THESIS

MIXED POPULATIONS FLOOD FREQUENCY ANALYSIS IN THE MID-ATLANTIC REGION OF THE UNITED STATES

Submitted by

Avital Breverman

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2023

Master's Committee:

Advisor: Mazdak Arabi

Ryan Morrison
Steven Fassnacht

ABSTRACT

MIXED POPULATIONS FLOOD FREQUENCY ANALYSIS IN THE MID-ATLANTIC REGION OF THE UNITED STATES

In many parts of the United States, floods at a single site are caused by multiple mechanisms. Flood mechanisms can broadly be classified as meteorologic, land surface processes, and disturbances. 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. Without quantitative guidance on the treatment of mixed populations in flood frequency analyses, there is the potential for considerable variability in frequency-based flood estimates. The treatment of a flood series resulting from a mixed population violates the assumption that floods at a site are independent and identically distributed. To avoid this issue, separate statistical models should be fit to floods arising from different mechanisms and the resulting curves should be combined to produce flood quantiles. Mixed population and flood typing literature has focused primarily on the western United States. In comparison, the Mid-Atlantic region of the United States is characterized by complex meteorology and numerous flood causal mechanisms but has been studied less frequently in mixed population literature.

In addition to the absence of guidance of mixed populations, flood frequency approaches in the United States use the annual maximum series, or the maximum flow each year. The peaks-over-threshold, or partial duration series, approach to selecting flood events has the potential to increase the information content in a flood series which is important when subdividing flood events by causal mechanism. The objectives of the study are to examine: (1) flood typing based on gridded meteorologic products, (2) the

advantages of using partial duration series over annual maximum series when performing a flood frequency analysis with consideration of flood type, and (3) the difference in flood quantiles from the proposed combined population methodology compared to those from the current mixed flood frequency analysis.

An automated flood classification procedure was developed using gridded meteorologic products. The automated classification procedure was validated manually using historic storm publications. Flood frequency analyses were performed using both partial duration and annual maximum flood series. The method is applied within the Lehigh River watershed in eastern Pennsylvania. While the flood frequency analysis results varied across the watershed, separation of flood series by causal mechanism generally resulted in higher flood quantiles than those obtained from mixed flood series. Design floods based on the treatment of flood series as homogeneous are likely underpredicting event magnitudes. In addition, quantitative guidance on separation of flood events by causal mechanism and treatment of flood type subsets within frequency analyses is needed to produce more reliable flood estimates.

ACKNOWLEDGEMENTS

I am grateful to my advisor Dr. Mazdak Arabi for his guidance and support in my research. Thank you to my thesis committee members, Drs. Ryan Morrison, and Steven Fassnacht for your time and for the opportunity to take interesting classes during my time at CSU.

I am thankful to my undergraduate advisor at Drexel University, Dr. Jonathan Cheng, for motivating me to pursue graduate research. Thank you to Dr. Jory Hecht, Dr. Nancy Barth, and Michelle Irizarry-Ortiz at the U.S. Geological Survey for your insightful discussions.

I am indebted to several colleagues and mentors at the U.S. Army Corps of Engineers. Thank you to Laura Bittner, Jake Helminiak, and Steve England at the Philadelphia District for providing me with challenging opportunities and for teaching me to "turn excuses into solutions." I am very grateful to have had the opportunity to be a part of the Hydrologic Engineering Center's Hydrology and Statistics Division while in graduate school. I would not have been able to complete this thesis without the mentorship of Mike Bartles and Greg Karlovits. Thank you for motivating me to examine flood frequency analysis from new perspectives. I am thankful to Greg for his assistance in learning the R programming language and broadening my understanding of hydrometeorology. I want to especially thank Mike who gave me the opportunity to sit in his office for a month in 2019 which sparked my interest in flood frequency analysis and mixed populations.

Finally, thank you to my family and friends for the encouragement and support I received over the past two years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Flooding is one of the costliest natural hazards (IPCC, 2012; Mishra et al., 2022). During the period 1972-2006, flooding caused an average of $2.67 billion in damages per year in the United States (Changnon, 2008). Flood risk and damages are expected to increase in response to urbanization and climate change (Slater & Villarini, 2016; Merz et al., 2020). The ability to estimate flood frequency and magnitude is crucial for flood risk management and flood damage reduction.

Flood frequency analysis relates the magnitude of flood events (typically instantaneous peak discharges) to their frequency, which is often expressed in terms of annual exceedance probability (AEP) and expected arrival time or return period. Flood frequency analysis is used in the design of hydraulic structures, flood insurance, and floodplain zoning and management. 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 Subcommittee on Hydrology recognized the need for "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, classification of flood events by causal mechanisms (e.g., snowmelt, rain-on-snow, tropical cyclones, etc.) and implementation of flood frequency analysis with flood type subsets has not been addressed in flood frequency analysis guidance.

Historically, flood frequency analysis has used the block maxima approach from extreme value theory to select flood events (Gumbel, 1958). The streamflow time series is partitioned into water year (WY) blocks, spanning October 1 through September 30. An analytical probability distribution is fit to the annual peak flows, or the annual maximum series (AMS), at a streamflow gage. In fitting statistical

models to data, a common assumption is that the data points are independent and identically distributed (IID). Annual maximum flows are used to ensure independence between events (unless consecutive annual maxima occur close to the WY demarcation). Beyond ensuring independence, the use of annual maxima in flood frequency analysis results from historic stream gaging methods. Prior to the implementation of digital recording gages in the 1960s, streamflow measurements were made using crest-stage gages (Carter et al., 1963). A crest-stage gage typically consists of a pipe, a graduated staff, and granulated cork (Linsley et al., 1982). When the stream level rises, the cork adheres to the staff at the highest level reached by the stream (Friday, 1965). While crest-stage gages offer an economical alternative to digital recording gages, they can only be used for a single streamflow event between manual inspections. Crest stage gage peak flow measurements were often collected only once per year.

While the AMS approach to flood frequency analysis has advantages, discharge in a river is a random process and does not follow an annual cycle. An alternative to the block maxima approach is the use of partial duration series (PDS), or the peaks-over-threshold (POT) method (Balkema & de Haan, 1974; Pickands, 1975). The PDS approach is particularly relevant for rivers in arid and semi-arid regions, where the largest flow in a dry year may be so small that it provides no relevant information for determination of flood magnitudes (Wang, 1991). The PDS approach postulates that only discharge events above a certain threshold provide information about floods (Todorovic & Zelenhasic, 1970; Todorovic & Rouselle, 1971; Wang, 1991). The PDS approach to flood frequency analysis has the potential to increase the predictive capability of flood frequency analysis techniques, partially due to larger sample size of the flood time series. This consideration is particularly important when dividing the flood time series by causal mechanisms.

In Bulletin 17C, the Subcommittee on Hydrology acknowledged that flooding arises from multiple 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). Mixed population analyses allow for examination of changing trends within a single flood subset. Most importantly, separation of flood events by causal mechanism improves the likelihood that the random sample of flood events used in flood frequency analysis are 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 mechanism 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, ice-jam, or debris dam breakup (U.S. Geological Survey, 2019).

Frameworks for classifying flood events include season-based, hydrograph-based, hydrological, and hydroclimatic (Tarasova et al., 2019). The National Weather Service (NWS) uses a hydrograph-based approach to classify floods. NWS subdivides floods into two categories, flash floods and river floods, based on the rate of water level rise. Flash floods are those that develop within 6 hours of precipitation or other causes, such as ice and debris jams (National Oceanic and Atmospheric Administration, 2023). Hydrological classifications typically use antecedent precipitation and snowmelt, measured from change in snow water equivalent (SWE), to differentiate between snow and rain dominated floods. Many studies have investigated rainfall-dominated and snowmelt-dominated floods in the Colorado Front Range. Jarrett and Costa (1982) developed generalized skew coefficients and frequency curves for rainfall-runoff and snowmelt-runoff in the Colorado Front Range. Yu et al. (2021, 2022) and Kampf and Lefsky (2016) used antecedent volumetric input from precipitation and snowmelt to discriminate between rainfall, rain-on-snow, and snowmelt events in the Colorado Front Range. Waylen and Woo (1982) used antecedent precipitation to separate rainfall/rain-on-snow and snowmelt dominated flood events in the Cascade Mountains. Others have developed hydrological classifications based on the temporal intensity of rainfall (Oppel & Fischer, 2020) and the ratio of runoff volume to runoff flood flow (Gaál et al., 2015; Fischer et

al., 2019). These approaches typically use unsupervised learning techniques to group flood events by temporal indices.

In recent years, process-based approaches to flood estimation have been motivated by the understanding that rainfall alone does not explain flood response (Turkington et al., 2016; Berghuijs et al., 2019; Merz et al., 2020; Yu et al., 2021). A variety of temporal and spatial indices have been used to classify floods. Merz and Blöschl (2003) proposed five classifications of floods (long-rain, short-rain, flash floods, rain-on-snow, and snowmelt) in Austria based on the following process indicators: day of year of flood, storm duration, rainfall depths, SWE, runoff coefficient, and spatial extent of floods. Hirschboeck (1987) presented one of the first hydroclimatic analyses of mixed populations, using a decision tree approach. 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 events 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). A daily database of storm types was developed for the Trinity River watershed in Texas using point-based and gridded meteorologic datasets that describe precipitation characteristics (Martin et al., 2018).

In Bulletin 17C, the Subcommittee on Hydrology explicitly discouraged separation by season or calendar period (England et al., 2019) since 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.

While many studies have examined classification schemes for flood causal mechanisms, few have investigated fitting analytical probability distributions to flood subsets for the purpose of flood frequency analysis. Murphy (2001) compared single- and mixed-population statistics for floods in Massachusetts using Bulletin 17B procedures, classifying floods into “ordinary,” tropical cyclone, and ice-jam-release events. Barth et al. (2019) used a simulation framework to perform flood frequency analyses in terms of

flood series comprised of atmospheric river and non-atmospheric river driven flood events in the western United States. Further complicating the lack of research on mixed population flood frequency analysis is the relatively infrequent nature of certain meteorologic events, such as tropical cyclones, that control the upper tail of flood frequency curves. Flood samples resulting from these events are often too small to reasonably fit analytical probability distributions.

This study presents an examination of approaches to flood frequency analysis with consideration of flood causal mechanisms. The flood frequency analysis approaches are applied to the Lehigh River watershed in eastern Pennsylvania. The objectives of the study are to examine: (1) flood typing based on gridded meteorologic products, (2) the advantages of using partial duration series over annual maximum series when performing a flood frequency analysis with consideration of flood type, and (3) the difference in flood quantiles from the proposed combined population methodology compared to those from the current mixed flood frequency analysis.

2. DATA AND METHODS

The study method is comprised of the following primary steps: (1) data collection and preparation, (2) flood type classification, (3) temporal resolution conversion, and (4) application of statistical models for flood frequency analysis. The Bulletin 17C framework utilizes multiple types of data, including systematic (gage) and historical data. In addition, Bulletin 17C procedures are not designed for use with streamflow records that are appreciably altered by upstream regulation. Streamflow records used in this study were collected from gaging sites, historic project documents, and post-storm reports. In addition, regulated streamflow records were transformed to unregulated records using hydrologic routing computations to remove the effects of reservoir regulation.

To assess the impact of mixed populations on flood flow quantiles, flood events were classified by seasons and by physical causal mechanisms. Seasonal, or day-of-year, classification approaches are the most common way to classify flood types. The relevant flood causal mechanisms were determined and classified manually using historic storm publications. Manual classification of flood types was used to develop and validate the automated classification, which was based on spatially distributed meteorologic products.

Mean daily streamflow data were used to avoid gaps in the period of record. Since flood flow quantiles are typically produced from instantaneous streamflow records, the mean daily streamflow data was converted to an instantaneous temporal resolution using quantile mapping.

Mixed population and combined frequency curves were computed. A mixed population curve is the frequency curve that is derived directly from flood flow data that have not been separated by causal mechanism (Figure 2-1 gray curve) while a combined population frequency curve is defined as the resultant frequency curve that is derived from two or more separate frequency curves (Figure 2-1 blue and green curves) and each curve is developed from a separate population (Figure 2-1 red curve) (Hydrologic Engineering Center, 1982). Two statistical modeling approaches were used to produced flood frequency

curves: (1) Bulletin 17C procedures were applied to the AMS and (2) the generalized Pareto (GP) distribution was applied to the PDS and analytically converted to the Generalized Extreme Value (GEV) distribution.

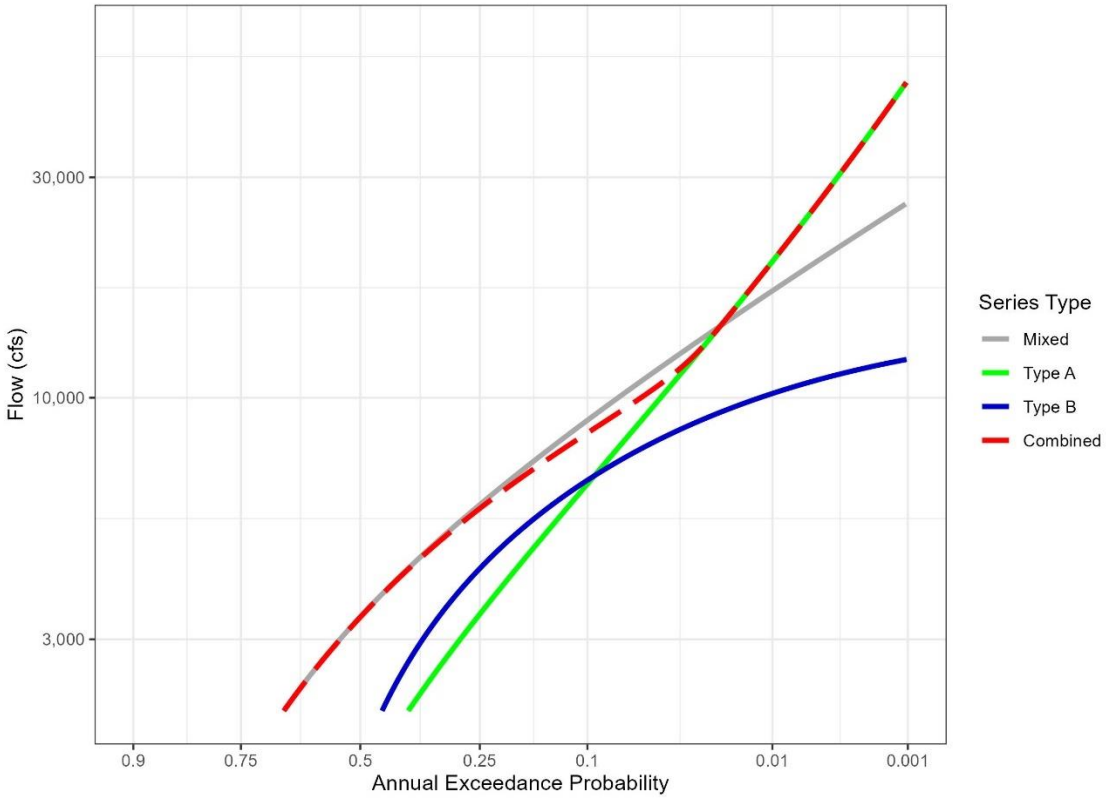


Figure 2-1. Mixed and combined flood frequency curves

Six (6) flow frequency analyses methods were developed (Table 2-1) to investigate the sensitivity of flood quantiles to consideration of mixed populations and flood event selection method.

Table 2-1. Flow frequency analyses modeling methods

Method	Flood Selection	Statistical Model	Mixed or Combined	Separation Approach	Base Temporal Resolution (Prior to Quantile Mapping Conversion)
AMH	**A**MS	B17C	**M**ixed	None	**H**ourly
AMD	**A**MS	B17C	**M**ixed	None	**D**aily
PMD	**P**DS	GP to GEV	**M**ixed	None	**D**aily
ACD_Seasonal	**A**MS	B17C	**C**ombined	Seasonal	**D**aily
ACD_Causal	**A**MS	B17C	**C**ombined	Automated classification by causal mechanism	**D**aily
PCD	**P**DS	GP to GEV	**C**ombined	Automated classification by causal mechanism	**D**aily

2.1. Study Watershed: Lehigh River

2.1.1. Watershed Description

The Lehigh River (Figure 2-2) is one of the three main tributaries to the Delaware River and includes approximately 1,300 mi^2 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 mi^2 of land area in four states: New York, New Jersey, Pennsylvania, and Delaware. The Delaware River basin is home to 8.3 million people. 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, Francis 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).

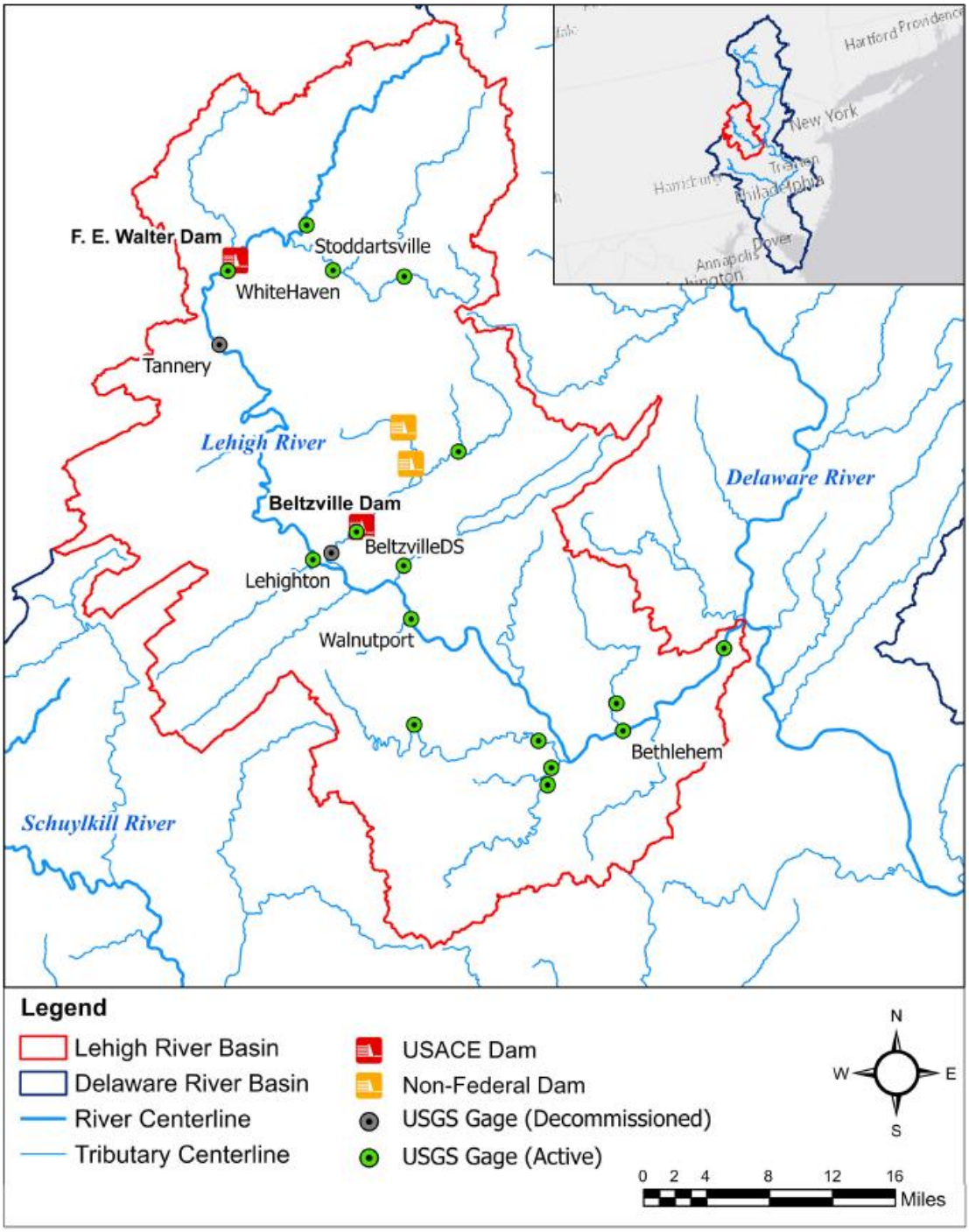


Figure 2-2. Lehigh River basin

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-feet and each regulate a drainage area of approximately 20 mi^2 (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). 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. Winter snow cover in the Mid-Atlantic typically lasts days to weeks. Mean annual maximum SWE values ranged from 2.5 to 5 inches during the 2003-2017 period (Cho & Jacobs, 2020).

