooding exceeding a threshold in a given year was computed using the probability of union:

$$P_c = 1 - (1-P_1)(1-P_2)(1-P_3) \quad (10)$$

where P_1, P_2, and P_3 are the probabilities of flooding caused by TC/TSR, rain-on-snow, and rainfall events, respectively.

Unlike Bulletin 17C procedures, the GP-GEV model does not incorporate historical and paleoflood information.

Results and Discussion

At most gages and AEPs, the combined annualized PDS frequency curve produced larger flood quantiles than the mixed Bulletin 17C curve (Figure 6). The difference between the two FFA methods is most pronounced at the tails of the distributions. Since an empirical probability distribution was used for the TC/TSR events, the frequency curves were not extended beyond the empirical plotting position of the flood of record. For the right tail of the distribution (0.02 and 0.01 AEP events or 50- to 100-year return periods), the difference in flood quantiles ranged from 9% to 85% (Figure 7). The Flood quantiles at these AEPs are controlled by TC/TSR events and are based on empirical distributions, meaning that they are dependent on record length. The AEPs of the TC/TSR events are likely rarer than the current record lengths and plotting positions imply. For example, the August 1955 event may remain the (unregulated) flood of record along the Lehigh River for the next 200 years, extending the gage record lengths by 200 years and reducing the AEPs of that event.

