The Pennsylvania State University

The Graduate School

College of Earth and Mineral Sciences

ANALYSIS OF MIXED-DISTRIBUTION STATISTICAL FLOOD FREQUENCY MODELS AND IMPLICATIONS FOR DAM SAFETY ASSESSMENTS

A Thesis in

Geosciences

by

Kenneth Joel Roop-Eckart

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2018

The thesis of Kenneth Joel Roop-Eckart was reviewed and approved* by the following:

Klaus Keller
Professor of Geosciences
Thesis Advisor

Elizabeth Hajek
Associate Professor of Geosciences

Tess Russo
Assistant Research Professor at the Earth and Environmental Systems Institute

Christopher Duffy
Professor of Civil and Environmental Engineering

Demian Saffer
Professor of Geosciences
Head of the Department of Geosciences

*Signatures are on file in the Graduate School

ABSTRACT

Water is a vital resource, but also a source of hazards. Flooding poses considerable hazards to human lives and property. Dams and levee systems are key components of modern flood defenses. However, these flood defenses can fail catastrophically. This thesis addresses mixed distributions in statistical flood frequency analysis and implications for dam safety assessments.

Previous studies in dam safety assessment have established a variety of statistical and physical modeling methods (Swain *et al*., 2006). Statistical flood frequency analysis represents a popular, low cost method (Swain *et al*., 2006). However, current flood frequency methods can neglect mixed distributions and under predict true flood risk. Here, I improve on the standard flood frequency methods (England *et al*., 2018) by:

A. implementing single and mixed distribution models to assess flood frequency analysis sensitivity to model choice and model structural uncertainty,
B. statistically test for the presence of mixed distributions in peak flow data, and
C. demonstrating the implications of accounting for mixed distribution peak flows in dam safety assessments.

I find that current methods in flood frequency analysis can lead analysts to disregard mixed distributions of peak flows. Goodness-of-fit metrics can be used to identify mixed distributions of peak flows at a location. Additionally, implementing mixed distribution statistical flood frequency analysis at mixed distribution peak flow sites can produce better fits (as judged by statistical tests) and can greatly increase predicted flood risk. These findings have potential safety implications for flood-frequency analysis based dam safety assessments.

TABLE OF CONTENTS

Appendix A

Appendix B

LIST OF FIGURES

LIST OF TABLES

PREFACE

This thesis includes a manuscript projected for submission to a peer-reviewed journal. Kenneth Joel Roop-Eckart, the candidate for Master of Science, is the first and corresponding author. The thesis advisor, Klaus Keller, who is the last author, together with Tess Russo, initiated the study. Kenneth Joel Roop-Eckart performed the analysis, drafted the paper, and designed and coded the likelihood functions. Benjamin Lee assisted in designing and coding the likelihood functions. Caitlin Spence provided research guidance, perspective, and problem framing. Tess Russo provided research guidance and early stage project framing. Klaus Keller provided logistical support, general oversight of this study, and research guidance. All authors contributed to the study design and discussed the results.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Keller research group for thoughtful discussions about this work, the Department of Geosciences for providing a supportive and encouraging learning environment, and to all those who contributed to this study.

This study was partially supported by the Department of Energy sponsored Program on Coupled Human Earth Systems (PCHES) under DOE Cooperative Agreement No. DE-SC0016162, the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507, and the Penn State Center for Climate Risk Management. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entities.

Special thanks to Benjamin Seiyon Lee for vital mentorship in statistics and programming, to Caitlin Spence for personal and professional mentorship and valuable insights into the world of hydrological engineering, to John F England, Jr. for sharing the TREX hydrographs and for insights into the field of dam safety assessments, and to Randy Miller for reviewing and testing the reproducibility.

And last but not least, I would like to thank my advisor, Klaus Keller, for spending many hours of his valuable time, both in group settings and individually, providing invaluable guidance and mentorship.

Chapter 1

Introduction: A brief review of the history of flood risk assessments

An age-old danger

Beginning at the end of the last ice age, approximately 12,000 years ago, humans began to farm along rivers (Bianchi, 2016). Roughly 5,000 years ago, cities and civilizations had grown up along rivers. These early settlers cultivated crops on the fertile floodplains and used the rivers for water, irrigation, and transportation, among other uses (Bianchi, 2016).