2.1.2. Streamflow Data

Streamflow data from five USGS gages along the Lehigh River and one USGS gage along the Pohopoco Creek were collected. The drainage areas and periods of measured discharge at the gages are provided in Table 2-2. The availability of streamflow data is shown in Figure 2-3. Hourly streamflow data were only available for WY 1990-2022. This study used mean daily streamflow from the selected gages. In addition, inflow to F. E. Walter and Beltzville Dams was computed at an hourly time step using change of reservoir storage:

$$Q_{in,t} = Q_{out,t} + (S_t - S_{t-1}) \quad (1)$$

where Q_{in} and Q_{out} are the reservoir inflow and outflow at time t and $S_t - S_{t-1}$ is the change in storage in the reservoir between the current time t and the previous time step $t-1$.

The calculated inflow was smoothed using a 5-hour center-moving average, which has historically provided appropriate smoothing to computed values (U.S. Army Corps of Engineers Philadelphia District, 2017).

Table 2-2. Lehigh River USGS Gages

Gage ID and Name	Gage Short Name	Drainage Area (mi^2)	Availability of Mean Daily Flow (WY)	Availability of Annual Peak Flows (WY)
USGS 01447500 – Lehigh River at Stoddartsville, PA	Stoddartsville	92	1944-2022	1942-2022
USGS 01447800 – Lehigh River below F. E. Water Reservoir near White Haven, PA	White Haven	290	1958-2022	1955-2022
USGS 01448000 – Lehigh River at Tannery, PA	Tannery	322	1917-1959	1915-1959
USGS 01449000 – Lehigh River at Lehighton, PA	Lehighton	591	1983-2022	1982-2022
USGS 01451000 – Lehigh River at Walnutport, PA	Walnutport	889	1947-2022	1942-2022
USGS 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

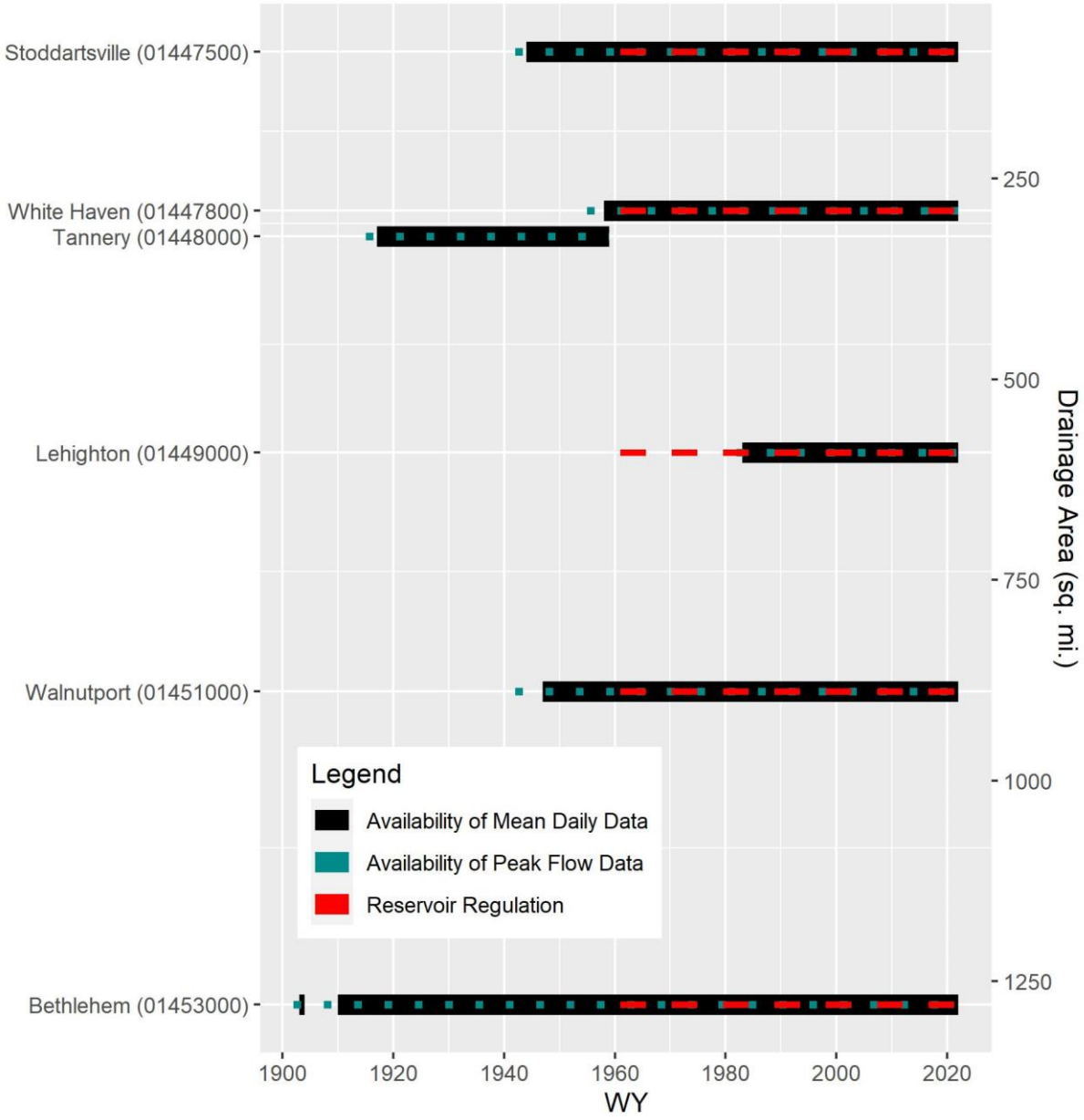


Figure 2-3. Lehigh River systematic streamflow data availability

The Lehigh River below F. E. Walter Reservoir near White Haven, PA gage (USGS 01447800) is located 3,300 feet downstream of F. E. Walter Dam and is treated as a conduit release gage. Streamflow data at the Lehigh River at Tannery, PA gage (USGS 01448000) is available through WY 1959, prior to the construction of F. E. Walter Dam. The Tannery and White Haven gages have two years (WY 1957-1959)

of overlapping streamflow data. The period of record from the Tannery gage was combined with the White Haven gage (USGS 01447800) using a drainage area adjustment:

$$\hat{Q}_1 = \mathrm{Q}_2(A_1/A_2)^{\Phi} \quad (2)$$

where $\hat{Q}_1$ is the estimated daily discharge at site 1, Q_2 is the measured discharge at site 2, A_1 is the drainage area at site 1, A_2 is the drainage area at site 2, and Φ is a regional coefficient.

The regional coefficient Φ was computed by minimizing the sum of squared errors between the computed and measured flow at the White Haven site over the observed record:

$$\text{argmin} \left\{ \sum \left[\hat{Q}_1 - Q_2 \left(\frac{A_1}{A_2} \right)^{\Phi} \right]^2, \Phi \right\} \quad (3)$$

where Q_W and Q_T represent daily measured discharge at the White Heaven and Tannery sites. A regional exponent of 0.90 was used to adjust the measured streamflow at the Tannery site for use at the White Haven site. Maintenance of variance (MOVE) record extension techniques were not used because Bulletin 17C guidance recommends a minimum of 10 years of overlapping record (England et al., 2019). In addition, only two years of streamflow records at the White Haven gage were unregulated.

2.1.3. Estimation of Unregulated River Discharge

The Bulletin 17C framework utilizes multiple types of data: systematic, historical, and paleoflood records, including censored values. Gage record lengths are significantly shorter than the return period of design events. Therefore, flood frequency analysis should leverage all available data. Flood frequency analysis procedures, such as those documented in Bulletin 17C, involve fitting analytical distributions to streamflow records that constitute "a representative time sample of random, homogeneous events." The first step in flood frequency analysis is "[a]ssessment of the adequacy and applicability of flood records." Bulletin 17C procedures are designed for streamflow records not appreciably altered by upstream regulation. Bulletin 17C identified the development of "national guidance on methods for estimating flood flow frequency curves at stream locations affected by varying degrees of regulation" as an area of

future work. Pre-regulation streamflow records are not typically long enough to use in a flow frequency analysis. Therefore, regulated streamflow records often need to be transformed to unregulated streamflow for flood frequency analysis.

To extract unregulated streamflow from a hydrologic model, a period of record simulation could be performed. However, this requires continuous meteorological forcing data of sufficient spatial and temporal resolution to simulate flood events in drainage areas of varying sizes. In addition, a period of record simulation is time-consuming to perform. This study used hydrologic routing to deregulate (i.e. remove the effects of reservoir regulation) streamflow data. 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. The procedure used to deregulate flow data (U.S. Army Corps of Engineers, 1998; Pearson, 2021) is described in Appendix A.

Streamflow for the four gages located downstream of the F.E. Walter and Beltzville Dams were deregulated using a simple hydrologic routing model in the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) version 4.10. The Modified Puls routing method was used. Previously developed storage-outflow relationships were used in the Modified Puls routing method (U.S. Army Corps of Engineers Philadelphia District, 2017). The deregulated streamflow records and data types are tabulated in Appendix B.

2.1.4. Snow Data

Snow-related gridded products typically have shorter temporal coverage than other gridded meteorologic products, such as precipitation and temperature. As a result, multiple snow datasets were used in this analysis for flood type classification. The University of Arizona (UA) (Broxton et al., 2019) and 20th Century Reanalysis Project version 3 (20CRV3) (Compo et al., 2011; Slivinski et al., 2019) datasets were used for SWE estimates. The UA dataset provides daily estimates of SWE and snow depths over the conterminous United States for October 1981 – near present, with a horizontal spatial scale of 4 km x 4

km (2.5 mi x 2.5 mi). 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 20CRV3 dataset is an ensemble of four-dimensional weather and atmospheric condition maps, produced by assimilating only surface pressure reports and observed sea-surface temperature (Compo et al., 2011; Slivinski et al., 2019). Version 3 of the dataset spans 1836 – 2015 and has a spatial resolution of 1° x 1° (approximately 60 mi x 45 mi in the Lehigh River basin, shown in Figure 2-4).

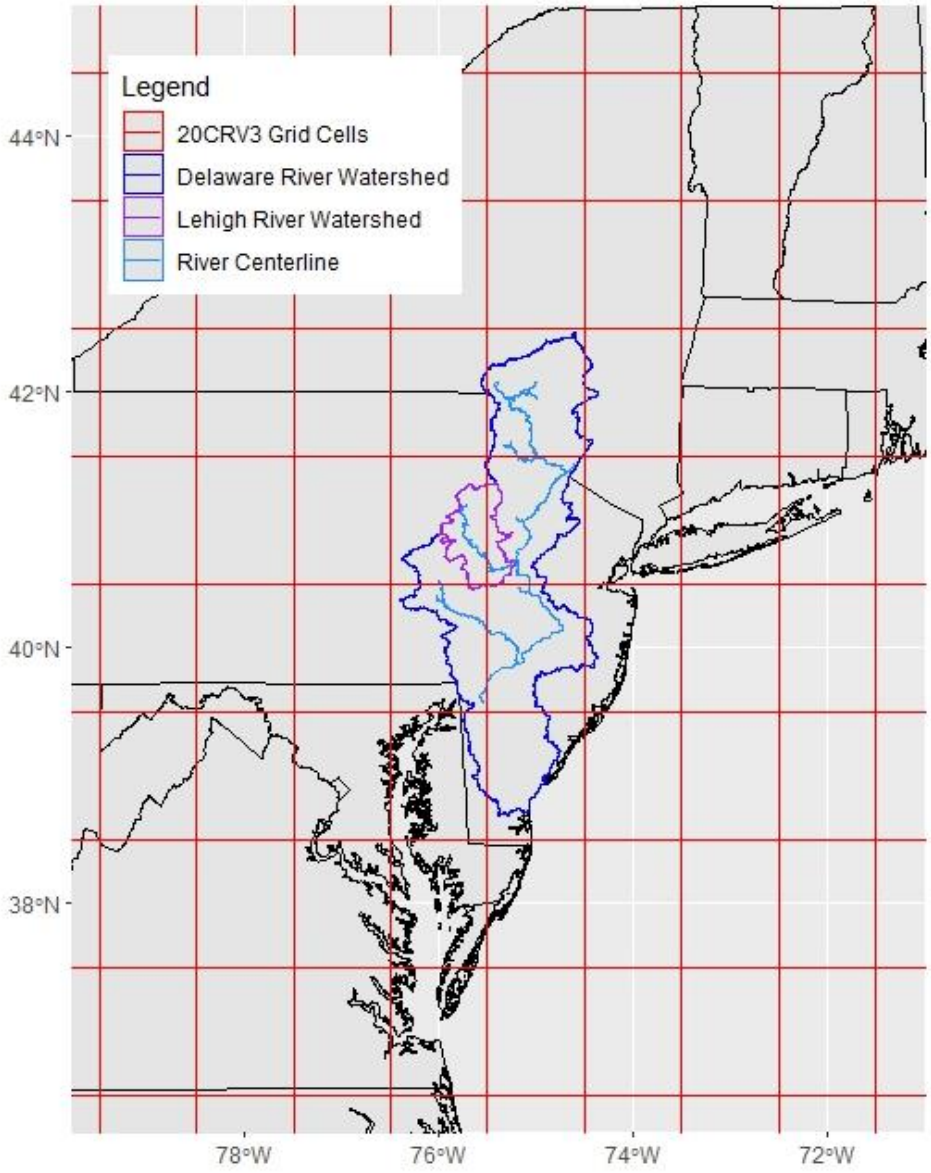


Figure 2-4. 20CRV3 grid cell coverage in the Delaware and Lehigh River watersheds

2.1.5. Tropical Cyclone Track Data

Tropical cyclones are synoptic-scale 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 tropical cyclone and tropical storm remnant (TC/TSR) tracks (Knapp

et al., 2010; Knapp et al., 2018). Datasets used to classify flood events by causal mechanism are shown in Table 2-3.

Table 2-3. Flood causal mechanism indices and datasets

Dataset Name	Meteorologic/ Land Surface Indices	Temporal Availability	Temporal Resolution	Spatial Resolution
International Best Track Archive for Climate Stewardship (IBTrACS) v4	North Atlantic cyclone tracks	23 June 1851 – 11 November 2022	Interpolated to 3 hourly from 6 hourly data	0.1° x 0.1° (6 mi x 4.5 mi)
University of Arizona (UA) Gridded SWE and Snow Depth	Snow water equivalent (SWE)	01 October 1981 – 30 September 2021	Daily	4 km x 4 km (2.5 mi x 2.5 mi)
20th Century Reanalysis Project V3 (20CRV3)	Water equivalent of snow depth (WEASD)	01 January 1836 – 31 December 2015	Daily	1.0° x 1.0° (60 mi x 45 mi)

2.2. Classification of Flood Events by Causal Mechanism

Flood causal mechanisms described in the literature are summarized in Table 2-4. Non-stationarities resulting from urbanization, water control structure operation/regulation, and agriculture are not considered flood causal mechanisms in this study. In addition, the combination of multiple causal mechanisms, such as high antecedent soil moisture coupled with a tropical cyclone, can cause catastrophic flooding.

Table 2-4. Flood causal mechanisms

Flood Causal Mechanism Category	Spatial Scale	Examples
Meteorological	Global	El Niño–Southern Oscillation Atlantic Multidecadal Oscillation North Atlantic Oscillation
	Synoptic	Tropical cyclones Extratropical cyclones Atmospheric rivers
	Mesoscale	Fronts Convection Large thunderstorms
	Local	Local thunderstorms Local turbulence
Land surface process	Varying	Rain-on-snow Snowmelt Antecedent soil moisture Antecedent frozen ground Ice-jams
Land disturbance/discontinuity	Varying	Wildfires

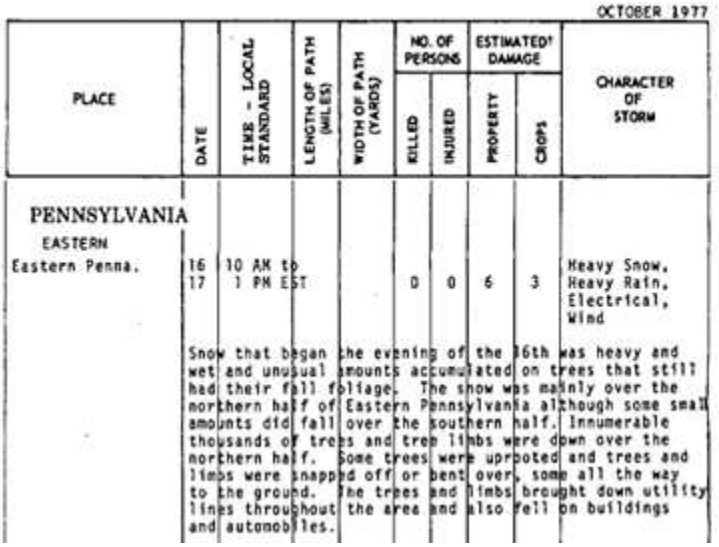
2.2.1. Seasonal Classification

When considered in flood frequency analysis, mixed populations are commonly classified using seasonal, or day-of-year, approaches (Todorovic & Rouselle, 1971; Veatch & Villarini, 2022). Seasonal classifications typically differentiate between summer and winter, or rainfall- and snowmelt-driven, flood events. Despite the variability of the date of the maximum annual flow in many parts of the United States (Berghuijs et al., 2016), seasonal classifiaction is common in practice because of its ease of use. This classification method does not separately account for extreme meteorologic events, such as tropical cyclones. In the Lehigh River watershed, the winter season was defined as December 1st through March 31st and the summer season was defined as April 1st through November 31st. These dates generally discriminate between rain-on-snow (ROS) and rainfall driven flood events.