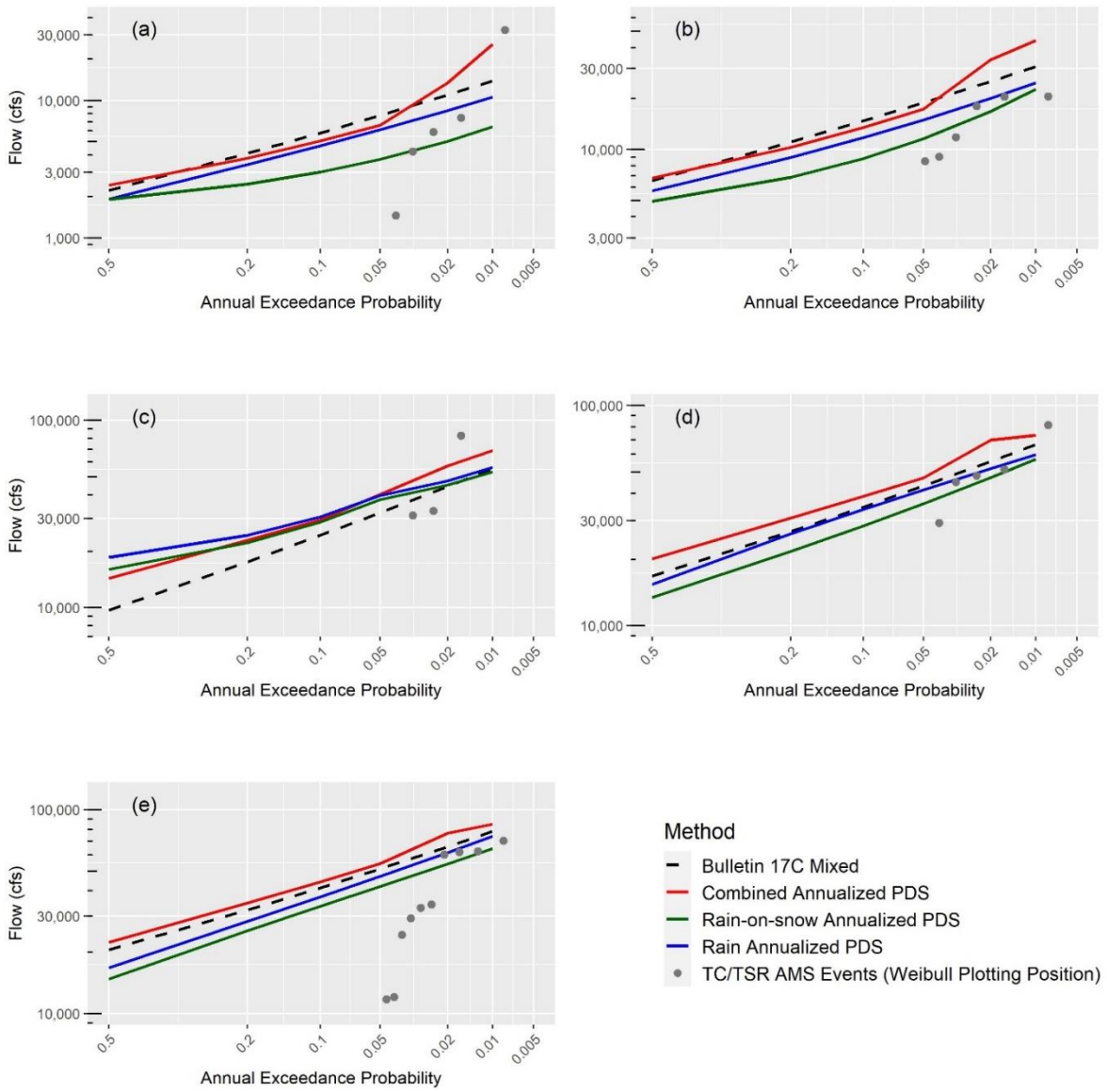

Figure 6. Flood frequency curves for (a) Stoddartsville, (b) White Haven/Tannery, (c) Lehighton, (d) Walnutport, and (e) Bethlehem gages

Flood quantiles for more frequent events (0.50 to 0.05 AEP events or 2- to 20-year return period) are controlled by the rain-on-snow and rainfall frequency curves. In the Lehigh River basin, non-TC/TSR rainfall events typically produce larger floods than rain-on-snow events. The differences in flood quantiles for more frequent events are less pronounced between the two FFA methods. For the AEPs in the middle of the range, the flood quantiles from the combined frequency curve were lower than the mixed curve for the Stoddartsville and White Haven/Tannery gages. In addition, the flood quantiles produced by the combined curve were substantially higher than those produced by the mixed curve for the same gages at AEPs of 0.02 and 0.01. Therefore, drainage area and the scale of flood responses produced by different flood mechanisms are important in the identification and treatment of flood mechanisms in mixed population analysis.

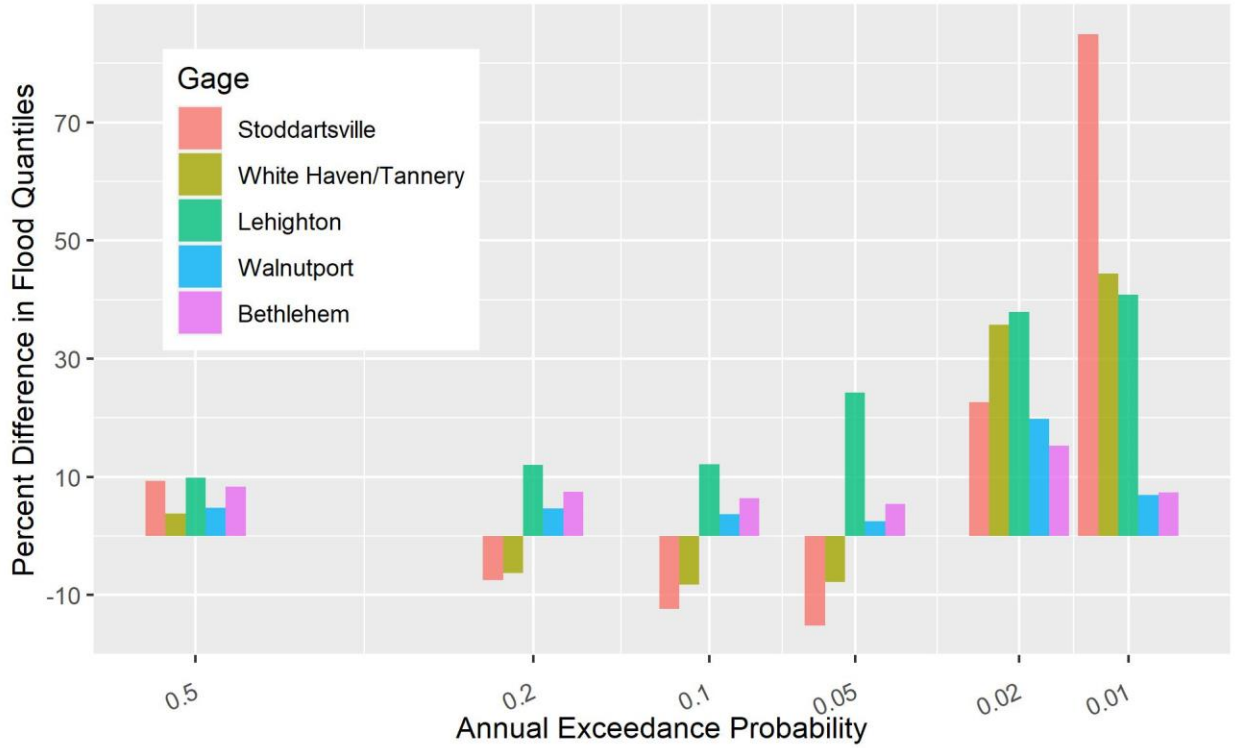

Figure 7. Percent difference in flood quantile: annualized combined PDS relative to mixed AMS implemented using B17C procedures