Humanity's relationship with floods is as old as civilization itself. Early settlements could do little to stem the floodwaters when they came. Early water-management structures, including diversion and retention dams and irrigation ditches were built for agricultural purposes. As early as 5,100 years ago, early Chinese were already building a "large-scale complex of dams, levees, ditches, and other water controlling features" (Liu *et al*., 2017). Similar water management structures were implemented in Mesopotamia and Egypt (Bazza, 2007).

In early dynastic China, approximately 4,000 years ago, concerted flood control efforts, in the form of earthen levees, began along the Yellow river (Clark, 1982). These are some of the earliest known flood control measures organized and implemented on a national scale. Flood control along the Yellow river has been a struggle with many successes and failures ever since (Chen *et al*., 2012).

An early history of risk assessment

Contemporary flood control is driven by risk assessment. Flood control infrastructure is built to specific standards based on these risk assessments. These standards could be based on loss of life (England *et al*., 2006), a standard return period (Robinson *et al*., 2004), a cost benefit analysis (Van Dantzig, 1956), or another standard. Regardless of the standard chosen, each standard depends on flood risk assessments. Risk assessment itself, however, has a long and varied history that only in the 300 years has matured into the probability theory based risk assessments used today.

The most basic form of Risk analysis can be found as early as the Asipu in ancient Mesopotamia, circa 5,200 years ago (Covello and Mumpower, 1984). The Asipu were consulted on important decisions, such as alliances or royal marriages, considered the possible outcomes, consulted the gods, and produced and risk assessment reports (Covello and Mumpower, 1984). Unlike modern risk assessment, however, The Asipu were priest-like interpreters of the gods (Covello and Mumpower, 1984). Current understanding of risk is deeply associated with probability theory, and thus did not develop about until advances in probability theory in the 1700s (Covello and Mumpower, 1984).

Discussions of probability may have begun with Plato's Phaedo, published in the 4th century B.C., and a number of works following it (Covello and Mumpower, 1984). The works focused on risk in the afterlife associated with one's actions on Earth. These works addressed the understanding of uncertainty associated with the afterlife and qualitative analysis of risk based on actions (Covello and Mumpower, 1984).

One of the first applications of expectation maximization, an important concept in modern risk analysis, was introduced by Arnobius the Elder in 4th century C.E. in North Africa (Covello and Mumpower, 1984). When Arnobius converted to Christianity, he published a 2x2 matrix argument. The choices in the matrix are “remain pagan” and “become Christian”, and the states of the world are “(the Christian) God exists” and “(the Christian) God does not exist” (Covello and Mumpower, 1984). Arnobius argued that given the states of the world, choice to become Christian posed the highest expected value for the individual (Covello and Mumpower, 1984).

Pascal in the 17th century A.D. revised Arnobius’ argument using a modern understanding of probability (Covello and Mumpower, 1984). The following century experienced an explosion of probabilistic thinking, culminating in Laplace’s prototype of modern quantitative risk analysis, an analysis of the likelihood of death by whether someone received the Smallpox vaccine or not, in 1792 (Covello and Mumpower, 1984).

The reason for this explosion of understanding and investigation in the area of probability, or the lack of knowledge prior to Pascal is not well understood (Covello and Mumpower, 1984). For this thesis it will is sufficient to understand that quantitative risk analysis in thought experiment or reality was not possible prior to the period from Pascal’s 17th century investigations to Laplace’s analysis in 1792 (Covello and Mumpower, 1984).

While probability theory forms the basis of modern quantitative risk analysis and thus modern flood frequency analysis (England *et al*., 2018; Swain *et al*., 2006), it is not the only way risk may be considered and handled in financial situations. An example is insurance in the ancient world. Thousands of years prior to the development of the modern understanding of probability,

trade in Babylon and later Greece and the Roman Empire needed insurance for loans (Covello and Mumpower, 1984). This was typically charged as an insurance premium on top of the value of the loan. These were so important Hammurabi's code of laws circa 1950 B.C. included regulations for trade insurance (Covello and Mumpower, 1984). The Romans further extended insurance to early forms of health and life insurance (Covello and Mumpower, 1984). Thus the ancient world possessed an intimate understanding of risk and developed detailed application to real world decisions despite no recorded understanding of probability theory itself.

Origins of statistical flood frequency analysis

Currently (2018), statistical flood frequency analysis is an efficient method of assessing flood risk. The method uses a statistical distribution fitted to peak flow data to create a flood frequency curve relating peak flows to return periods. The curve is then used to estimate the peak flows of uncommon (typically greater than 50 year return period) flooding events, and therefore the size of the necessary flood control infrastructure.