2.2.2. Manual Classification

Robustness checks of flood type classifications are rarely performed (Tarasova et al., 2020). In this study, floods were classified manually using the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) storm data publications to improve the quality of

the automated classification and reduce uncertainty in the selection of the snowmelt threshold used to discriminate between ROS and rainfall events in the automated classification. Monthly issues of the NOAA NCEI storm publications are available online and "contain a chronological listing, by state, of storm occurrences and unusual weather phenomenon" (National Oceanic and Atmospheric Administration, 2022). Monthly storm publications are available January 1959 - November 2018. A sample storm publication description for an October 1977 event in eastern Pennsylvania is shown in Figure 2-5. Based on the storm publication description, this event was classified as a ROS event. However, using a seasonal classification as described in Section 2.2.1, this event was classified as a rainfall event.



OCTOBER 1977

PLACE	DATE	TIME - LOCAL STANDARD	LENGTH OF PATH (MILES)	WIDTH OF PATH (YARDS)	NO. OF PERSONS KILLED	NO. OF PERSONS INJURED	ESTIMATED DAMAGE PROPERTY	ESTIMATED DAMAGE CROPS	CHARACTER OF STORM
PENNSYLVANIA									
EASTERN									
Eastern Penna.	16 17	10 AM to 1 PM EST			0	0	6	3	Heavy Snow, Heavy Rain, Electrical, Wind

Snow that began the evening of the 16th was heavy and wet and unusual amounts accumulated on trees that still had their fall foliage. The snow was mainly over the northern half of Eastern Pennsylvania although some small amounts did fall over the southern half. Innumerable thousands of trees and tree limbs were down over the northern half. Some trees were uprooted and trees and limbs were snapped off or bent over, some all the way to the ground. The trees and limbs brought down utility lines throughout the area and also fell on buildings and automobiles.

Figure 2-5. Sample storm publication description used to manually classify flood types (National Oceanic and Atmospheric Administration, 1977)

2.2.3. Automated Classification

The automated flood type classification used SWE and TC/TSR datasets to classify flood events caused by (1) TC/TSR, (2) ROS, and (3) rainfall (Figure 2-6). The automated classification procedure used an antecedent time period of:

$$d = \lceil 5 + ln(A) \rceil \quad (4)$$

where $\lceil x \rceil$ denotes the ceiling function and A is the drainage area in square miles.

This equation, provided in Bulletin 17 (U.S. Water Resources Council, 1976), is used as a time separation criterion to ensure independence between PDS flood events. The equation was selected to relate the antecedent period to the drainage area of the site.

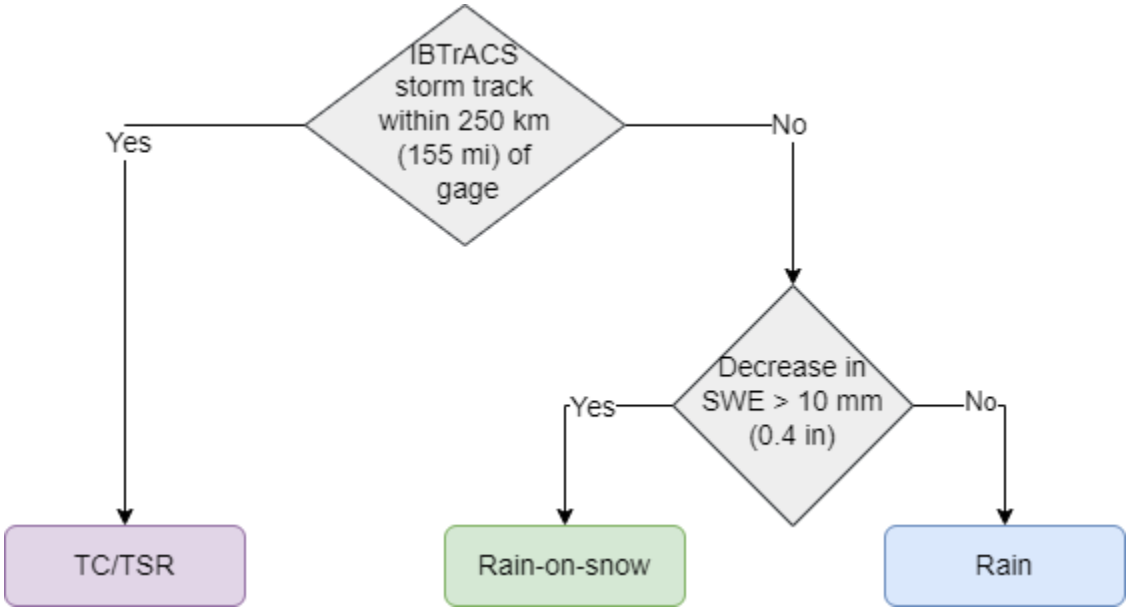


Figure 2-6. Flood typing decision tree

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 (Eqn. 4) of the flood event and (2) the TC/TSR track passed within 250 km (155 mi) of the gage location. Smith et al. (2011) attributed flood events to tropical cyclone when 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. A smaller distance was selected in this study based on agreement with the manual classification of TC/TSR driven floods (discussed in Section 2.2.2).

Volumetric contributions from liquid precipitation and SWE are typically used to differentiate between rainfall- and snowmelt-driven flood events (Kampf & Lefsky, 2016; Yu et al., 2021; Yu et al., 2022). 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 rain-on-snow and snowmelt events, using a minimum volume input of 10 mm, from either precipitation and/or snowmelt. In this study, event snowmelt was defined as the

maximum decrease in SWE in the d days (Eqn. 4) prior to a flood. Snowmelt thresholds ranging from 5 to 50 mm were evaluated (Figure 2-7). For snowmelt thresholds less than 10 mm (0.4 in), very few flood events were classified as ROS events for four of the five gages. A threshold of 10 mm was selected based on agreement with the manual classification. For the period when both UA and 20CRV3 estimates were available, events were classified as ROS if the snowmelt contribution from either data source exceeded 10 mm. 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 events were classified as rainfall events. While rainfall-induced floods vary in properties (Merz & Blöschl, 2003; Oppel & Fischer, 2020), a single classification was used in this study for a number of reasons. First, historic storm publications do not contain sufficient meteorologic descriptions to manually validate the classification of rainfall events by type (i.e. convective, frontal, orographic, etc.). The automated classification was developed based on the ability to manually validate flood types in a straightforward way that could be expanded to other regions. Other meteorologic measurements, such as balloon soundings and surface pressure maps, would allow for manual validation of different rainfall mechanisms. Second, while the 20CRV3 product contains multiple measures of convection (including convective available potential energy and convective precipitation accumulation), the product has a coarse spatial resolution and was development from the assimilation of surface measurements (Compo et al., 2011; Slivinski et al., 2019). In developing an ingredients-based approach for identifying and forecasting flash flood events, Doswell et al. cautioned that "[t]hresholds can be intellectual traps for the unwary" (Doswell et al., 1996). Finally, subdivision of rainfall-driven floods into multiple subsets reduces the sample size of each subset and can produce instability in probability distribution parameter estimates. The purpose of classification of flood events in this study was to ascertain hydrologic, rather than meteorologic, differences in floods.

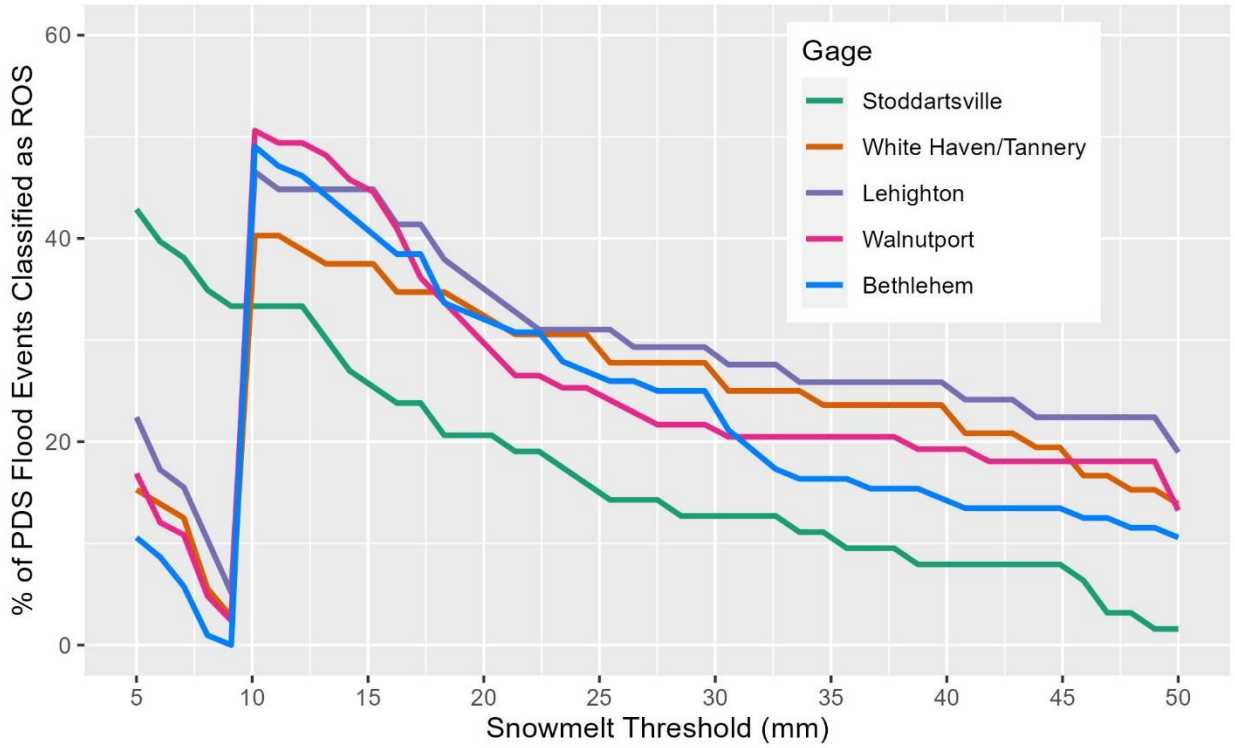


Figure 2-7. Percent of PDS flood events classified as rain-on-snow (ROS) vs. snowmelt threshold

The automated and manual flood type classifications were compared by computing the number of PDS flood events in each category during 1958-2018, the period when NOAA NCEI historic storm publications were available. The manual flood classifications were used to validate the automated classification procedure.

2.3. Temporal Resolution Conversion

Mean daily streamflow was used to select flood events. However, frequency-based flood flow estimates are typically made using instantaneous or sub-daily data. The available record of instantaneous and sub-daily data (e.g. 15-, 30, and/or 60-minute streamflow) is typically shorter than that of mean daily streamflow. In addition, sub-daily streamflow records are not always continuous, resulting in data gaps.

Quantile mapping was used to develop a daily-to-instantaneous streamflow relationship. Quantile mapping is a bias correction method to simulated data such that its distribution matches that of the observed data (Piani et al., 2010). A transfer function (T) is derived by calculating the cumulative distribution functions of the simulated and observed datasets and then associating each simulated value to

the observed value such that the theoretical cumulative distribution function ($\hat{F}$) of the simulated response ($\hat{y}$) is equivalent to the cumulative distribution function (F) of the observed response (y) (Piani et al., 2010):

$$\hat{F}(\hat{y}) = F(y) \quad (5)$$

and hence,

$$\hat{y}_{corrected} = \hat{F}^{-1}[F(y)] \quad (6)$$

where $\hat{F}^{-1}(.)$ represents the quantile estimation equation for the simulated response of the system. The term $\hat{F}^{-1}[F(.)]$ is the quantile mapping transfer function. Quantile mapping is advantageous as it is an established bias correction method and independent of statistical distribution.

The R library "qmap" (Gudmundsson, 2016) was used to fit a parametric transformation to the quantile-quantile relationship of the observed and modeled values. The daily-to-instantaneous flow quantile relationship followed a power function:

$$\hat{F}^{-1}(p) = c[F^{-1}(p)]^d \quad (7)$$

where c and d are model parameters.

Pairs of mean daily and instantaneous/hourly flows, corresponding to the same flood event, were developed from pre-regulation or deregulated periods. All flow pairs were used for quantile mapping model development. The quantile mapping performance was evaluated by randomly selecting one-third of the flow pairs at each gage in 100 iterations. 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):

$$\text{RSR} = \frac{\text{RMSE}}{\text{STDEV}_{\text{obs}}} = \left[\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}\right]^{1/2} \quad (8)$$

2.4. Statistical Models for Flood Frequency Analysis

2.4.1. Bulletin 17C Flood Frequency Analysis Procedures Using Annual Maximum Series

Bulletin 17C assumes the use of block maxima, where the block is defined as a WY. In the United States, a WY spans October 1 to September 30 of the following year and is named by the year in which the measurements end (e.g., WY 2022 spans October 1, 2021 through September 30, 2022). In Bulletin 17C, the Pearson Type III distribution with log transformation of the flood data (referred to as log-Pearson Type III, or LP-III) is recommended for describing the AMS. The Bulletin 17C Expected Moments Algorithm (EMA) framework enables flood quantiles 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 typically the result of anecdotal evidence, such as high-water marks on bridge abutments, rather than from instrumentation. Finally, periods of missing data are described using perception thresholds. Perception thresholds describe the range of observable or perceptible discharges. The missing values are within a range, but the exact value is unknown.

EMA uses an iterative approach to fitting of the LP-III distribution to flood data (England et al., 2019). First, LP-III distribution parameters (mean, standard deviation, and skew), computed using the method of moments, are estimated from the systematic (or exactly known) data. Next, distribution parameters are estimated using both systematic and censored (historical and perception threshold) data. Since the moments of the censored data aren't exactly known, the procedure must be repeated until the moments converge upon a stable solution.

The Bulletin 17C procedure recommends the use of a weighted average of station skew and regional skew since station skew estimates are sensitive to outliers. Regional skew information for the Lehigh River basin was obtained from a previous study (Konieszczki & Goldman, 2009). A generalized least squares constant regional average skew and skew mean square error of 0.165 and 0.030, respectively, were used

with to stabilize station skew coefficients. Regional skew information for the Lehigh River was developed from flood flows without separation by causal mechanism (Konieszczki & Goldman, 2009). Within Bulletin 17C analyses, the empirical exceedance estimates of flood data are used to evaluate the adequacy of the estimated flood frequency relationships. The Hirsch-Stedinger (Hirsch & Stedinger, 1987) plotting position formula is recommended to incorporate all available data, including censored data:

$$P(i) = \frac{n_e}{n}\left(\frac{i - A}{k + 1 - 2A}\right) \quad (9)$$

where $P(i)$ is the exceedance probability of the i^{th} observed annual maxima, i is the rank of the annual maxima flood in a descending order of magnitude (i = 1 corresponds to the largest flood event), A is a constant, n is the total number of years of record including the threshold, and n_e is the number of annual maxima floods above the threshold.

The pre-regulation and unregulated hourly and instantaneous flow values were used to compute empirical AEP using the Hirsch-Stedinger plotting position formula (Eqn. 9). In some cases, the causal mechanisms for the hourly AMS could not be determined because unregulated flow data from previous studies and historical flow events were recorded by WY, without the complete date of occurrence.

The Hydrologic Engineering Center's Statistical Software Program (HEC-SSP) version 2.3 was used to perform flood flow frequency analyses in accordance with Bulletin 17C. Historical data for the Lehigh River gages was available from a previous basin study (U.S. Army Corps of Engineers, 1985). Specifically, the May 1942 and August 1955 flood events were the largest flows since at least 1786 for the Bethlehem gage and 1902 for the remaining gages. Perception thresholds were used to describe the period of missing data during the systematic record. A perception threshold of 32,000 cubic feet per second (cfs) was selected for the Bethlehem gage, shown in Figure 2-8, 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 would have been recorded had they occurred). The perception threshold at each gage was assumed to be constant over the historical period. Paleoflood data were not available in

the Lehigh River basin. In HEC-SSP, the Hirsch-Stedinger plotting position is computed with constant A set to 0 (Hydrologic Engineering Center, 2023).