Conclusion

Out results demonstrate that consideration of mixed populations within a flood series is important, particularly for estimation of rare events considered in flood management. Treatment of a mixed flood series as homogeneous in FFA violates the assumption that flood events are identically distributed. While the instantaneous annual maximum discharges at a site have historically been used in flood frequency, the use of a PDS of mean daily discharge has the potential to increase the information content of a flood series and reduce the need for the treatment of censored flows.

The automated flood type classifications used in this study show promise. While a manual classification approach is more efficient for a site-specific or watershed-scale FFA, automated procedures allow for efficient classification of larger regions. While this study focused on the Mid-Atlantic region of the United States, the indices used in the classification algorithm could be expanded and the algorithm could be extended to other regions. The right tail behavior of frequency curves is often controlled by rare meteorologic events, such as tropical cyclones, with sample sizes too small for methods based on analytical probability distributions to be feasible. Flood frequency analysis should be combined with precipitation frequency analysis and hydrologic modeling to produce robust estimates of rare meteorologic and hydrologic events (Yu et al., 2021; Yu et al., 2022). As usual with FFA, there is uncertainty in predicting low AEP floods due to short record lengths.

References

Barth, N., Villarini, G., & White, K. (2019). Accounting for Mixed Populations in Flood Frequency Analysis: Bulletin 17C Perspective. *Journal of Hydrologic Engineering*.

Beard, L. R. (1962). *Statistical Methods in Hydrology*. Davis, CA: U.S. Army Corps of Engineers.

Berghuijs, W., Woods, R., Hutton, C., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters, 43*(9), 4382-4390.

Bethlehem Authority. (2023). *Penn Forest and Wild Creek Reservoirs*. Retrieved from Bethlehem Authority: www.bethlehemauthority.org/reservoirs

Broxton, P., Zeng, X., & Dawson, N. (2019). *Daily 4 km Gridded SWE and Snow Depth from Assimilated In-Situ and Modeled Data over the Conterminous US, Version 1 User's Guide*. Boulder, CO: NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:https://doi.org/10.5067/0GGPB220EX6A

Cho, E., & Jacobs, J. M. (2020). Extremevalue snow water equivalent and snowmelt for infrastructure design overthe contiguous United States. *Water Resources Research, 56*.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Compo, G.P., et al. (2011). The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc., 137*(654), 1-28.

Delaware River Basin Commission. (2023, March 7). *Basin Information*. Retrieved from Delaware River Basin Commission: www.nj.gov/drbc/basin/

England, J., Cohn, T., Faber, B., Stedinger, J., Thomas, W., Veilleux, A., . . . Mason, R. (2019). *Guidelines for Determining Flood Flow Frequency Bulletin 17C*. Reston, VA: U.S. Geological Survey.

Fischer, S., & Schumann, A. (2021). Regionalisation of flood frequencies based on flood type-specific mixture distributions. *Journal of Hydrology X, 13*.

Grote, T. (2017). Hydroclimatological Identification of Flood Seasonality and Generating Mechanisms in the Upper Allegheny River Basin. *Middle States Geographer, 50*, 74-81.

Gudmundsson, L. (2016). Statistical Transformations for Post-Processing Climate Model Output. R package. Version 1.0-4. Retrieved from https://cran.r-project.org/web/packages/qmap/qmap.pdf

Hirschboeck, K. (1987). Hydroclimatically-defined mixed distributions in partial duration flood series. In V. Singh, *Hydrologic frequency modeling* (pp. 199-212). Boston, MA: D. Reidel Publishing Company.

Hirschboeck, K. (1991). *Hydrology of floods and droughts, climate and floods: U.S. Geological Survey Water Supply Paper 2375*.

Historical Research Associates, Inc. (2012). *Responsiveness and Reliability: A History of the Philadelphia District and the Marine Design Center, U.S. Army Corps of Engineers, 1972-2008*. Philadelphia: U.S. Army Corps of Engineers.