Physical rainfall-flood relationships were derived as early as 1851 (Rossi *et al*., 1994), however statistical analyses of flood frequency were not investigated in detail until the early 20th century. Fuller (1913), Foster (1924), Hazen (1930), Gibrat (1932), and Supino (1934) laid much of the groundwork for modern flood frequency analysis, including identifying identical and independent distribution of data and stationarity as important assumptions in flood frequency analysis (Rossi *et al*., 1994).

Early flood frequency analysis was viewed in a functional context. Distributions were chosen for their goodness-of-fit. Theoretical statistical justifications for one distribution over

another were typically not considered. Flood frequency models were chosen and used by how well they fit the data, not underlying theoretical justifications (Rossi *et al*., 1994). However, Gumbel (1941) proposes that annual peak flows resulted from a particular "parent" distribution, the Gumbel or Extreme Value distribution type 1 (EV1) distribution, which is rooted in extreme value theory as the maximum of a sufficiently large number of draws from identically and randomly distributed variables (Rossi *et al*., 1994). Later the EV1 was shown to be a special case of the more general, and aptly named, Generalized Extreme Value (GEV) distribution (McFadden, 1978).

The presence of outliers that could not be reconciled by these distributions lead to the Wakeby distribution (Houghton, 1978), Two-Component-Extreme-Value distribution (Rossi *et al*., 1984), a mixed lognormal distribution (Singh and Sinclair, 1972), and many other mixed distributions. These distributions accounted for outliers by using multiple shape parameters to be more flexible than past methods in the case of the Wakeby, or proposed that annual peak flows were in fact the result of two processes that could be modeled by a basic and outlier component (Rossi *et al*., 1994).

An incomplete history of stream gaging

Accurate peak flow records are essential for flood frequency analysis. The bulk of peak flow data for flood frequency analysis comes from stream gages. Stream gages typically operate by constructing a stream stage to discharge relationship known as a rating curve from coincident stream stage and flow measurements. The stream stage is then used to estimate stream flows. The relationship is typically updated on some interval to maintain or improve measurement accuracy.

Stream gaging has a long but sparse history. Measurements of yearly maximum stream stage are as old as the first Egyptian dynasty roughly 3,000 B.C.E. (Bell, 1970). However, the first stream gage to apply a rating curve and calculate daily discharge may have been in Basel, Switzerland, from 1809 to 1821 (Follansbee, 1919). The Basel stream gage and other early ventures in stream gaging used the slope method of calculating flow, though the Basel stream gage was supplemented with some flow measurements (Follansbee, 1919). The first daily discharge measurements by stream stage and velocity were conducted on the Ohio River in the summer and fall of 1849 (Follansbee, 1919). Finally, the first United States Geological Survey (USGS) stream gage in the United States was established in 1888 in Embudo, New Mexico, and provides the longest stream gage record in the United States (Frazier and Heckler, 1972).

Water scarcity and importance of water budgeting for irrigation uses drove stream gage development across the western United States. While the Embudo stream gage served as testing ground for USGS stream gaging techniques, the states of California and Colorado independently experimented with stream gaging (Follansbee, 1919). The United States government approved funds for the USGS to begin stream gaging operations with no location restrictions in 1894 (Frazier and Heckler, 1972). The stream gage on the Arkansas River in Pueblo, Colorado, was established the same year (Follansbee, 1919).

Development of flood control and frequency analysis in the United States

Flood control was originally a local problem in the United States (Wright, 2000). The Constitution neither authorized nor prohibited federal funding for internal improvements, including flood control infrastructure. As a result, local governments or organizations were left with the responsibility of flood control.

In the 1824 *Gibbons v. Ogden* decision, the United States Supreme Court ruled that the federal government could fund internal infrastructure projects, including flood prevention under the Commerce Clause. Almost immediately congress authorized the United States Army Corps of Engineers and appropriated funds, but river navigation projects were preferred over large flood control infrastructure projects (Wright, 2000).

The flood control debate in the United States concentrated on the Mississippi and Ohio rivers, where flooding in 1903, 1912, and 1913 in killed many people and caused considerable damages. The 1913 flood alone killed 415 people and caused roughly $200 million in property damages (Wright, 2000). The river basins were economically valuable, but the local communities did not have the resources to protect themselves from large floods (Wright, 2000). It became clear local authorities could not control the Mississippi. The Federal government became increasingly involved in local flood control with the Flood Control Acts of 1917 and 1936. It became clear local authorities could not control the Mississippi (Wright, 2000).