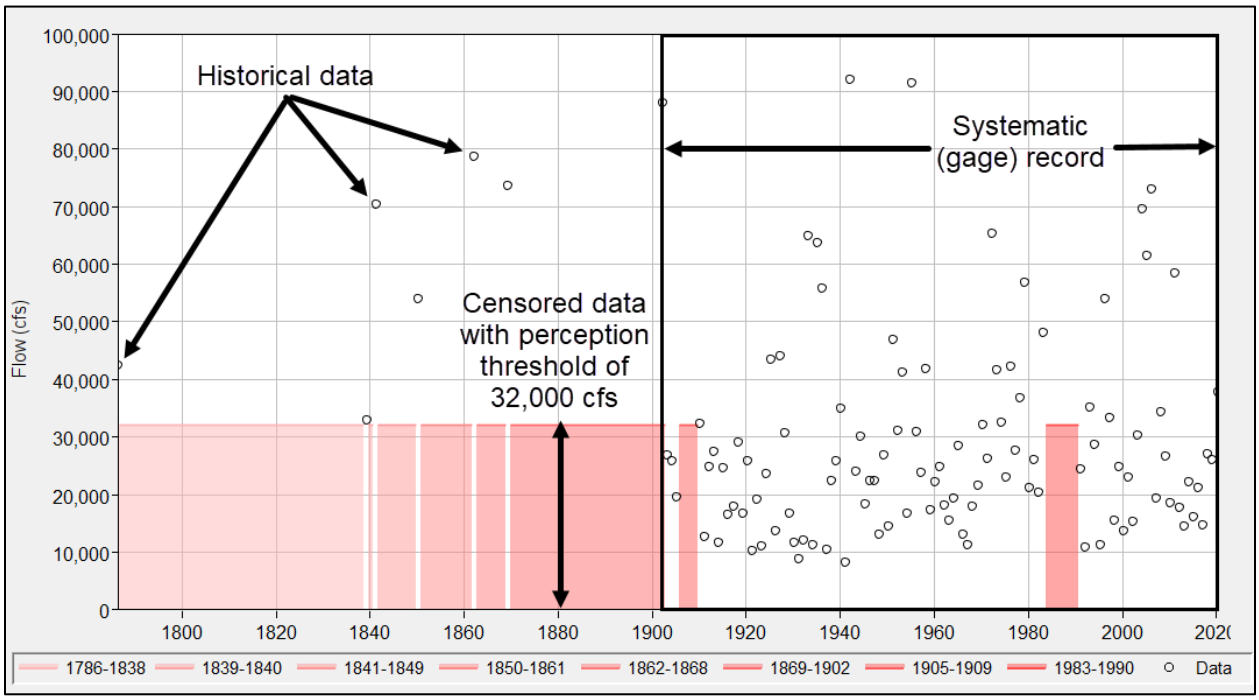


Figure 2-8. Treatment of systematic, historical, and censored data in a Bulletin 17C Analysis in HEC-SSP

2.4.2. Flood Frequency Analysis Using the Partial Duration Series

The development of a PDS requires the selection of a statistical threshold, above which flow values are considered extreme events. To model exceedances above a threshold, a sufficiently high threshold u is selected such that the values of X that exceed u are considered extreme events (Naghettini, 2017). The exceedances of X over u are designated as $W = X - u$ for $X > u$. The magnitude of the value above the threshold u is referred to as an "excess." The GP distribution was used to model the excesses.

The non-exceedance probability for the GP distribution is given by (Madsen et al., 1997b):

$$F(x) = Pr(X \leq x \,|x > u) = 1 - \left[1 - \frac{\kappa}{\alpha}(x - u)\right]^{1/\kappa} \quad (10)$$

where u, κ, and α represent location (threshold), scale, and shape parameters, respectively.

The GP distribution becomes the 2-parameter exponential distribution when the shape parameter κ^* is zero. Selection of the threshold parameter u for the GP distribution is an important consideration. The smallest threshold value such that the GP distribution is a reasonable approximation of the tail behavior is selected (Naghettini, 2017). Three diagnostic graphs were used to select the flow threshold: (1) GP distribution shape parameter, computed using L-moments (Hosking & Wallis, 1997), (2) mean excess, and (3) number of exceedances.

The expected value of the magnitude of exceedances above the threshold u, given a threshold of $u > u_0$ and following a GP distribution with scale and shaper parameters α and $\kappa > -1$, is given by:

$$E[X - u \,|X > u] = \frac{\alpha}{1 + \kappa} + \left(\frac{-\kappa}{1 + \kappa}\right) u \qquad (11)$$

Therefore, the expected value must be a linear function of u, with slope coefficient $-\kappa/(1 + \kappa)$. The resulting exploratory method is referred to as the mean residual life plot, or mean excess plot, and depicts varying values of the threshold u and the corresponding mean exceedance values. Typically, the threshold value is selected when the mean residual life plot's behavior becomes nonlinear, or there is a break is the slope (Naghettini, 2017).

Another method to select a GP distribution threshold is through evaluation of parameter stability. If excess values can be modeled by a GP distribution above a threshold u_0, then excesses of a higher threshold u can also be modeled using a GP distribution and the two distributions will have the same shape parameter (Coles, 2001). Therefore, the smallest threshold value corresponding to a stable shape parameter is selected (Naghettini, 2017).

Finally, one event per year of record is often used a GP distribution threshold selection heuristic. The number of exceedances was plotted as a function of the threshold. Diagnostic plots used to select the GP distribution threshold for each gage are shown in Appendix C.

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

$$\lambda = N/t \quad (12)$$

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

The GP distribution was fit to the PDS flood events at each gage using the L-moments method (Hosking & Wallis, 1997). The r^{th} L-moment of a sample $\{x_1, x_2, \ldots, x_n\}$ is defined as (Hosking, 1990):

$$l_r = r^{-1}\binom{n}{r}^{-1} \sum_{x_1 < \cdots < x_j < \cdots < x_r} (-1)^{r-j} \binom{r-1}{j} x_j \quad for\ r = 1, \ldots, n \quad (13)$$

where $x_1 < \cdots < x_j < \cdots < x_r$ are the r -element ordered subsets of the sample. The R package "lmom" was used to compute the sample L-moments (Hosking J. R., 2022). The L-moment estimator of the shape parameter of the GP distribution is:

$$\hat{k} = \frac{l_1 - u}{l_2 - 2} \quad (14)$$

Similarly, the scale parameter is estimated by:

$$\hat{\alpha} = \frac{1 + k}{l_1 - u} \quad (15)$$

The R package "lmom" (Hosking, 2022) was used to estimate distribution parameters for each flood type using L-moments.

2.4.3. Conversion from PDS and GP Distribution to GEV Distribution

Extreme value distributions are the limiting distributions for the minimum and maximum values of a large sample of random events. The GEV distribution family is used to model IID random block maxima events (Jenkinson, 1955) while the GP distribution is used to model random IID threshold exceedances. The GP distribution was used to model the PDS flood events at each gage. However, the GP distribution

describes conditional exceedance, not AEP. AEPs are needed in flood frequency analysis to compute annual flood losses and benefits in the planning and design of water management systems (Foster, 1941) 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 the Poisson process (Madsen et al., 1997a):

$$\lambda = N/t \quad (16)$$

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 parameters location u, scale α, and shape κ and GEV distribution parameters location ξ, scale α^*, and shape κ^* are (Madsen et al., 1997b):

$$\xi = u + \frac{\alpha}{\kappa}(1 - \lambda^{-\kappa}) \quad (17)$$

$$\alpha^* = \alpha\, \lambda^{-\kappa} \quad (18)$$

$$\kappa^* = \kappa \quad (19)$$

The non-exceedance probability for the GEV distribution is given by (Naghettini, 2017)

$$F(x) = exp\left\{-\left[1 - \kappa\left(\frac{x - \xi}{\alpha}\right)\right]^{1/\kappa}\right\} \quad (20)$$

2.4.4. Flood Frequency Analysis Methods

Since regional skew information for the Lehigh River was developed from a mixed flood series, the information was only used in the mixed frequency analyses to obtain a weighted skew. The combined frequency analyses used at-site skew information.

Bulletin 17C recommends a minimum systematic record length of 10 years to warrant statistical analysis (England et al., 2019). TC/TSR events have caused some of the most severe floods in the Mid-Atlantic but constitute less than 15% of annual flood peaks (Smith et al., 2011; Sun et al., 2021). As a result, the

sample sizes of TC/TSR floods are often too small to fit reasonable analytical probability distributions. In addition, the TC/TSR flood events at the 5 Lehigh River gages were produced by the same storm events. Consequently, regionalization was not pursued because the few sample flood events were correlated.

The AEP of TC/TSR flood events was computed using an empirical plotting position formula, as in Murphy (2001).

$$P(Q > q) = P(Q > q \mid T_3)P(T_3) \quad (21)$$

where q is a flood quantile (cfs) and T_3 is floods caused by mechanism type 3 (TC/TSR).

The empirical AEP was computed using the Gringorten plotting position, which is optimized for a Gumbel, or type I Extreme Value distribution (Maidment, 1993):

$$P(Q > q \mid T_3) = \frac{i-0.44}{n_3+0.12} \quad (22)$$

where i is the rank of the annual maxima flood in descending order of magnitude (i = 1 corresponds to the largest flood event) and n_3 is the number of annual peak floods caused by mechanism 3 (TC/TSR).

The probability of a flood caused by mechanism 3 is computed as:

$$P(T_3) = \frac{n_3}{n} \quad (23)$$

In the combined frequency analysis methods (Table 2-1), the TC/TSR, ROS, and rainfall statistical models were combined using the probability of union. The combined probability of flooding exceeding a threshold in a given year from m causal mechanisms is computed as:

$$P_c = 1 - \prod_{i=1}^{m}(1 - P_i) \quad (24)$$

where P_i denotes the annual probability of flooding from the i^{th} causal mechanism. In this study, m is 3 corresponding to the TC/TSR, ROS, and rainfall events. Floods caused by one mechanism are assumed to be independent of floods caused by the other mechanisms. The combined flood frequency analysis

method that used PDS (PCD Table 2-1) was computed by analytically converting GP distribution parameters for the ROS and rainfall floods to GEV/AMS models and then combining exceedance probabilities with those from the TC/TSR empirical probability distribution.

Unlike Bulletin 17C procedures, the GP-GEV model, fit using L-moments, does not incorporate historical and paleoflood information. For the analyses that used empirical plotting positions, the frequency curves were not extrapolated beyond the empirical exceedance probability of the largest observed TC/TSR flood. Extension of empirical probability distributions poses a challenge and is often coupled with hydrologic modeling and precipitation frequency analyses to enable credible extrapolation to rare AEPs.

3. RESULTS

3.1. Flood Type Classification

The automated and manual flood type classifications showed agreement (Figure 3-1). Four flood type categories were used in the manual classification. Flood events that were missing a description in the monthly historic storm publications were categorized under the "No Data" flood type (Figure 3-1). Validation of the automated classification for these flood events was not possible. The automated classification correctly categorized all TC/TSR floods and was successful at differentiating between rainfall and ROS floods.

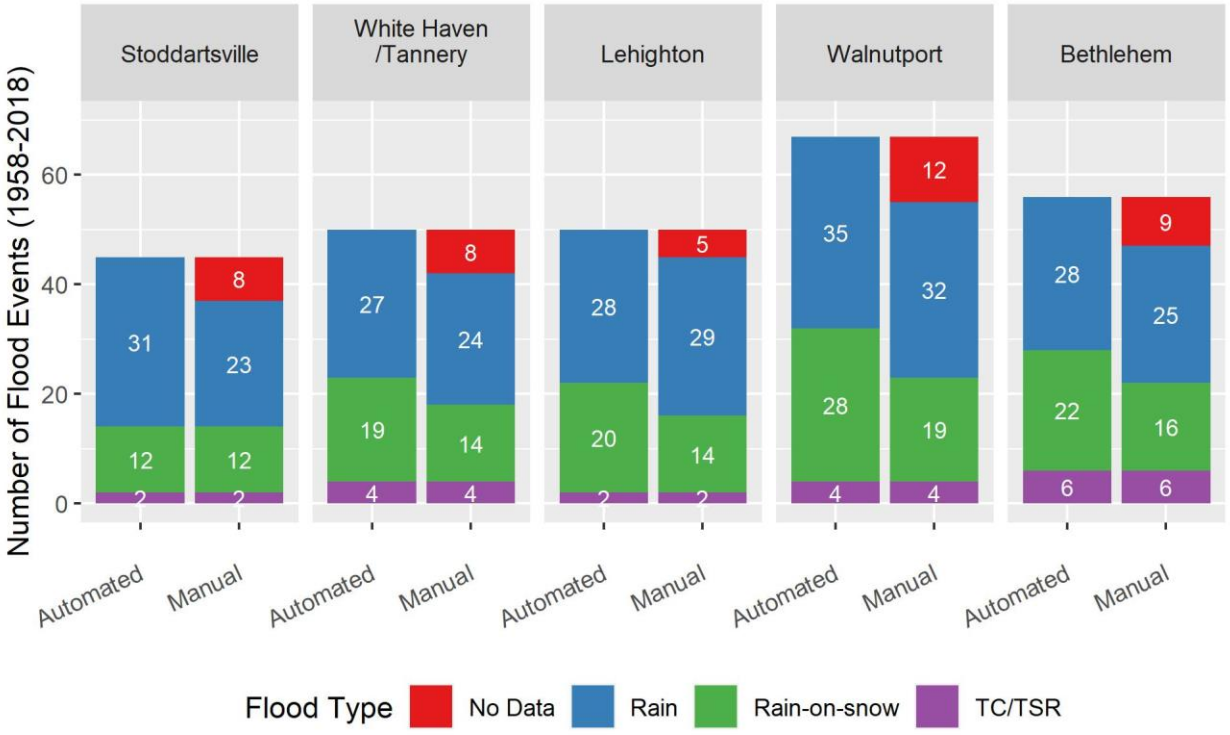


Figure 3-1. 1958-2018 PDS flood events by flood type using automated and manual classification methods

Rain flood events comprised 50-70% of the PDS floods at the Lehigh River gages while 25-40% and 5-10% of flood events were caused by rain-on-snow and TC/TSR, respectively (Figure 3-2).

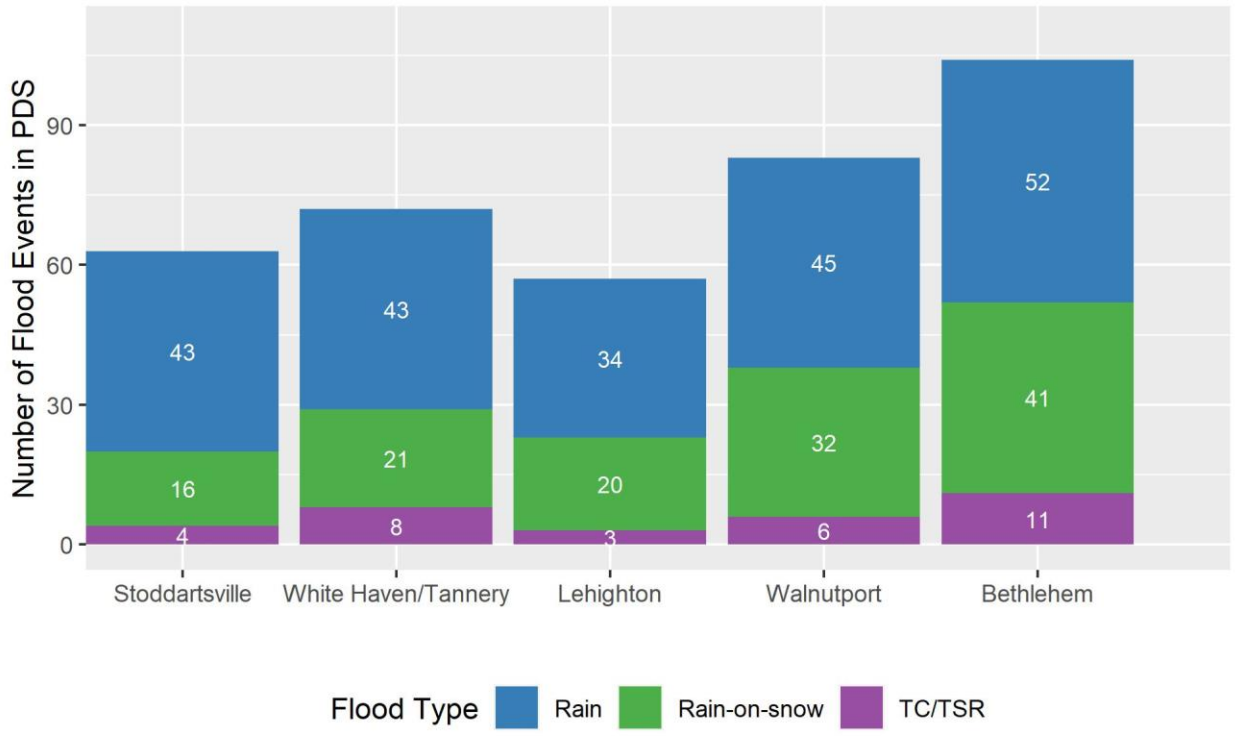


Figure 3-2. PDS flood events by flood type using automated classification methods

3.2. Temporal Resolution Conversion

The quantile mapping model parameters and validation results are shown in Table 3-1. For each of the 100 iterations, the RSR of the instantaneous flow from the quantile mapping was computed (Figure 3-3). 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 longer records.

Table 3-1. Daily-to-Instantaneous flow quantile mapping model parameters and validation results

Gage Short Name	Total Number of Flood Pairs	Parameter c	Parameter d	Model RSR
Stoddartsville	63	0.71	1.09	0.09
White Haven/Tannery	100	0.44	1.12	0.07
Lehighton	23	0.83	1.05	0.18
Walnutport	37	1.61	0.98	0.13
Bethlehem	66	0.90	1.03	0.09

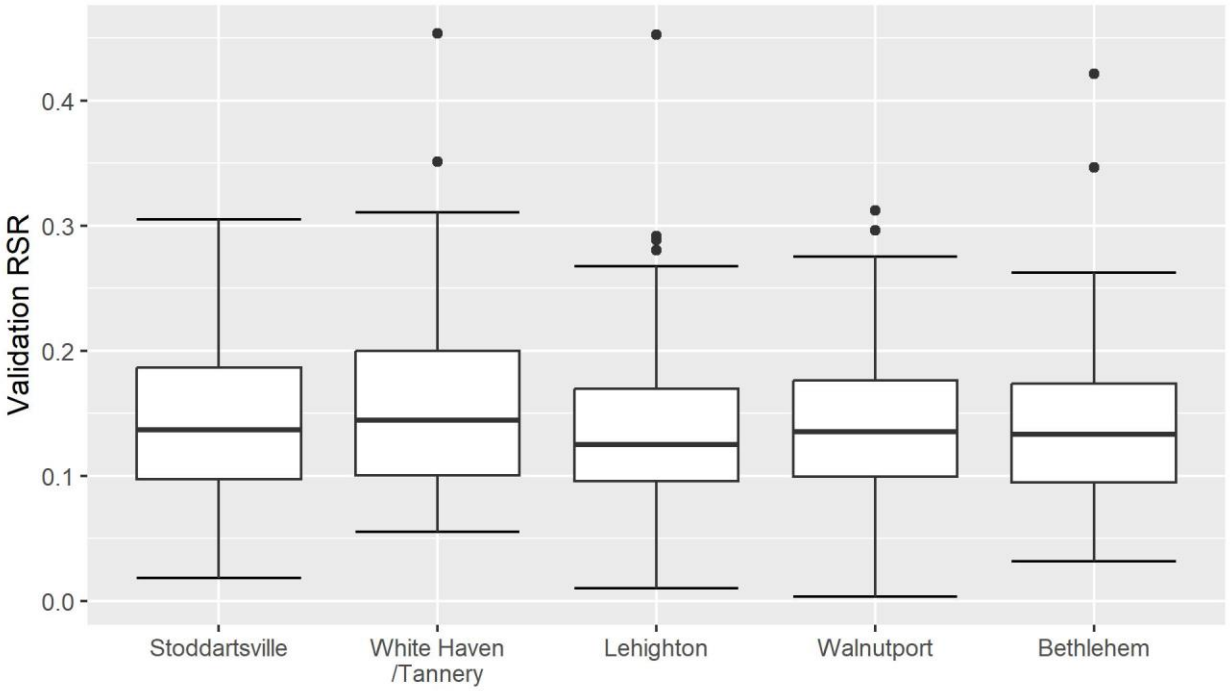


Figure 3-3. Boxplot of validation RSR values

3.3. Flood Frequency Analyses

The results of the six (6) flow frequency analyses methods (Table 2-1) are presented in this section. TC/TSR events produced some of the largest flow events at each of the gages (Figure 3-4). However, at the two most downstream gages, the largest flow event was produced by non-TC/TSR rainfall. Many of the smallest annual peak flows along the Lehigh River were caused by rain-on-snow events. The Hirsch-Stedinger empirical flood quantiles are substantially higher than the mixed LP-III analytical flood quantiles for AEPs of 0.02 and 0.01 (50- and 100-year return periods) at the Stoddartsville, White Haven/Tannery, and Walnutport gages.