Hosking, J. R. (2022). L-Moments. R package. version 2.9. Retrieved from https://cran.r-project.org/web/packages/lmom/lmom.pdf

Hosking, J., & Wallis, J. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge, UK: Cambridge University Press.

Hydrologic Engineering Center. (1982). *Mixed Population Frequency Analysis*. Davis, CA: U.S. Army Corps of Engineers.

Hydrology Subcommittee Interagency Advisory Committee on Water Data. (1982). *Guidelines for Determining Flood Flow Frequency Bulletin 17B*. Reston, VA: U.S. Geological Survey.

Knapp, K., Diamond, H., Kossin, J., Kruk, M., & Schreck, C. (2018). *International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4*. Retrieved from NOAA National Centers for Environmental Information.

Knapp, K., Kruk, M., Levinson, D., Diamond, H., & Neumann, C. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bulletin of the American Meteorological Society*, 363-376.

Langbein, W. B. (1949). Annual Floods and the Partial-duration Flood Series. *American Geophysical Union, 30*(6), 879-881.

Madsen, H., Pearson, C., & Rosbjerg, D. (1997b). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. *Water Resources Research, 33*(4), 759-769.

Madsen, H., Rasmussen, P., & Rosbjerg, D. (1997a). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling. *Water Resources Research, 33*(4), 747-757.

Maidment, D. R. (1993). *Handbook of Hydrology*. New York: McGraw-Hill, Inc.

Martin, D., Caldwell, R., Parzybok, T., Bahls, V., Crow, B., & Gibson, W. (2018). *Trinity River Hydrologic Hazards Project Task 2 Report—Storm Typing for the Trinity River Basin*. MetStat, Inc.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *American Society of Agricultural and Biological Engineers*, 885-900.

Naghettini, M. (2017). *Fundamentals of Statistical Hydrology*. Cham, Switzerland: Springer International Publishing.

National Oceanic and Atmospheric Administration. (2022). *Storm Data Publication*. Retrieved from National Centers for Environmental Information Image and Publications System: https://www.ncdc.noaa.gov/IPS/sd/sd.html

National Weather Service. (2022). Retrieved from Weather Prediction Center's Surface Analysis Archive: https://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive.php

O'Grady, J. G., Stephenson, A. G., & McInnes, K. L. (2022). Gauging mixed climate extreme value distributions in tropical cyclone regions. *Scientific Reports, 12:4626*.

Pearson, C. E. (2021). *Flood Hazard Methodology for Semi-Quantitative Risk Assessment*. Lakewood, CO: U.S. Army Corps of Engineers.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., & Haerter, J. O. (2010). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology, 395*, 199-215.

Slivinski, L. et al. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Q. J. R. Meteorol Soc., 145*, 2876-2908.

Smith, J. A., Villarini, G., & Baeck, M. L. (2011). Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States. *Journal of Hydrometeorology, 12*(2), 294-309.

Tarasova, L. et al. (2019). Causative classification of river flood events. *WIREs Water, 6*(4).

U.S. Army Corps of Engineers. (1985). *Modification of the Francis E. Walter Dam and Reservoir General Design Memorandum*. Philadelphia, PA: Department of the Army.

U.S. Army Corps of Engineers. (1993). *Engineer Manual 1110-2-1415 Hydrologic Frequency Analysis*. Washington, D.C.: Department of the Army.

U.S. Army Corps of Engineers. (1998). *American River, California Rain Flood Flow Frequency Analysis*. Sacramento, CA: U.S. Army Corps of Engineers.

U.S. Army Corps of Engineers Philadelphia District. (2017). *Delaware River Basin Corps Water Management System Report*. Philadelphia: U.S. Army Corps of Engineers.

U.S. Geological Survey. (2019, March 3). Retrieved from USGS Streamgages By the Numbers: https://www.usgs.gov/mission-areas/water-resources/science/usgs-streamgages-numbers

Welty, J., & Zeng, X. (2021). Characteristics and Causes of Extreme Snowmelt over the Conterminous United States. *Bulletin of the American Meteorological Society, 102*(8), E1526–E1542.
Yu, G., Wright, D. W., & Davenport, F. V. (2022). Diverse physical processes drive upper-tail flood quantiles in the US mountain west. *Geophysical Research Letters, 49*(10).
Yu, G., Wright, D., & Holman, K. (2021). Connecting Hydrometeorological Processes to Low-Probability Floods in the Mountainous Colorado Front Range. *Water Resources Research, 57*(4).