In 1965 the Water Resources Council was created as an independent agency with secretaries from six federal agencies. The council's purpose was to address water resource planning and coordination, including flood related issues on a national scale (Wright, 2000). In 1967, the Water Resources Council published Bulletin 15, A Uniform Technique for Determining Flood Flow Frequencies. The methods established in Bulletin 15 became the national standard for flood frequency analysis. The methods in Bulletin 15 have been updated in Bulletin 17, Bulletin 17A, Bulletin 17B, and most recently Bulletin 17C (England *et al*., 2018).

The United States Bureau of Reclamation

While flooding in the Ohio and Mississippi river valleys is mostly addressed by levees and floodplain management practices, the arid climate of the west required a different approach. Local and state governments lacked the necessary resources or skills to implement the large scale water management projects needed by local governments (United States Bureau of Reclamation, 2011).

In 1902, The United States Reclamation Service was created under the United States Geological Survey to build irrigation, water management, and flood control projects in the western United States. Reclamation Service's mission was to "reclaim" the arid and difficult to farm regions in the western United States for development (United States Bureau of Reclamation, 2011).

The arid nature of the west, where snowpack and seasonal rainfall are often the main sources of water, means that dams, rather than levees, are needed for year around water management (Untied States Bureau of Reclamation, 2011). Dams protect against flooding in the wet season, and retain water for irrigation, industrial, and private use in the dry season.

After a period of growth and learning from early difficulties, the Reclamation Service was separated into its own agency, The United States Bureau of Reclamation, under the Department of the Interior in 1923 (United States Bureau of Reclamation, 2011). The Bureau of Reclamation now oversees more than 180 water management projects in the seventeen western states, and for the most part has transitioned from construction and planning to maintenance and

management of existing projects. This includes carefully monitoring the safety of all dams under the Bureau's oversight (United States Bureau of Reclamation, 2011).

The Pueblo Dam

The Pueblo Dam is the second largest Bureau of Reclamation dam in Colorado and the largest and terminal dam in the Fryingpan-Arkansas project (Rogers, 2006). The project was first considered by congress in 1952, but was not approved until 1962. The Pueblo Dam was finished in 1975 (Rogers, 2006). The Pueblo Reservoir provides water to Pueblo, Colorado, for industry, agriculture, and municipal use, and flood control to the city (Rogers, 2006).

The Pueblo Dam was built to route the Probable Maximum Flood (PMF) without overtopping (United States Bureau of Reclamation, 1968a, 1968b). The PMF for Pueblo was revaluated in 1991 as part of the ongoing Safety of Dams Program (England *et al*., 2006). The Pueblo dam failed to safely route the PMF estimated in 1991 (Bullard and Leverson, 1991).

New dam safety investigations were required, because the dam failed the PMF test. England *et al*. (2010) investigates paleofloods and historical floods throughout the upper Arkansas watershed above Pueblo, Colorado, and hypothesized a mixture of snowmelt and rainfall floods at Pueblo that complicated statistical flood frequency analysis. England *et al*. (2014) uses stochastic storm transpositioning with a rainfall runoff model to assess the overtopping return period of the dam. England *et al*. (2014) concludes the dam was safe to the regulation return period 400,000 years (England *et al*., 2006). This method avoids the complications of flood frequency analysis with mixed distributions by modeling rainfall floods,

which are hypothesized by England *et* al. (2010) to dominate the most extreme floods. However, the method is a great deal more expensive than flood frequency analysis from a worker hour and computational perspective (Swain *et al*., 2006) and is subject to watershed and storm characteristic uncertainties (England *et al*., 2014).

References

Bianchi, Thomas S. *Deltas and Humans: A Long Relationship Now Threatened by Global Change*. New York, New York: Oxford University Press, 2016.

Bullard, KL. Leverson, V. 1991. “Pueblo Dam, Fryingpan-Arkansas Project Probable Maximum Flood (PMF) Study.” *Bureau of Reclamation*.

Bureau of Reclamation. 1968a. Supplement to Final Inflow Design Flood Study of January 16, 1967 by Region 7, Pueblo Dam site, Fryingpan - Arkansas Project, Memorandum from Donald L. Miller to Head, Flood Hydrology Section, dated January 24, 1968.