The AMD and AMH analyses produced similar quantile estimates (Figure 3-5 and Figure 3-6 AMD), indicating that the use of quantile mapping to convert observed mean daily discharges to instantaneous discharges yields excellent results. The ability to convert the temporal resolution of streamflow data yields longer data records, which is beneficial when predicting the magnitude of rare flood events. At small AEPs (rare events), the AMD curve produced smaller flood quantiles than the AMH curve at the Stoddartsville gage and larger flood quantiles than the AMH curve at the other four gages. Watershed runoff response is typically flashy in small drainage areas. The assumption that the non-exceedance probability of mean daily and the corresponding instantaneous discharge is equivalent may not be appropriate for small drainage areas.

Flood flow quantiles produced by combined frequency methods were larger than those produced from mixed frequency methods, particularly at small AEPs. The combined frequency analysis methods (ACD_Seasonal, ACD_Causal, and PCD) yielded the largest 1% quantile estimates (Figure 3-6 top row and Figure 3-7). At the 1% AEP, the quantile estimates from the ACD_Causal and PCD methods were the same since they are based on the empirical plotting positions of observed TC/TSR flows.

The ACD_Seasonal method groups TC/TSR flows with other summer (April 1st through November 31st) annual peak flow values. Using seasons to subset flood types resulted in larger sample sizes than the analyses that subset floods by causal mechanisms and allowed for the use of analytical probability distributions.

The AMH analysis produced the smallest 1% AEP flow quantile at most gages. Flood frequency analyses are typically performed with the AMH approach, using instantaneous or hourly annual maxima flows without separation by flood causal mechanism. These results imply the traditional flood frequency analysis approach underpredicts flood quantiles, particularly those corresponding to small AEPs. The relative rank of the PMD flow quantiles, with respect to the other analysis methods, varied by gage, primarily due to the length of record. For example, the Lehighton gage has the shortest record length of the gages analyzed (WY 1983-2022). In addition, the Lehighton systematic record does not include major flood events in WY 1936, 1942, and 1955. The inability of the GP-GEV statistical model, fit using L-

moments, to include historical information results in an underprediction of the 1% AEP quantile estimate from the PMD analysis. On the other hand, the PMD flow quantiles are higher than those estimated from the AMH analysis for the remaining four gages, which have longer systematic record lengths (Figure 3-6 PMD). Therefore, the use of PDS, rather than AMS, has the potential to increase the predictive capability of a flood series.

LP-III distribution parameters from the analyses using Bulletin 17C procedures were computed. Station skew values for both mixed analyses (AMH and AMD) were high, with all but one value exceeding 0.5 and some values exceeding 1.0 (Table 3-2). This is expected because strongly positive skew values are typically indicative of mixed populations, often including large floods from rare meteorologic events (Beard, 1962). However, even after flood type separation, some station skew estimates remained above 1.0. There were significant differences in station skew estimates between the two combined frequency Bulletin 17C analyses, ACD_Causal and ACD_Seasonal (Table 3-3). For example, winter flood skew estimates from the ACD_Seasonal analyses (0.01 – 0.30) were substantially lower than ROS flood skew estimates from the ACD_Causal analyses (0.54 – 1.3). The two fitted distributions also exhibit a tradeoff between standard deviation and skew: the seasonal classification resulted in lower station skew values than the causal mechanism-based classification, but the standard deviation values were higher. The difference in LP-III parameters between the seasonal and causal mechanism-based approaches to flood type separation further proves the inability of seasons to serve as proxies for flood causal mechanisms.

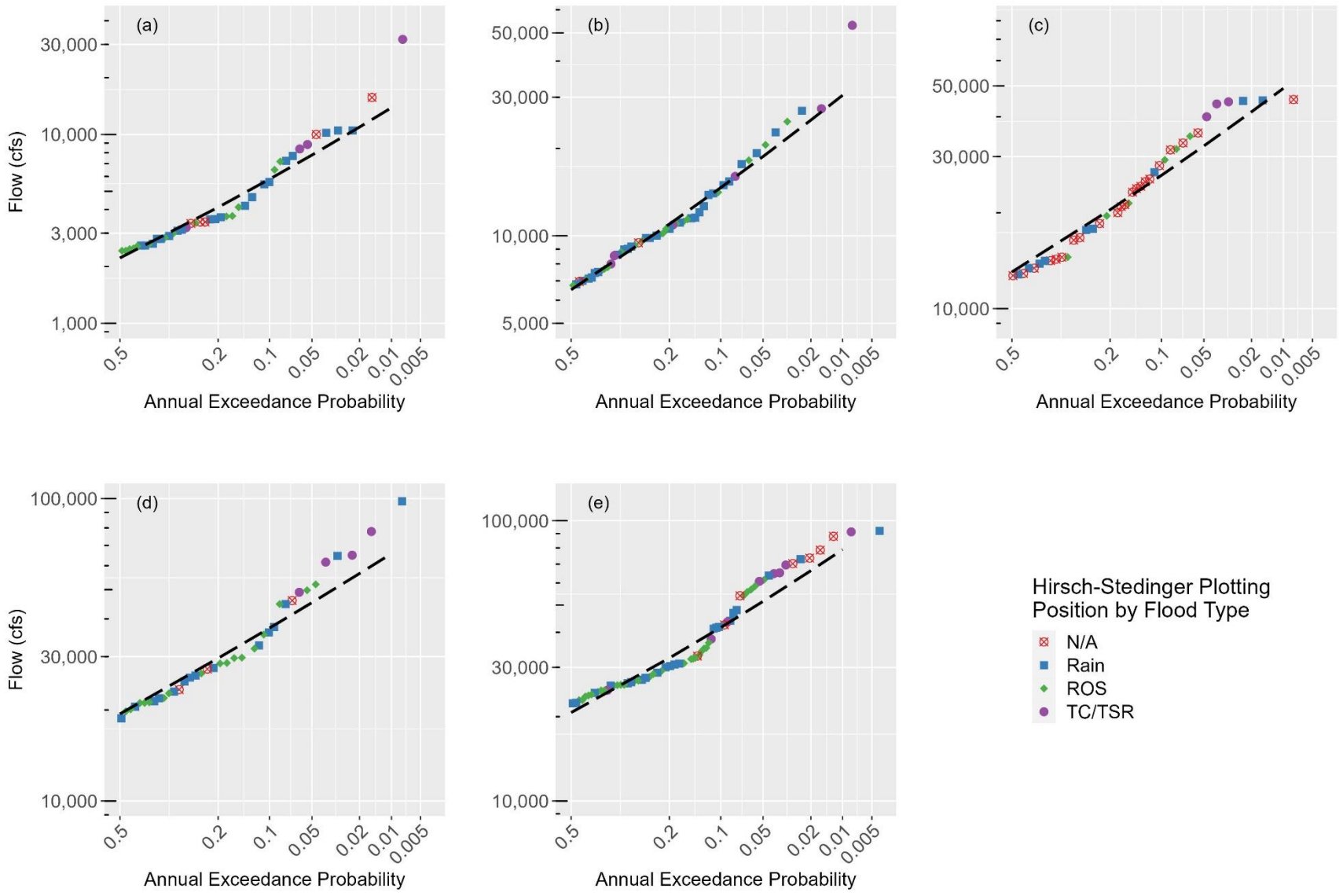


Figure 3-4. AMH flood frequency curves for (a) Stoddartsville, (b) White Haven/Tannery, (c) Lehighton, (d) Walnutport, and (e) Bethlehem with Hirsch-Stedinger plotting positions (H-S PP) by flood type

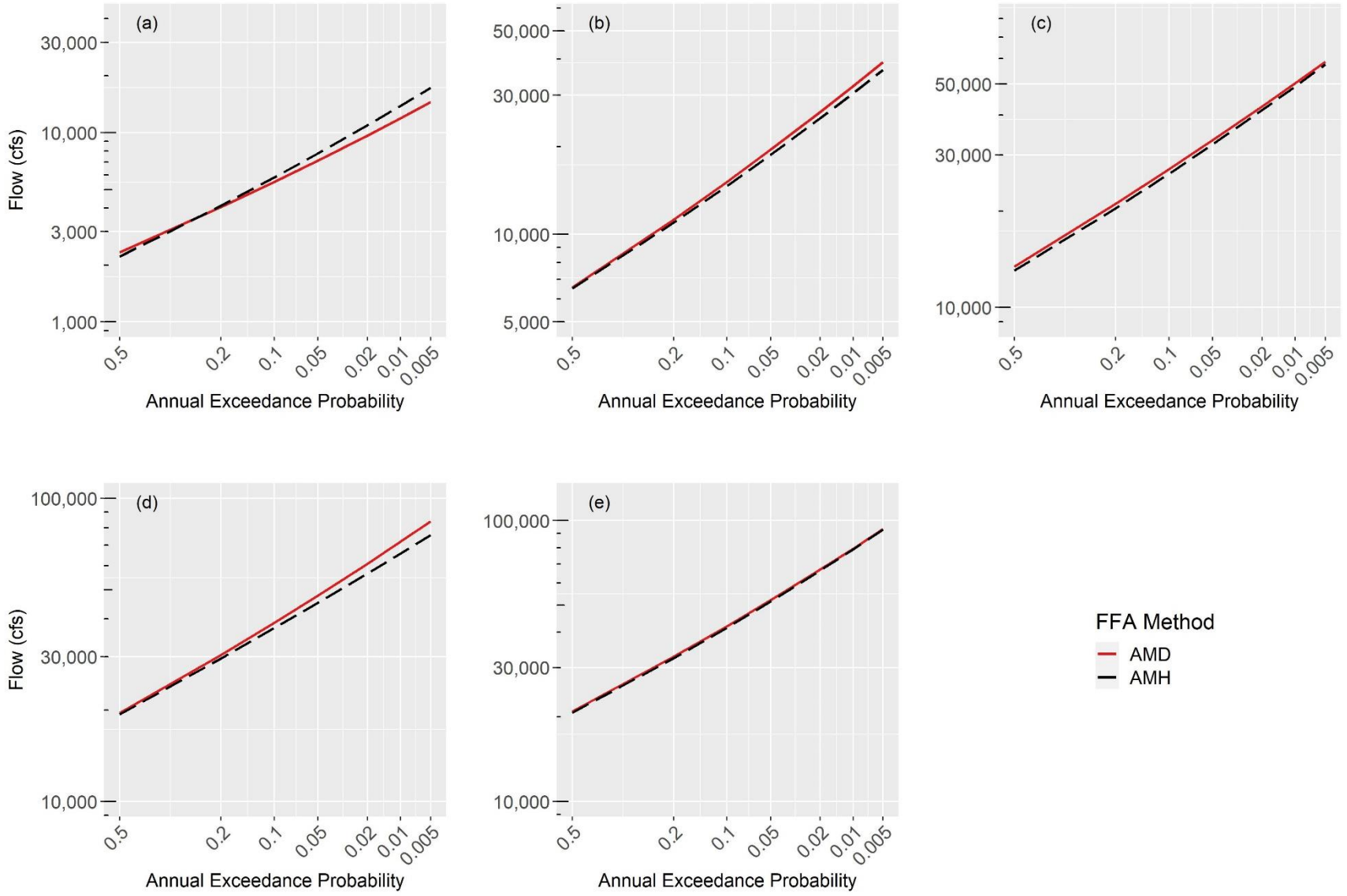


Figure 3-5. AMH and AMD flood frequency curves for (a) Stoddartsville, (b) White Haven/Tannery, (c) Lehighton, (d) Walnutport, and (e) Bethlehem

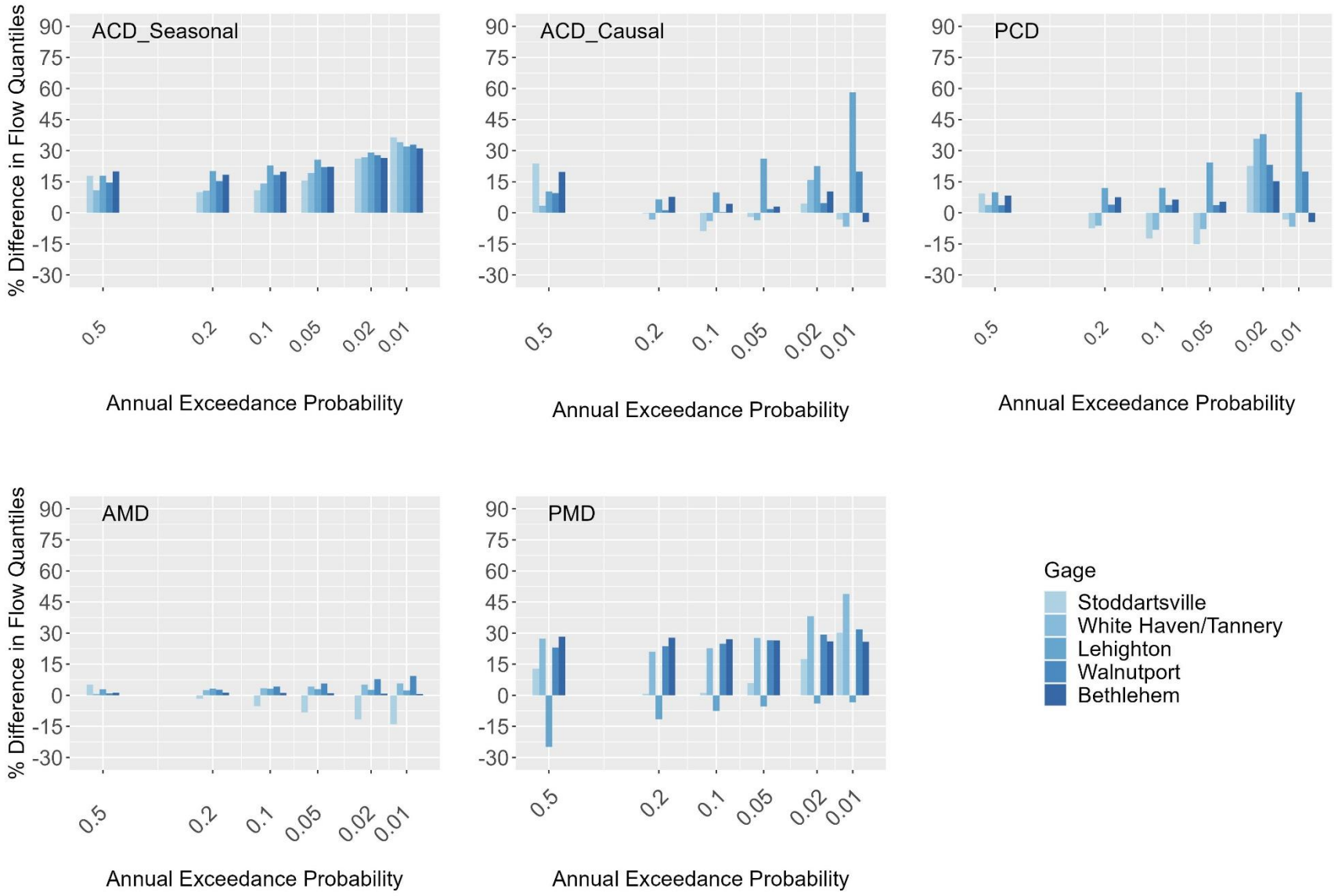


Figure 3-6. Percent difference in flow quantiles relative to AMH flow quantiles

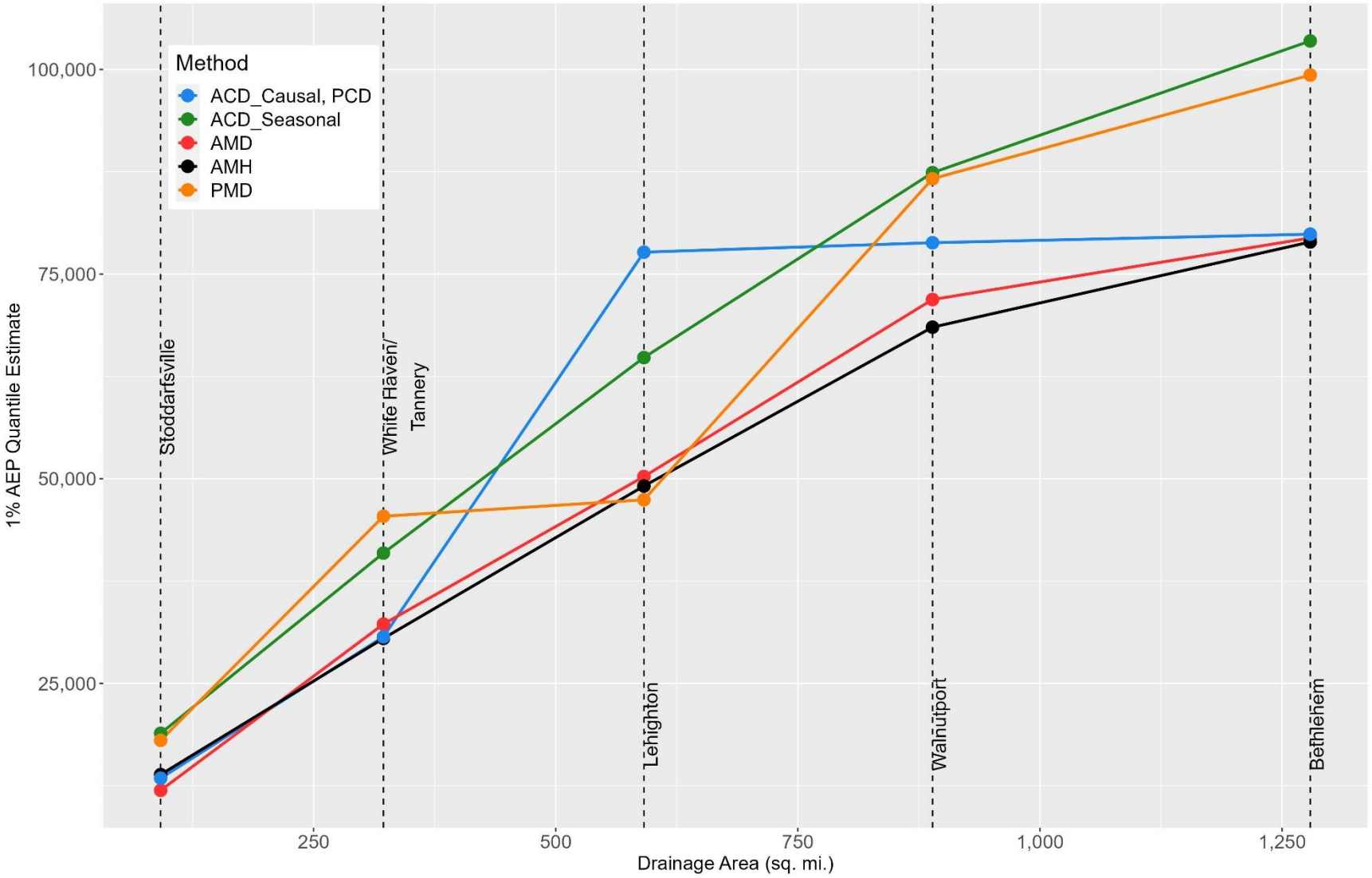


Figure 3-7. 1% AEP quantile estimate vs. drainage area

Table 3-2. LP-III parameters for mixed Bulletin 17C flood frequency analyses

Gage	AMH			AMD		
	Mean log(Q)	Standard Deviation log(Q)	Station Skew	Mean log(Q)	Standard Deviation log(Q)	Station Skew
Stoddartsville	3.361	0.307	0.725	3.381	0.274	0.698
White Haven/Tannery	3.827	0.259	0.630	3.830	0.268	0.570
Lehighton	4.125	0.224	0.647	4.136	0.227	0.389
Walnutport	4.294	0.214	1.002	4.302	0.220	0.560
Bethlehem	4.328	0.222	0.706	4.332	0.223	0.586

Table 3-3. LP-III parameters for combined Bulletin 17C flood frequency analyses

Gage	ACD_Causal			ACD_Seasonal		
	Mean log(Q)	Standard Deviation log(Q)	Station Skew	Mean log(Q)	Standard Deviation log(Q)	Station Skew
Stoddartsville	3.313 (ROS) 3.338 (Rain)	0.177 (ROS) 0.260 (Rain)	0.568 (ROS) 0.460 (Rain)	3.186 (W) 3.308 (S)	0.275 (W) 0.347 (S)	0.046 (W) 0.646 (S)
White Haven/Tannery	3.641 (ROS) 3.767 (Rain)	0.223 (ROS) 0.238 (Rain)	1.325 (ROS) 0.495 (Rain)	3.648 (W) 3.746 (S)	0.268 (W) 0.317 (S)	0.305 (W) 0.485 (S)
Lehighton	4.032 (ROS) 4.065 (Rain)	0.246 (ROS) 0.196 (Rain)	0.725 (ROS) 1.097 (Rain)	3.965 (W) 4.067 (S)	0.284 (W) 0.394 (S)	0.011 (W) 0.183 (S)
Walnutport	4.199 (ROS) 4.259 (Rain)	0.204 (ROS) 0.176 (Rain)	0.732 (ROS) 0.738 (Rain)	4.175 (W) 4.221 (S)	0.239 (W) 0.285 (S)	0.073 (W) 0.214 (S)
Bethlehem	4.293 (ROS) 4.298 (Rain)	0.192 (ROS) 0.189 (Rain)	0.538 (ROS) 0.756 (Rain)	4.234 (W) 4.255 (S)	0.239 (W) 0.291 (S)	0.041 (W) 0.334 (S)

W = Winter (December 1st - March 31st); S = Summer (April 1st - November 31th); ROS = rain-on-snow

Table 3-4. GEV parameters for PDS flood frequency analyses

Gage	PMD			PCD		
	Location, ξ	**Scale, α**	**Shape, κ**	**Location, ξ**	**Scale, α**	**Shape, κ**
Stoddartsville	2,140	901	-0.495	1,776 (ROS) 1,508 (Rain)	327 (ROS) 1,055 (Rain)	-0.426 (ROS) -0.248 (Rain)
White Haven/Tannery	7,107	3,048	-0.382	4,504 (ROS) 4,815 (Rain)	1,070 (ROS) 2,280 (Rain)	-0.481 (ROS) -0.254 (Rain)
Lehighton	7,660	5,451	-0.351	5,106 (ROS) 8,984 (Rain)	6,683 (ROS) 5,671 (Rain)	-0.11 (ROS) -0.159 (Rain)
Walnutport	20,207	9,708	-0.163	11,102 (ROS) 12,076 (Rain)	6,107 (ROS) 8,996 (Rain)	-0.199 (ROS) -0.061 (Rain)
Bethlehem	22,276	11,237	-0.163	11,600 (ROS) 13,432 (Rain)	8,426 (ROS) 8,764 (Rain)	-0.127 (ROS) -0.166 (Rain)

4. DISCUSSION

The automated flood type classification developed in this study performed well in the Lehigh River watershed. The automated classification was based on extensive review of historic storm publications and post-flood reports and utilized multiple spatially distributed meteorologic products. When identifying and classifying floods by causal mechanism in a flood frequency analysis, multiple lines of evidence should be used. Flood type classifications developed solely from automated approaches are less accurate and scalable than those that are based on manual classification. For example, the discrepancies in the number of rainfall and ROS events between the manual and automated classifications (Figure 3-1) were attributed to the spatial coarseness of the 20CRV3 dataset. The 1,300 mi^2 Lehigh River watershed is covered by two 20CRV3 dataset grid cells (Figure 2-4).

The results of the temporal resolution conversion show that quantile mapping can be leveraged to expand the available annual peak flow data at a site, thereby increasing the information content of a flood series. This is particularly important when using PDS approaches to flood frequency since USGS does not provide an official PDS product at gage sites. In addition, expanding the amount of flood information at a site is crucial when considering mixed populations, since flood typing results in multiple subsets with reduced sample sizes.

In this study, most flow quantiles produced from combined frequency analyses were larger than those obtained from mixed frequency analyses. Consideration of mixed populations and computation of a combined frequency curve produces higher flood quantiles for low AEPs. The GEV shape parameters obtained from the mixed and individual flood type samples were negative (Table 3-4), indicating heavy-tailed distributions (i.e. no upper bound). For three of the gages, the 1% AEP flow quantiles from the combined analyses using PDS, PCD, were lower than from the PDS mixed analyses, PMD (Figure 3-7). This result is explained by a statistical phenomenon termed “skew separation” (Dawdy & Gupta, 1995) which results in increased skew estimates when a single probability distribution is fit to samples from

multiple distributions (e.g. flood causal mechanisms). The GEV distribution shape parameter is indicative of skewness. In dividing Colorado Front Range flood events into rainfall-, snowmelt, and rain-on-snow-driven subsets, Yu et al. (2022) observed a similar pattern.

In the absence of quantitative guidance on implementation of mixed population flood frequency analysis, there is considerable variability in flood flow quantiles. In addition, day-of-year approaches to assigning flood types may be justified in the western United States, where rainfall and snowmelt driven floods exhibit strong seasonality (Berghuijs et al., 2016), but are inappropriate in the eastern United States. These results have important implications for infrastructure investment.

5. CONCLUSION

Flood frequency analyses at five sites along the Lehigh River in Pennsylvania were performed to assess the impact of consideration of mixed populations on computed flood quantiles. The results of this study demonstrate that consideration of mixed populations within a flood series is important. Treatment of a mixed flood series as homogeneous in flood frequency analysis not only violates the assumption that flood events are identically distributed but underpredicts frequency-based flood estimates. Flood frequency guidelines have been documented since 1967, with the objective of promoting "uniform application" of the procedures (England et al., 2019). Despite the acknowledgement that the "identification and treatment" of mixed populations warrants future study, federal flood frequency guidelines have not been updated to reflect this research topic.

While the instantaneous annual maximum discharges at a site have historically been used in flood frequency, the use of a PDS and mean daily discharge have the potential to increase the information content of a flood series. However, when fit using L-moments, the PDS/GP and AMS/GEV models do not incorporate historical and paleoflood information. This poses a serious disadvantage compared to the Bulletin 17C framework, particularly for sites with substantial pre-systematic information.

The automated flood type classifications used in this study show promise. A manual classification approach is more efficient for a site-specific or watershed-scale flood frequency analysis but automated procedures are scalable and 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 in the United States.

This topic of research would benefit from additional studies performed on other hydrologic regions in the United States. 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. Additional meteorologic and land surface process indices are needed to expand the flood type classification in other geographic regions and for optimization of the snowmelt threshold used to

discriminate between rainfall and snowmelt events. The manual classification and validation of the automated classification should be expanded to include meteorologic measurements. In addition, regionalization of event types that do not occur every year (e.g. tropical cyclones, atmospheric rivers, etc.) is needed to extend flood frequency curves to rare events but a method to do so has not yet been established.

Flood frequency analysis should be combined with precipitation frequency analysis and hydrologic modeling (Yu et al., 2021; Yu et al., 2022) to produce robust estimates of rare meteorologic and hydrologic events. Statistical approaches for flood frequency analysis, such as those used in this study, "do not explicitly consider the physical processes that produce floods" (Wright et al., 2014). Process-based flood frequency analysis methods are advantageous as they combine multiple meteorologic and hydrologic processes that produce floods. In addition, some process-based flood frequency analysis methods are multi-scale, meaning that they produce flood quantile estimates throughout a watershed rather than at a gage, or point, location (Wright et al., 2020). Finally, the use of stochastic simulation techniques in flood and precipitation frequency would allow for assessment of the uncertainty in probability distribution parameters and in flow quantile estimates.

REFERENCES

Balkema, A. A., & de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability, 2*(5), 792-804.

Barrett, A. (2003). *National Operational Hydrologic Remote Sensing Center SNOw Data Assimilation System (SNODAS) Products at NSIDC. NSIDC Special Report 11.* Boulder, CO: National Snow and Ice Data Center. Digital media.

Barth, N., Villarini, G., & White, K. (2019). Accounting for Mixed Populations in Flood Frequency Analysis: Bulletin 17C Perspective. *Journal of Hydrologic Engineering, 24*(3).

Barth, N., Villarini, G., Nayak, M., & White, K. (2017). Mixed populations and annual flood frequency estimates in the western United States: The role of atmospheric rivers. *Water Resources Research, 53*(1), 257-269.

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

Berghuijs, W., Harrigan, S., Molnar, P., Slater, L., & Kirchner, J. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research, 55*, 4582-4593.

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

Carter, R. W., Anderson, W. L., Isherwood, W. L., Rolfe, K. W., Showen, C. R., & Smith, W. (1963). *Automation of Streamflow Records Geological Survey Circular 474.* Washington, D.C.: U.S. Geological Survey.

Changnon, S. (2008). Assessment of Flood Losses in the United States. *Journal of Contemporary Water Research & Education*(138), 38-44.

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., Whitaker, J. S., Sardeshmukh, N., Matsui, N., Allan, R. J., Yin, X., . . . Worley, S. J. (2011). The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc., 137*(654), 1-28.

Dawdy, D. R., & Gupta, V. K. (1995). Multiscaling and skew separation in regional floods. *Water Resources Research, 31*(11), 2761-2767.

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

Doswell III, C. A., Brooks, H. E., & Maddox, R. A. (1996). Flash Flood Forecasting: An Ingredients-Based Methodology. *Weather and Forecasting, 11*, 560-581.

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*.

Fischer, S., Schumann, A., & Bühler, P. (2019). Timescale-based flood typing to estimate temporal changes in flood frequencies. *Hydrological Sciences Journal, 64*(15), 1867-1892.

Foster, E. (1941). Evaluation of Flood Losses and Benefits. *American Society of Civil Engineers Transactions*.

Friday, J. (1965). *The operation and maintenance of a crest-stage gaging station.* Portland, OR: U.S. Geological Survey.

Gaál, L., Szolgay, J., Kohnová, S., Hlavcová, K., Parajka, J., Viglione, A., . . . Blöschl, G. (2015). Dependence between flood peaks and volumes: a case study on climate and hydrological controls. *Hydrological Sciences Journal, 60*(6), 968-984.

Gershunov, A., Shulgina, T., Ralph, F., Lavers, D., & Rutz, J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters, 44*.

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

Gumbel, E. J. (1958). *Statistics of Extremes.* New York: Columbia University Press.

Hirsch, R. M., & Stedinger, J. R. (1987). Plotting positions for historical floods and their precision. *Water Resources Research, 23*(4), 715-727.

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. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society*, 105-124.

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.

Hydrologic Engineering Center. (2023). HEC-SSP User's Manual version 2.3: Plotting Positions for Interval Data. Retrieved from https://www.hec.usace.army.mil/confluence/pages/viewpage.action?pageId=111837612

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

IPCC. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.* A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NJ: Cambridge University Press.

Jarrett, R., & Costa, J. (1982). Multidisciplinary approach to the flood hydrology of foothill streams in Colorado. *International Symposium on Hydrometeorology* (pp. 565-569). Denver, CO: American Water Resources Association.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.*

Kampf, S., & Lefsky, M. (2016). Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. *Water Resources Research, 52*, 407-422.

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.

Konieszki, M., & Goldman, D. (2009). *Delaware River Basin Regional Skew Analysis PR-70.* Davis, CA: Hydrologic Engineering Center.

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

Linsley, R., Kohler, M., & Paulhus, J. (1982). *Hydrology for Engineers.* New York: McGraw-Hill, Inc.

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.

Menne, M., Durre, I., Vose, R., Gleason, B., & Houston, T. (2013). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology, 29*, 897-910. doi:10.1175/JTECH-D-11-00103.1

Merz, R., & Blöschl, G. (2003). A process typology of regional floods. *Water Resources Research, 39*(12).

Merz, R., Tarasova, L., & Basso, S. (2020). The flood cooking book: ingredients and regional flavors of floods across Germany. *Environmental Research Letters, 15*(11).

Mishra, A., Mukherjee, S., Merz, B., Singh, V., Wright, D., Villarini, G., . . . Stedinger, J. (2022). An Overview of Flood Concepts, Challenges, and Future Directions. *Journal of Hydrologic Engineering, 27*(6).

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.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Transactions of the American Society of Agricultural and Biological Engineers, 58*(6), 1763-1785.

Murphy, P. (2001). Evaluation of mixed-population flood-frequency analysis. *Journal of Hydrologic Engineering, 6*(1), 62-70.

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

National Oceanic and Atmospheric Administration. (1977, October). Storm Data. *19*(10).

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 Oceanic and Atmospheric Administration. (2023). *Flood Related Hazards*. Retrieved from National Weather Service: https://www.weather.gov/safety/flood-hazards

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

Oppel, H., & Fischer, S. (2020). A New Unsupervised Learning Method to Assess Clusters of Temporal Distribution of Rainfall and Their Coherence with Flood Types. *Water Resources Research, 56*(5).

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.

Pickands III, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics, 3*(1), 119-131.

Slater, L., & Villarini, G. (2016). Recent trends in U.S. flood risk. *Geophysical Research Letters, 43*(24).

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.

Smith, J., Cox, A., Baeck, M., Yang, L., & Bates, P. (2018). The Upper Tail of Flood Peaks in the United States. *Water Resources Research, 54*, 6510-6542.

Sun, N., Wigmosta, M. S., Judi, D., Yang, Z., Xiao, Z., & Wang, T. (2021). Climatological analysis of tropical cyclone impacts on hydrological extremes in the Mid-Atlantic region of the United States. *Environmental Research Letters, 16*.

Szulczewski, W., & Jakubowski, W. (2018). The Application of Mixture Distribution for the Estimation of Extreme Floods in Controlled Catchment Basins. *Water Resources Management, 32*(10), 3519-3534.

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

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020). A Process-Based Framework to Characterize and Classify Runoff Events: The Event Typology of Germany. *Water Resources Research, 56*.

Todorovic, P., & Rouselle, J. (1971). Some Problems of Flood Analysis. *Water Resources Research, 7*(5).

Todorovic, P., & Zelenhasic, E. (1970). A stochastic model for flood analysis. *Water Resources Research, 6*(6), 1641-1648.

Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. (2016). *Water Resources Research, 52*, 407-422.

Turkington, T., Breinl, K., Ettema, J., Alkema, D., & Jetten, V. (2016). A new flood type classification method for use in climate change impact studies. *Weather and Climate Extremes, 14*, 1-16.

Turkington, T., Breinl, K., Ettema, J., Alkema, D., & Jetten, V. (2016). A new flood type classification method for use in climate change impact studies. *Weather and Climate Extremes, 14*, 1-16.

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. (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. (2012). *USGS Hydro-Climatic Data Network 2009 (HCDN-2009).* U.S. Department of the Interior.

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

U.S. Water Resources Council. (1976). *Guidelines for Determining Flood Flow Frequency, Bulletin No. 17.* Washington, D.C.: U.S. Water Resources Council, Subcommittee on Hydrology.

Veatch, W., & Villarini, G. (2022). Modeling riverine flood seasonality with mixtures of circular probability density functions. *Journal of Hydrology, 613*.

Wang, Q. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology, 129*, 263-280.

Waylen, P., & Woo, M. (1982). Prediction of Annual Floods Generated by Mixed Processes. *Water Resources Research, 18*(4), 1283-1286.

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.

Wright, D. B., Smith, J. A., & Baeck, M. L. (2014). Flood frequency analysis using radar rainfall fields andstochastic storm transposition. *Water Resources Research, 50*(1592-1615).

Wright, D. B., Yu, G., & England, J. F. (2020). Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects. *Journal of Hydrology, 585*(124816).

Yu, G., Wright, D. B., & Davenport, F. V. (2022). Diverse physical processes drive upper-tail flood quantiles in the US mountain west. *Geophysical Research Letters, 49*.

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).