Bureau of Reclamation. 1968b. Volume Inflow Design Flood, Pueblo Dam site, Fryingpan - Arkansas Project, Memorandum from Donald L. Miller to Head, Flood Hydrology Section, dated March 18, 1968.

Clark, Champ. *Planet Earth Flood*. Alexandria, VA: Time-Life Books, 1982.

Covello, Vincent T., and Jeryl Mumpower. 1984. “Risk Analysis and Risk Management: An Historical Perspective.” *Risk Analysis* 5: 103–120. https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1539-6924.1985.tb00159.x.

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas,W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr. 2018. Guidelines for determining flood flow frequency—Bulletin 17C: United States Geological Survey Techniques and Methods, book 4, chap. B5. *United States Geological Survey.*

England, John F, Jeanne E Klawon, Ralph E Klinger, and Travis R Bauer. 2006. “Flood Hazard Study Pueblo Dam, Colorado.” *United States Bureau of Reclamation*.

Follansbee, Robert. 1919. “A History of the Water Resources Branch, U. S. Geological Survey : Volume I, From Predecessor Surveys To June 30, 1919.” *United States Geological Survey.* https://pubs.usgs.gov/msb/7000087/report.pdf.

Foster, H. Alden. 1924. "Theoretical Frequency Curves and Their Application to Engineering Problems." *Proceedings of the American Society of Civil Engineers* 49 (5): 825–55.

Frazier, A.H., and W. Heckler. 1972. “Embudo, New Mexico, Birthplace of Systematic Stream Gaging.” *United States Geological Survey*. https://pubs.usgs.gov/pp/0778/report.pdf.

Fuller, W. E. 1914. “Flood flows”. *Transactions of the American Society of Civil Engineers* 77: 564-617

Gibrat, R. 1932. "Amenagement hydroelectrique des cours d'eau. Statistique mathematique et calcul des probabilities." *Revue general de I' electricite* 32 (15, 16).

Gumbel, E. J. 1941. “The Return Period of Flood Flows.” *The Annals of Mathematical Statistics* 12 (2): 163–190. http://www.jstor.org/stable/2235766.

Hazen, A. 1930. *Flood Flows, A Study of Frequencies and Magnitudes*, John Wiley and Sons, Inc., New York.

Houghton, John C. 1978. “Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows.” *Water Resources Research* 14 (6): 0–4.

James M. Wright. 2000. “The Nation’s Responses To Flood Disasters: A Historical Account.” *Association of State Floodplain Managers*. Madison, WI.

Liu, Bin, Ningyuan Wang, Minghui Chen, Xiaohong Wu, Duowen Mo, Jianguo Liu, and Shijin Xu. 2018. “Earliest Hydraulic Enterprise in China, 5,100 Years Ago.” *Proceedings of the National Academy of Sciences* 115 (52): 13637–13642. doi:10.1073/pnas.1722309115.

McFadden, Daniel. 1978. “Modeling the Choice of Residential Location.” *Transportation Research Record* (673): 72–77. http://onlinepubs.trb.org/Onlinepubs/trr/1978/673/673-012.pdf.

United States Bureau of Reclamation. 2011. “Brief History Bureau of Reclamation.” *United States Bureau of Reclamation*. https://www.usbr.gov/history/2011NEWBRIEFHISTORY.pdf.

Robinson, Michael F., John R. Sheaffer, Richard Krimm, Francis V. Reilly, Rutherford H. Platt, Firas Makarem, Vincent Parisi, et al. 2004. “Reducing Flood Losses : Is the 1% Chance (100-Year) Flood Standard Sufficient?” *2004 Assembly of the Gilbert F White National Flood Policy Forum*: 1–145. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/16/nrcs143_009401.pdf.

Rogers, Jedediah S. 2006. “Fryingpan-Arkansas Project.” *United States Bureau of Reclamation*.

Supino, G. 1934. “Probabilita e statistica nella previsione delle portate e delle piogge”, *Bollettino del Sindacato Ingegneri di Bologna*.

Van Dantzig, David. 1956. “Economic Decision Problems for Flood Prevention.” *Econometrica* 24 (3): 276–87. http://www.jstor.org/stable/1911632.