APPENDIX A: PROCEDURE TO DEREGULATE STREAMFLOW RECORDS

The following is a general procedure to compute unregulated streamflow:

1. Route the discharge leaving the reservoir(s), or reservoir outflow, from the reservoirs to the downstream gage of interest. The routed reservoir outflow hydrograph at the gage represents the contribution of flow at the gage from the reservoir storage.
2. Compute the local flow hydrograph as the difference between the systematic record at the gage and the routed reservoir outflow hydrograph. The local flow hydrograph represents the contribution of flow at the gage from the intervening drainage area, between the reservoir outlet and the gage.
3. Route the reservoir inflow (i.e. unregulated flow), assuming there is no regulation, to the gage of interest. The routed reservoir inflow represents the flow that the gage would have measured, had the dam not exited.
4. Compute the sum of the routed reservoir inflow and the local flow hydrographs at the gage. The sum represents (1) the contribution of flow from upstream of the dam location, without regulation (shaded yellow in Figure A-1) and (2) the contribution of flow from the drainage area between the dam and the gage (shaded purple in Figure A-1).

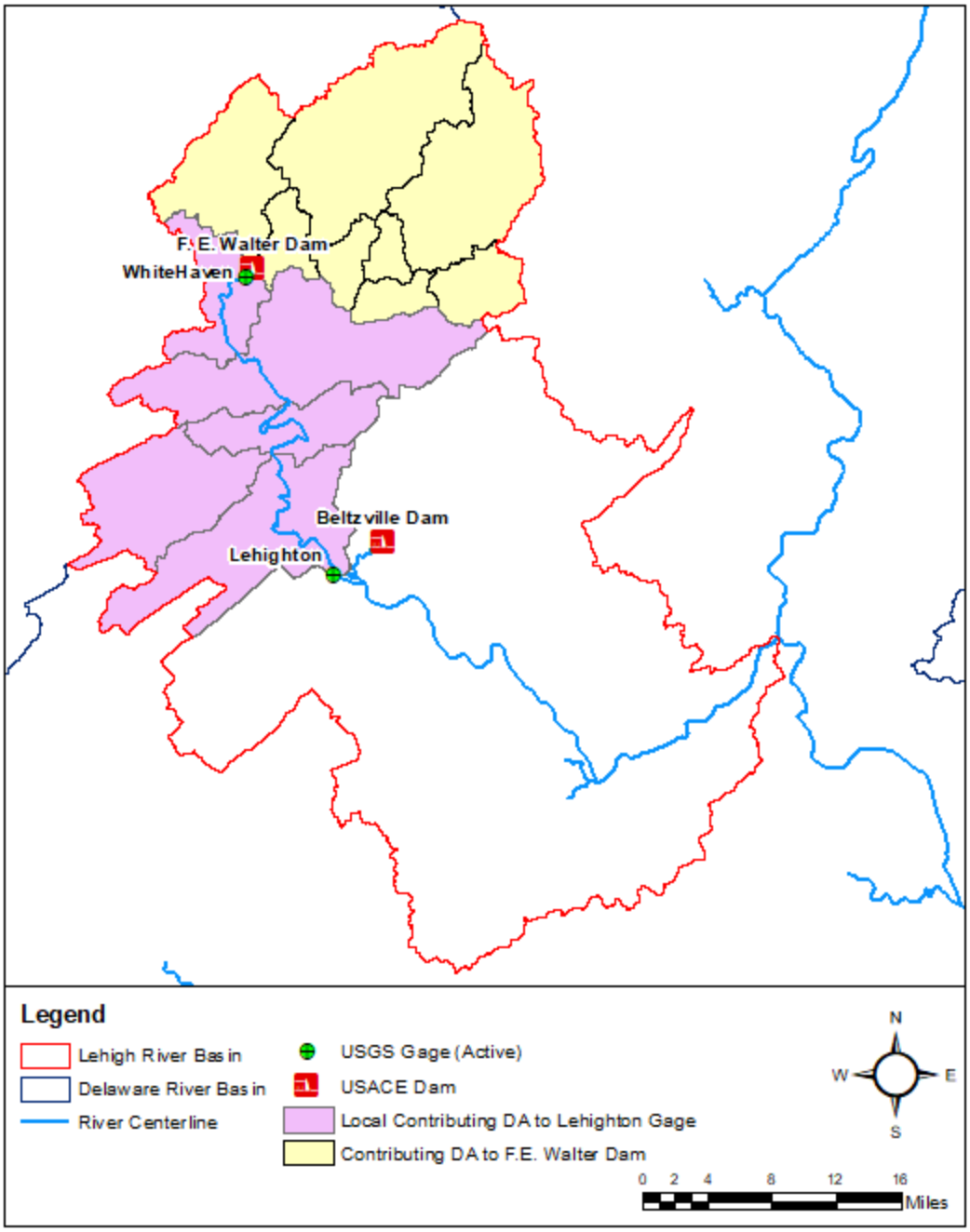


Figure A-1. Contributing drainage area at streamflow gage

APPENDIX B: UNREGUATED INSTANTANEOUS AND HOURLY ANNUAL PEAK FLOWS FOR LEHIGH RIVER GAGES

Table B-1. Legend for Table B-2

Key	Type of Data	Data Source
	Historical	Previous dam modification study (U.S. Army Corps of Engineers, 1985)
	Systematic	Pre-dam; USGS peak flow file
	Systematic	National Weather Service (NWS) stage converted to discharge based on USGS gage rating table (U.S. Army Corps of Engineers, 1985)
	Systematic	Deregulated for previous dam modification study and date of peak unregulated flow not recorded (U.S. Army Corps of Engineers, 1985)
	Systematic	Deregulated for this study using hydrologic routing and hourly streamflow records (see Section **Error! Reference source not found.**)
	Systematic	Gage is not subject to regulation; USGS peak flow file
	Censored	Historical information from previous dam modification study (U.S. Army Corps of Engineers, 1985)

Table B-2. Hourly/instantaneous AMS flows at Lehigh River gages

WY	Stoddartsville		White Haven/Tannery		Lehighton		Walnutport		Bethlehem	
	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)
1786										42,500
1839										32,900
1841										70,300
1850										54,000
1862										78,600
1869										73,600
1902						45,300			2/28/1902	88,000
1903									12/22/1902	26,800
1904									10/10/1903	25,800
1905									1/7/1905	19,600
1906										
1907										

WY	Stoddartsville		White Haven/Tannery		Lehighton		Walnutport		Bethlehem	
	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)
1908										
1909										
1910									1/22/1910	32,200
1911									8/31/1911	12,700
1912									3/15/1912	24,800
1913									3/27/1913	27,500
1914									1/31/1914	11,600
1915			1/19/1915	3,413					1/13/1915	24,500
1916									7/26/1916	16,400
1917			3/28/1917	4,050					3/28/1917	17,900
1918			10/30/1917	5,051					2/20/1918	29,100
1919			3/9/1919	2,958					3/9/1919	16,800
1920			3/12/1920	4,150					3/5/1920	25,800
1921			3/10/1921	3,595					3/4/1921	10,200
1922			3/8/1922	4,350					3/8/1922	19,100
1923			7/28/1923	3,868					3/17/1923	11,100
1924			9/30/1924	9,465					4/7/1924	23,600
1925			2/12/1925	5,151					10/1/1924	43,500
1926			11/13/1925	2,776					2/26/1926	13,600
1927			11/16/1926	11,558					11/17/1926	44,000
1928			6/30/1928	9,192					6/30/1928	30,700
1929			3/15/1929	3,950					3/6/1929	16,600
1930			6/11/1930	2,812					10/3/1929	11,600
1931			3/29/1931	5,734					2/18/1931	8,820
1932			4/1/1932	4,641					4/1/1932	12,000
1933			8/24/1933	16,018		35,600			8/24/1933	64,800
1934			4/1/1934	2,739					9/17/1934	11,300
1935			7/10/1935	17,656		33,100			7/10/1935	63,700

WY	Stoddartsville		White Haven/Tannery		Lehighton		Walnutport		Bethlehem	
	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)
1936			3/12/1936	18,202		25,500			3/12/1936	55,700
1937			4/6/1937	3,094					1/26/1937	10,500
1938			10/23/1937	7,190					10/23/1937	22,400
1939			12/6/1938	5,734		12,900			12/6/1938	25,900
1940			3/31/1940	11,012					3/15/1940	35,000
1941			4/6/1941	3,167					4/6/1941	8,210
1942	5/22/1942	15,700	5/22/1942	26,939		45,000		98,000	5/23/1942	92,000
1943			12/30/1942	6,434					12/31/1942	24,000
1944	11/9/1943	2,800	11/9/1943	9,019		23,200			11/9/1943	30,000
1945	7/19/1945	3,090	7/19/1945	6,143					7/15/1945	18,300
1946	5/28/1946	2,640	5/28/1946	5,588		13,400			5/28/1946	22,300
1947	7/8/1947	5,600	7/8/1947	9,829		20,000	7/8/1947	23,000	7/8/1947	22,300
1948	3/20/1948	1,950	3/20/1948	4,933			3/20/1948	11,100	4/15/1948	13,100
1949	12/30/1948	2,490	12/30/1948	6,753		16,400	12/30/1948	21,900	12/31/1948	26,900
1950	4/5/1950	1,870	3/28/1950	5,406			3/29/1950	12,600	3/29/1950	14,400
1951	12/4/1950	7,250	12/4/1950	13,833		31,500	12/4/1950	46,000	12/5/1950	46,900
1952	3/11/1952	2,310	7/10/1952	7,508		11,100	3/11/1952	27,300	3/12/1952	31,000
1953	12/11/1952	3,640	12/11/1952	9,829		20,900	11/22/1952	36,100	11/22/1952	41,200
1954	12/7/1953	1,200	12/7/1953	3,413			12/7/1953	11,700	12/7/1953	16,600
1955	8/19/1955	31,900	8/19/1955	53,059		43,900	8/19/1955	77,800	8/19/1955	91,300
1956	10/16/1955	4,650	10/15/1955	10,011			10/16/1955	25,600	10/16/1955	30,900
1957	4/6/1957	1,360	4/6/1957	5,306			4/6/1957	18,400	4/6/1957	23,800
1958	12/21/1957	3,550	12/21/1957	12,650		23,800	12/21/1957	37,600	12/21/1957	41,800
1959	1/22/1959	1,850	1/22/1959	4,951			3/6/1959	14,000	3/6/1959	17,300
1960	5/22/1960	3,440	4/1/1960	7,570			4/5/1960	16,400	4/5/1960	22,100
1961	2/26/1961	1,640		9,600				20,100		24,800
1962	1/7/1962	1,120		5,100				21,100		18,200
1963	3/27/1963	2,220		7,800		14,500		15,500		15,400

WY	Stoddartsville		White Haven/Tannery		Lehighton		Walnutport		Bethlehem	
	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)
1964	3/10/1964	2,270		7,700		14,300		17,100		19,300
1965	2/12/1965	495		2,200		5,670		16,000		28,400
1966	3/5/1966	560		2,800		9,760		11,100		13,100
1967	3/29/1967	1,000		3,500		9,760		12,000		11,300
1968	3/18/1968	960		3,500		10,200		16,500		17,900
1969	7/28/1969	1,660		11,500		10,430		20,500		21,600
1970	4/2/1970	1,830		7,100		12,700		28,500		32,100
1971	3/16/1971	1,180		3,700		7,640		17,500		26,300
1972	6/23/1972	3,210		10,900		24,200		49,000		65,200
1973	6/29/1973	7,700		14,000		28,100		32,700		41,500
1974	12/21/1973	2,840		11,200		8,820		26,400		32,400
1975	2/25/1975	2,150		10,100		9,090		19,800		22,900
1976	1/27/1976	2,420		11,400		18,500		28,600		42,200
1977	10/9/1976	1,990		8,800		16,700		22,700		27,700
1978	1/9/1978	2,990		11,100		25,000		31,900		36,700
1979	1/25/1979	2,470		9,300		21,200		35,500		56,800
1980	3/22/1980	2,330		10,200		8,200		21,200		21,200
1981	2/12/1981	1,350		8,610				26,000		26,000
1982	4/18/1982	1,140		5,220		9,100		13,800		20,300
1983	4/16/1983	4,190		15,385		26,800		44,800		48,000
1984	4/6/1984	3,700								
1985	9/27/1985	8,380								
1986	3/15/1986	4,120								
1987	9/13/1987	3,380								
1988	5/20/1988	1,450								
1989	5/6/1989	3,130								
1990	10/20/1989	1,040								
1991	11/11/1990	1,610	12/6/1990	4,427	12/4/1990	11,210	12/4/1990	21,876	12/4/1990	24,330

WY	Stoddartsville		White Haven/Tannery		Lehighton		Walnutport		Bethlehem	
	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)
1992	11/23/1991	991	11/23/1991	3,869	11/23/1991	7,046	6/1/1992	9,259	3/28/1992	10,793
1993	4/11/1993	3,680	4/11/1993	14,120	4/11/1993	21,411	4/11/1993	29,782	4/11/1993	35,185
1994	11/28/1993	2,420	11/29/1993	14,948	11/29/1993	17,671	11/28/1993	24,836	11/28/1993	28,702
1995	11/28/1994	1,040	1/16/1995	2,973	1/17/1995	5,522	11/29/1994	9,321	3/9/1995	11,321
1996	1/19/1996	7,190	1/20/1996	24,740	1/20/1996	29,293	1/20/1996	44,791	1/20/1996	54,010
1997	11/9/1996	3,440	11/9/1996	10,589	11/9/1996	17,820	12/2/1996	27,533	12/2/1996	33,235
1998	3/10/1998	1,350	3/10/1998	4,803	3/10/1998	7,650	3/10/1998	11,041	3/10/1998	15,552
1999	1/24/1999	2,560	9/17/1999	8,540	1/24/1999	14,512	1/24/1999	21,093	1/25/1999	24,856
2000	2/28/2000	1,170	2/28/2000	4,482	2/28/2000	8,390	2/29/2000	11,493	2/29/2000	13,764
2001	12/17/2000	1,610	12/17/2000	5,408	12/18/2000	11,013	12/18/2000	18,744	12/18/2000	22,903
2002	5/28/2002	5,440	5/29/2002	12,054	5/29/2002	13,397	5/29/2002	14,088	5/29/2002	15,233
2003	6/22/2003	2,800	6/22/2003	8,974	6/1/2003	14,147	9/23/2003	23,357	9/24/2003	30,313
2004	9/18/2004	10,200	9/18/2004	27,406	9/18/2004	44,572	9/18/2004	64,977	9/19/2004	69,528
2005	4/3/2005	6,500	4/3/2005	20,566	4/3/2005	34,757	4/3/2005	49,758	4/3/2005	61,514
2006	6/27/2006	10,500	6/28/2006	22,718	6/28/2006	44,879	6/28/2006	64,635	6/28/2006	72,903
2007	11/17/2006	2,580	11/17/2006	7,447	11/17/2006	14,127	11/17/2006	17,599	4/16/2007	19,402
2008	3/5/2008	3,410	3/5/2008	10,583	3/5/2008	19,555	3/5/2008	29,685	3/6/2008	34,276
2009	12/12/2008	2,580	12/12/2008	7,123	12/12/2008	12,809	12/12/2008	21,369	12/12/2008	26,536
2010	3/14/2010	2,620	3/14/2010	6,957	1/26/2010	11,316	3/14/2010	15,118	3/14/2010	18,516
2011	10/1/2010	10,500	10/1/2010	19,270	3/11/2011	31,677	3/11/2011	52,039	3/11/2011	58,437
2012	10/2/2011	973	11/25/2011	5,467	10/4/2011	8,365	9/19/2012	18,772	9/19/2012	17,692
2013	12/21/2012	1,510	12/22/2012	4,578	10/31/2012	8,699	12/21/2012	12,825	12/22/2012	14,527
2014	5/17/2014	979	5/17/2014	3,735	5/18/2014	7,319	5/1/2014	14,422	5/1/2014	22,180
2015	7/1/2015	711	7/1/2015	3,734	7/1/2015	11,244	7/1/2015	15,854	7/1/2015	16,124
2016	2/25/2016	1,360	2/25/2016	5,072	2/25/2016	9,951	2/25/2016	14,857	2/25/2016	21,082
2017	4/7/2017	1,790	4/7/2017	4,333	4/7/2017	8,688	4/7/2017	11,875	4/8/2017	14,751
2018	8/14/2018	3,560	7/26/2018	6,814	7/26/2018	13,845	7/26/2018	15,698	8/14/2018	27,093
2019	4/15/2019	2,250	12/22/2018	5,679	4/15/2019	11,308	1/25/2019	18,595	1/25/2019	25,966

WY	Stoddartsville		White Haven/Tannery		Lehighton		Walnutport		Bethlehem	
	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)	Date	Flow (cfs)
2020	6/11/2020	2,900	6/11/2020	7,002	6/11/2020	12,412	8/5/2020	14,452	8/4/2020	37,844
2021	9/2/2021	8,850	9/4/2021	8,000	9/2/2021	40,002	9/2/2021	61,618	9/2/2021	60,782
2022	4/8/2022	2,400	10/8/2021	4100	4/8/2022	12,421	4/8/2022	18,345	4/8/2022	21,134