Chapter 2

Effects of Mixed Distribution Statistical Flood Frequency Models on Dam Safety Assessments: A Case Study of the Pueblo Dam

Abstract

Statistical flood frequency analysis, coupled with hydrograph scaling, is commonly used for dam safety assessment. The results can be highly sensitive to the choice of statistical flood frequency model. Past studies typically use a single distribution model, often the Log Pearson Type III or Generalized Extreme Value distributions. Floods, however, may result from multiple physical processes such as snowmelt or intense rainstorms. Multiple processes can result in a mixed distribution of annual peak flows. Engineering design choices based on a single-distribution statistical model may hence be vulnerable to the effects of this potential structural model error. Here I analyze observations from Pueblo, Colorado, for model testing, where summer snowmelt and intense summer rainstorms are key drivers of annual peak flows. I analyze the potential implications for the annual probability of overtopping induced failure of the Pueblo Dam as a didactic example. I address the temporal and physical cause separation problems by building on previous work of fitting mixed distributions directly to mixed distribution peak flows. I first use hydrograph scaling and a flood routing model to determine the smallest flood to cause overtopping. I then analyze annual peak flows, historical floods, and paleoflood records through both single and mixed distribution statistical models to estimate overtopping flood return periods. I first identify mixed distributions of peak flows using statistical flood frequency models and robust model choice criteria. I then identify the Mixed Generalized Extreme Value distribution as the best model for mixed distribution flood frequency analysis. Finally, I show that accounting for mixed distributions can greatly increase predicted flood risk.

Introduction

People rely on dams for flood protection, and catastrophic dam failure can be devastating (Graham, 2009). Well known examples include the Johnstown Dam failure in 1977 that killed 40 people and caused millions of dollars in damages, or the Teton Dam failure in 1976 that killed 11 people and caused hundreds of millions of dollars in damages (Ellingwood *et al*., 1993). One of the primary modes of dam failure is overtopping (Foster *et al*., 2000). While this failure mode is most common in earthen embankment dams, it is a major cause of failure in all dam designs (Foster *et al*., 2000). The United States Bureau of Reclamation, uses seven methods for determining dam overtopping probability based on the available data and the safety needs of the dam (Swain *et al*., 2006). The probable maximum flood (PMF) is the initial dam safety assessment performed by the United States Bureau of Reclamation (England *et al*., 2011; Swain *et al*., 2006). The PMF represents the worst-case runoff scenario that can be reasonably expected to occur given the Probable Maximum Precipitation (PMP) (Swain *et al*., 2006). The PMP is the maximum precipitation event that can be reasonably expected to occur given current understanding of meteorological factors, and is derived from a meteorological assessment of the region (Bullard and Leverson, 1991). If a dam can route the PMF without overtopping, it is considered safe in all reasonably possible flood scenarios (England *et al*., 2011; Swain *et al*., 2006).

If the dam cannot route the PMF without overtopping, the dam cannot handle the largest reasonably expected flood, and methods to determine the overtopping return period are employed (Swain *et al*., 2006). A relatively quick and low cost approach is statistical flood frequency analysis (Swain *et al*., 2006). This approach fits a statistical distribution to a dataset of annual

peak flows, historical floods, and paleoflood bounds to create a flood frequency curve for peak flows (England *et al*., 2011; Swain *et al*., 2006).

Statistical flood frequency analysis creates flood frequency curves. Flood frequency curves are made for extrapolating to return periods exceeding the time spanned by the observational record. However, dam safety depends on both the peak flow and the volume delivered over time. Hydrologic hazard curves relate peak flow and volume for specified durations to return periods to inform dam-safety assessments (Swain *et al*., 2006). To produce hydrologic hazard curves for dam safety assessments, peak flows are used to scale representative flood hydrographs (Swain *et al*., 2006). The scaled hydrographs are then routed through the reservoir using a hydrograph routing model to determine peak reservoir elevation in a method called hydrograph scaling (Swain *et al*., 2006). The combination of statistical flood frequency analysis and hydrograph scaling represents a reasonably fast and efficient method to determine the overtopping return period, and thus the safety of dams (Swain *et al*., 2006).

Statistical flood frequency analysis typically operates under the assumption that annual peak flows, historical floods, and paleoflood bounds are derived from a single distribution (England *et al*., 2018). However, in some cases, a mixed distribution of peak flows can occur (England *et al*., 2018, 2010; Rossi *et al*., 1984).