APPENDIX C: DIAGNOSTIC PLOTS FOR GENERALIZED PAREO DISTRIBUTION THRESHOLD SELECTION

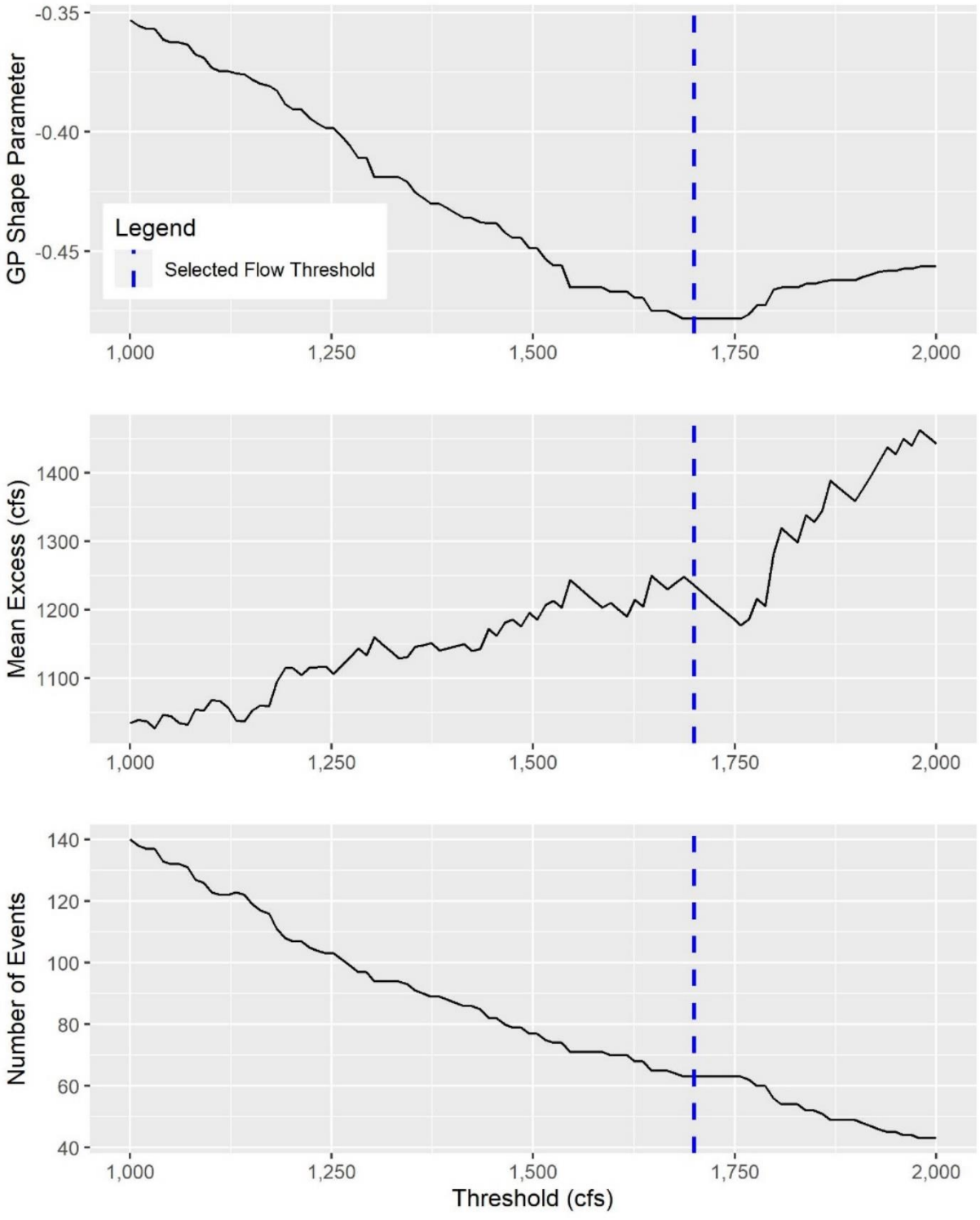


Figure C-1. Diagnostic plots for GP distribution threshold selection for Stoddartsville gage

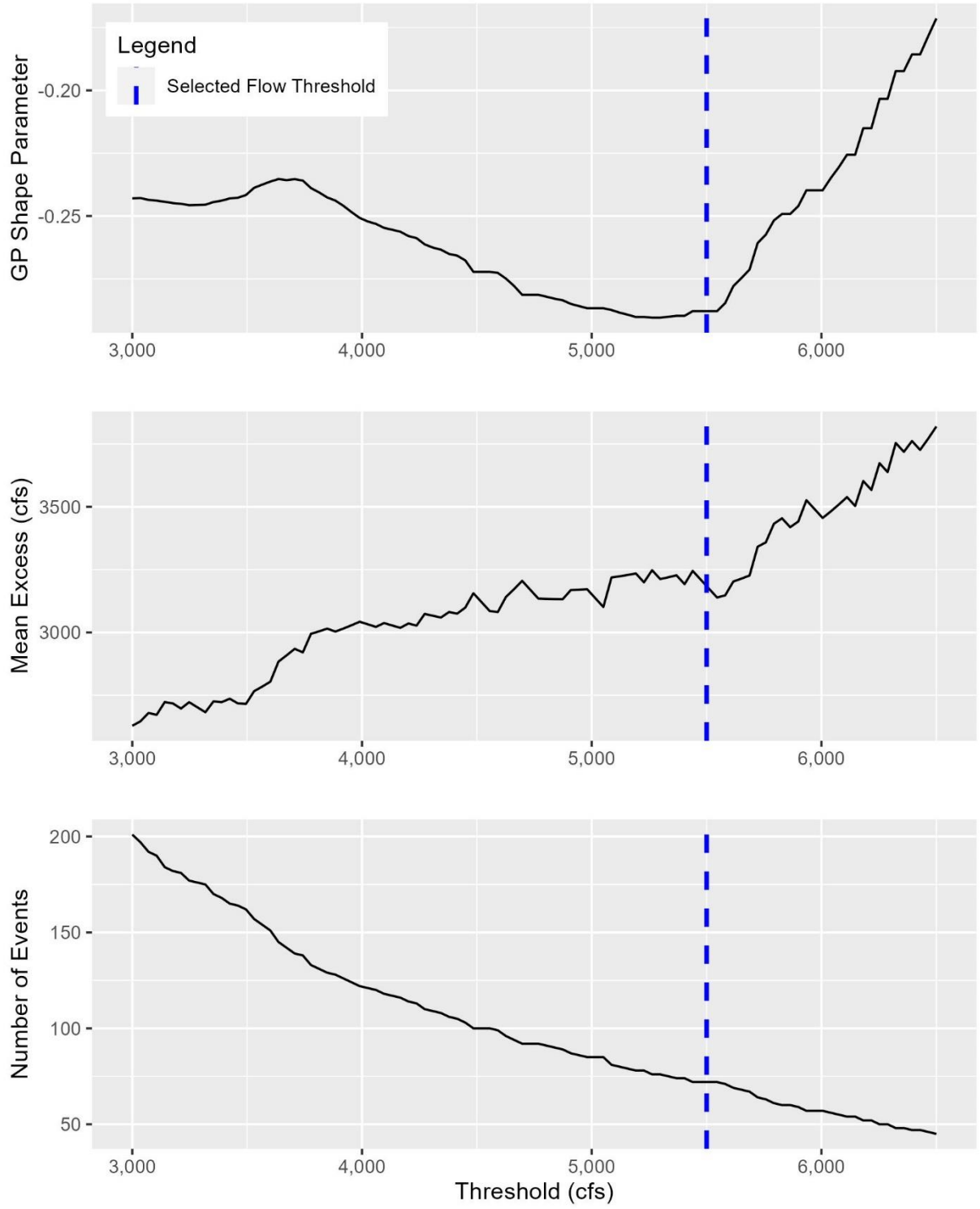


Figure C-2. Diagnostic plots for GP distribution threshold selection for White Haven/Tannery gage

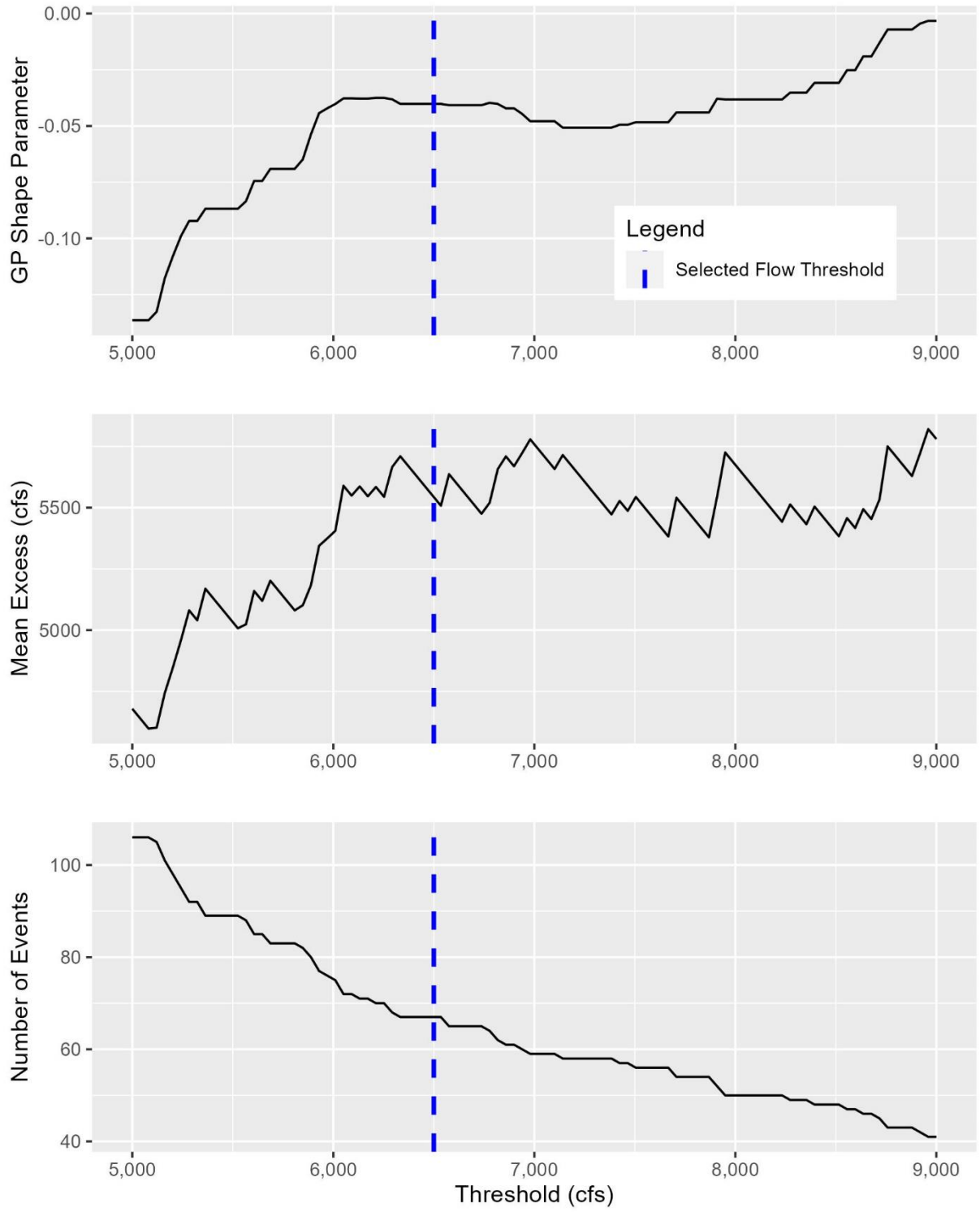


Figure C-3. Diagnostic plots for GP distribution threshold selection for Lehighton gage

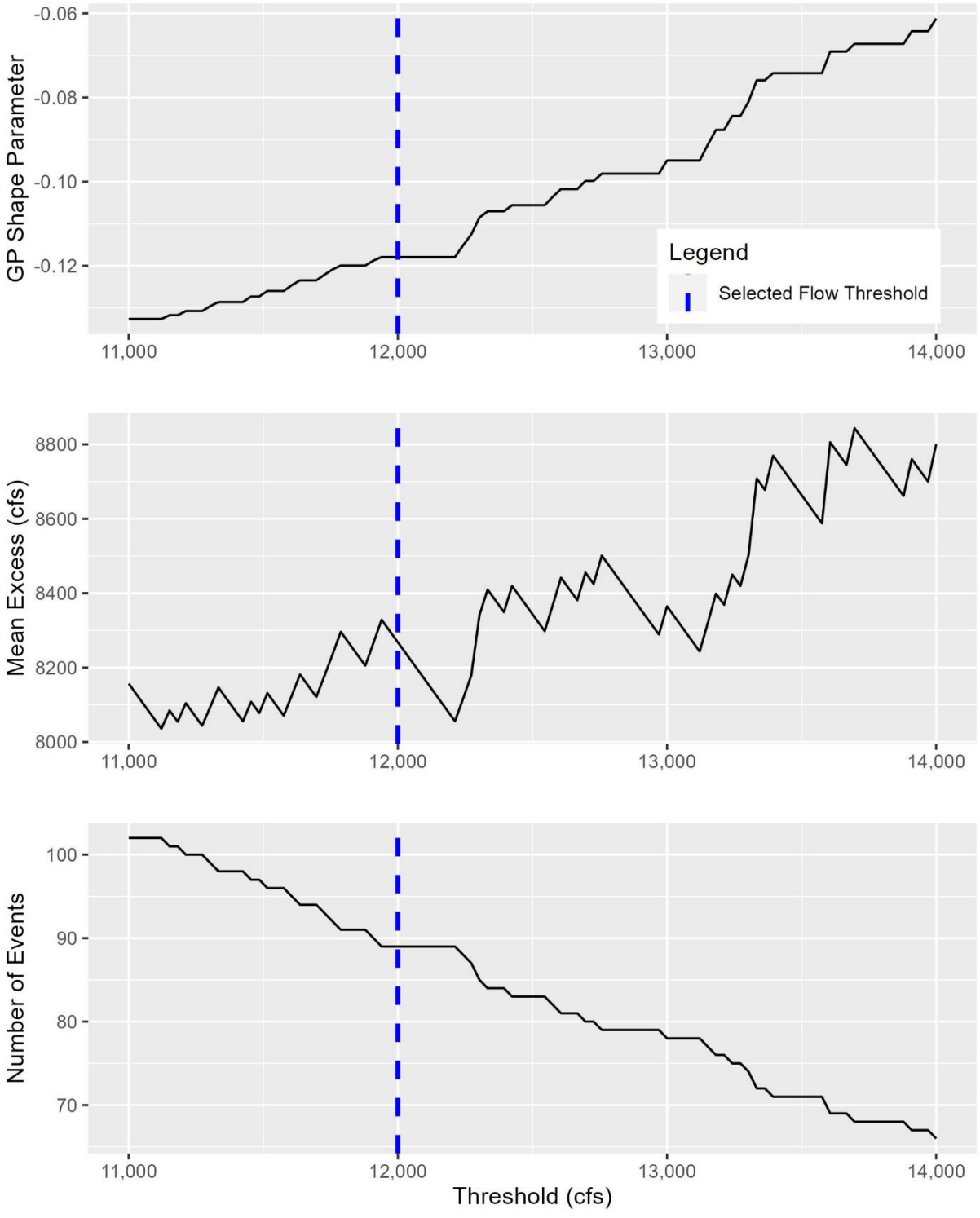


Figure C-4. Diagnostic plots for GP distribution threshold selection for Walnutport gage

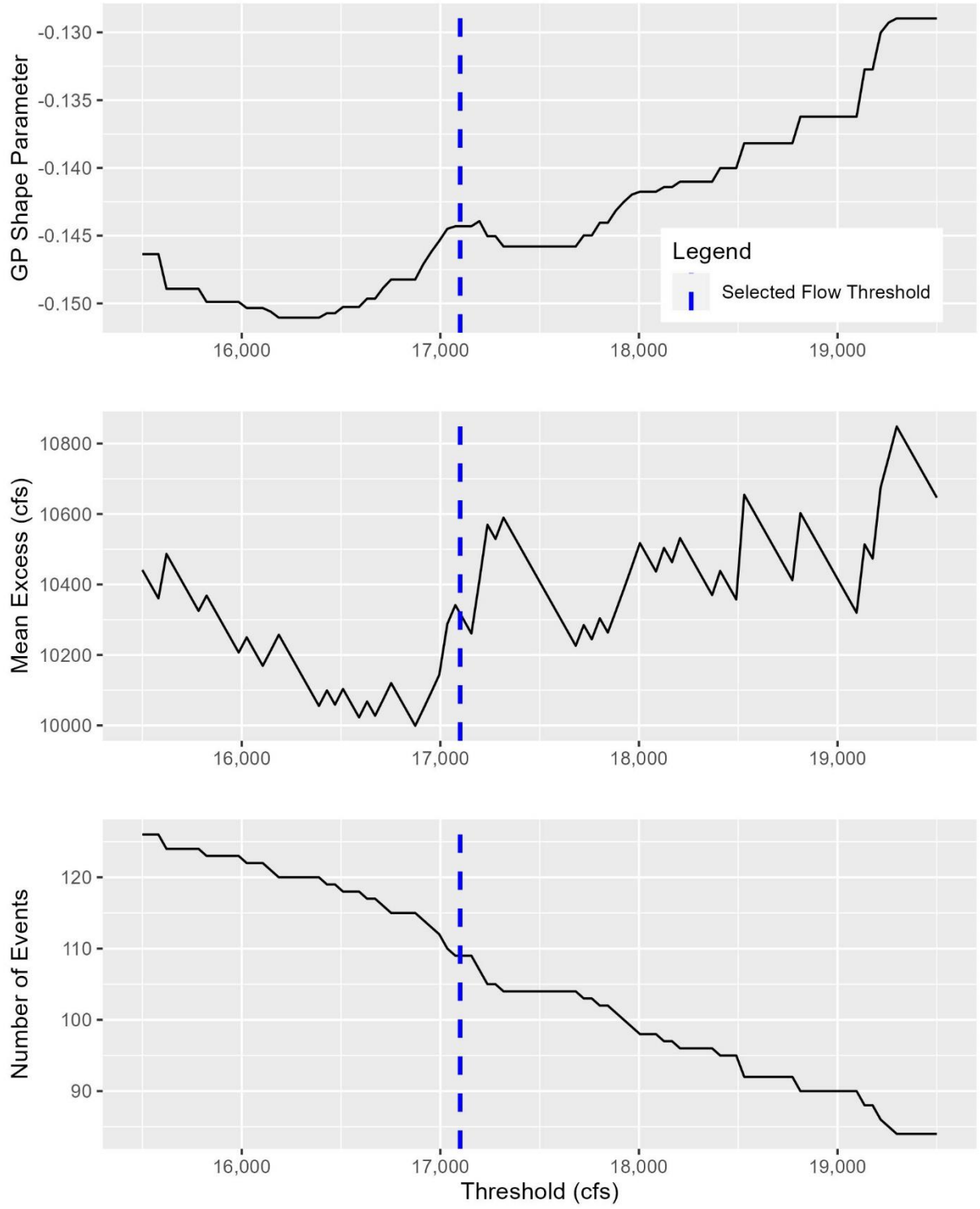


Figure C-5. Diagnostic plots for GP distribution threshold selection for Bethlehem gage

LIST OF ABBREVIATIONS

AEP	Annual exceedance probability
WY	Water year
AMS	Annual maximum series
IID	Independent and identically distributed
PDS	Partial duration series
POT	Peaks-over-threshold
USGS	United States Geological Survey
SWE	Snow water equivalent
GP	Generalized Pareto
GEV	Generalized extreme value
USACE	United States Army Corps of Engineers
MOVE	Maintenance of variance
HEC-HMS	Hydrologic Engineering Center's Hydrologic Modeling System
UA	University of Arizona
20CRV3	20th Century Reanalysis Project Version 3
SNOTEL	Snow Telemetry
COOP	Cooperative Observer Program
PRISM	Parameter-elevation Regression on Independent Slopes Model
IBTrACS	International Best Track Archive for Climate Stewardship
TC/TSR	Tropical cyclone/tropical storm remnant
ROS	Rain-on-snow
NOAA	National Oceanic and Atmospheric Administration
NCEI	National Centers for Environmental Information
RMSE	Root mean square error
RSR	Root mean square error-observed standard deviation ratio
LP-III	Log-Pearson Type III
EMA	Expected Moments Algorithm
HEC-SSP	Hydrologic Engineering Center's Statistical Software Program