There are several procedures to identify mixed distributions in flood frequency analysis (England *et al*., 2018). The current state-of-the art in the United States requires prior knowledge about the specific causes of each observed peak flow (England *et al*., 2018). The data is separated based on physical cause, and a distribution is fitted to each individual distribution (England *et al*., 2018). Then the two distributions are re-combined into a composite distribution (e.g., Jarrett and

Costa, 1988). However, in some cases this prior knowledge of physical cause may not be available due to sparse data records.

In the current guideline for statistical flood frequency analysis in the United States, Bulletin 17C, England *et al*. (2018) identifies the need for further work on the identification and treatment of mixed distribution flood flows in flood frequency analysis. This study addresses both identification and treatment of mixed distributions at a study area where previous methods of separation by physical cause are not adequate due to sparse data. The study uses mixed distributions and robust goodness-of-fit criteria to statistically establish the presence of a mixed distribution and perform flood frequency analysis of it.

This study demonstrates the method for the case of the Pueblo Dam at Pueblo, Colorado. The dam's calculated safety has come under scrutiny since its construction in the 1970s (England *et al*., 2014, 2010; Bullard and Leverson, 1991). The dam was designed for the original PMF (United States Bureau of Reclamation, 1968a, 1968b). However, a revised PMF, calculated with updated watershed and extreme storm characteristics, would overtop the dam (Bullard and Leverson, 1991). The United States Bureau of Reclamation requires all its infrastructure to be safe to one life lost per 1,000 years of service (England *et al*., 2006). A loss of life study conducted on the Pueblo Dam and determined that between 131 and 376 people would die from catastrophic failure due to overtopping depending on whether failure occurred at night or during the day (England *et al*., 2006). As a result, the overtopping return period for the Pueblo Dam must be greater than 131,000 to 376,000 years to meet current standards. Safety analyses of the dam typically round this return period to 400,000 years (England *et al*., 2014, 2006).

A flood frequency assessment examined the return periods of peak design inflows and outflows (England *et al*., 2010). England *et al*. (2010) hypothesizes the presence of a mixed distribution of annual peak flows due to both snowmelt and rainfall peak flow distributions. However, England *et al*., (2010) is silent on the effects a mixed distribution may have on the flood frequency analysis.

Statistical flood frequency analysis at the Pueblo Dam is complicated by the occurrence of both hypothesized distributions at the same time of year (England *et al*., 2010). Additionally, many of the peak flows are not explicitly attributed to a particular physical mechanism i.e. rainfall or snowmelt (England *et al*., 2010). Current methods require that mixed distributions be separable by individual cause, or by mechanisms that are separable by season (England *et al*., 2018).

In contrast to statistical flood frequency and PMF methods, England *et al*. (2014) uses stochastic storm transpositioning and the Two-dimensional Runoff, Erosion and Export (TREX) model. The study aims to represent a physically realistic watershed response to extreme rainfall storms in order to determine the safety of the Pueblo Dam from a physically-based perspective. England *et al*. (2014) concludes the Pueblo Dam overtopping return period meets Bureau of Reclamation safety standards. Physically-based rainfall-runoff models and stochastic storm transpositioning represent a more time consuming and computationally demanding method of safety assessment (England *et al*., 2014, 2011; Swain *et al*., 2006) and is subject to uncertainties in watershed and storm characteristics (England *et al*., 2014).

In this study, I assess the effectiveness of using mixed distribution statistical flood frequency models to model the hypothesized mixed distribution peak flows, and the ability of

these models to identify hypothesized mixed distributions by statistical goodness-of-fit criteria. In doing so, I addresses three main research questions:

1. **Is a statistically identifiable mixed distribution present at Pueblo?**
2. **Which distribution best fits the data?**
3. **How would a mixed distribution affect assessments of dam safety?**

This study builds on prior work in likelihood functions (Stedinger and Cohn, 1984; O'Connell *et al*., 2002) and mixed distributions (Rossi *et al*., 1984; Raynal-Villasenor, 2012). I focus on model uncertainty and identification of mixed distributions. I then apply the methods to a didactic safety assessment of the Pueblo Dam. The didactic safety assessment illustrates the importance of considering physical motivations for statistical models and accounting for model uncertainty.

Methods

England *et al*. (2010) hypothesizes annual peak flows and Pueblo, Colorado, consist of two distributions caused by two physical processes (Fig. 2.1, Fig. 2.2). The smaller floods, with discharge less than 283 m^3s^{-1} are predominantly due to summer snowmelt floods, while the larger floods, with discharge more than 283 m^3s^{-1} were predominantly due to summer rainstorms (Fig. 2.1, Fig. 2.2) (England *et al*., 2010). Based on the prior work, I formulate two main hypotheses:

(1) A mixed distribution flood frequency model fits the annual peak flow, historical flood, and paleoflood data better than current single distribution models, as measured by Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC).

(2) Accounting for the mixed distribution of peak flows at Pueblo with a mixed distribution model will decrease estimated dam safety by predicting larger rare (131,000 to 376,000 year return period) floods.

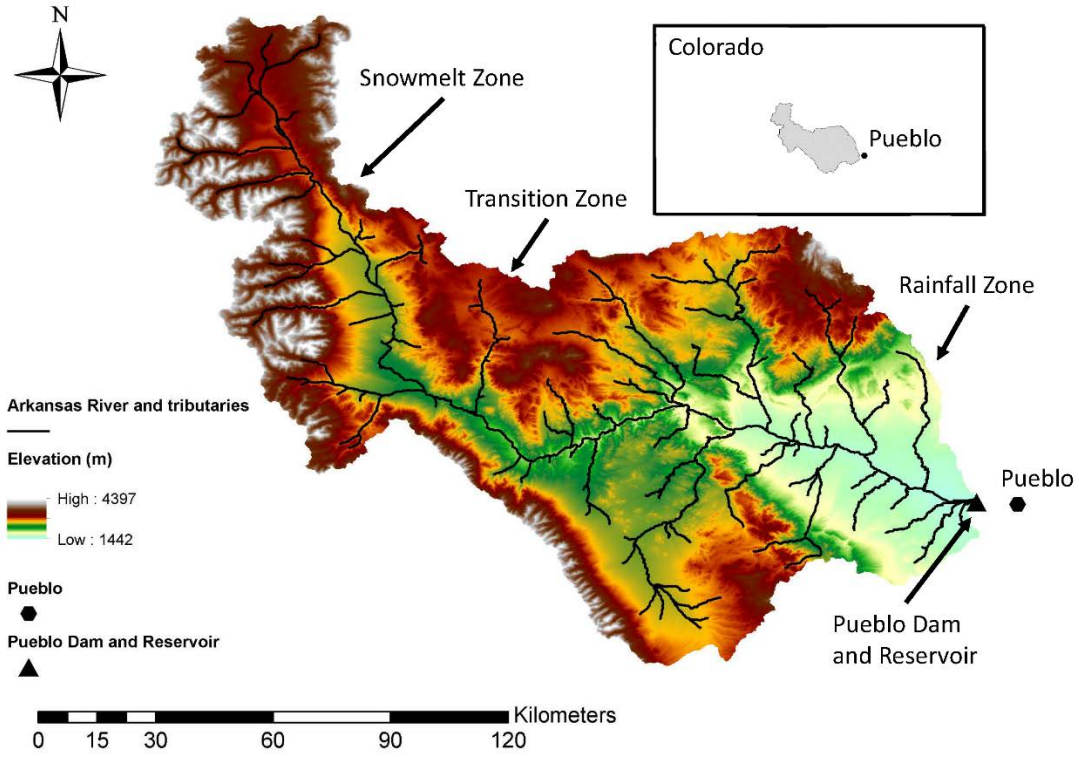

Figure 2.1: Map of the upper Arkansas River watershed, upstream of Pueblo, CO. The watershed is divided into three zones by the physical mechanism of the largest floods in the zone. In the mountain headwaters, summer snowmelt is the dominant cause of flooding. At transitional elevations summer rainstorms and snowmelt can both contribute to the largest floods. Where the narrow valley opens into the plains, snowmelt floods may still be observed, but the largest floods are rainfall driven (England *et al.*, 2014).

I collect published stream gage, historical flood, and paleoflood data at Pueblo, Colorado. The data at Pueblo, Colorado, (Fig. 2.2) consists of an 81 year daily and annual peak flow gage record, three historical floods, and one paleoflood bound. The 81 year gage record is from USGS

gage 07099500 at Pueblo, with records from 1895 to 1975 (USGS, 2018). The discharges for the three historical floods in 1864, 1893, and 1894 are calculated and reported in Campbell (1922), Follansbee and Jones (1922), Baker and Hafen (1927), Hafen (1948), and England *et al*. (2010) based on historical records of peak flood elevations. The paleoflood bound is calculated in England *et al*. (2010) using the HEC-RAS hydraulic flow model on seven channel cross sections.

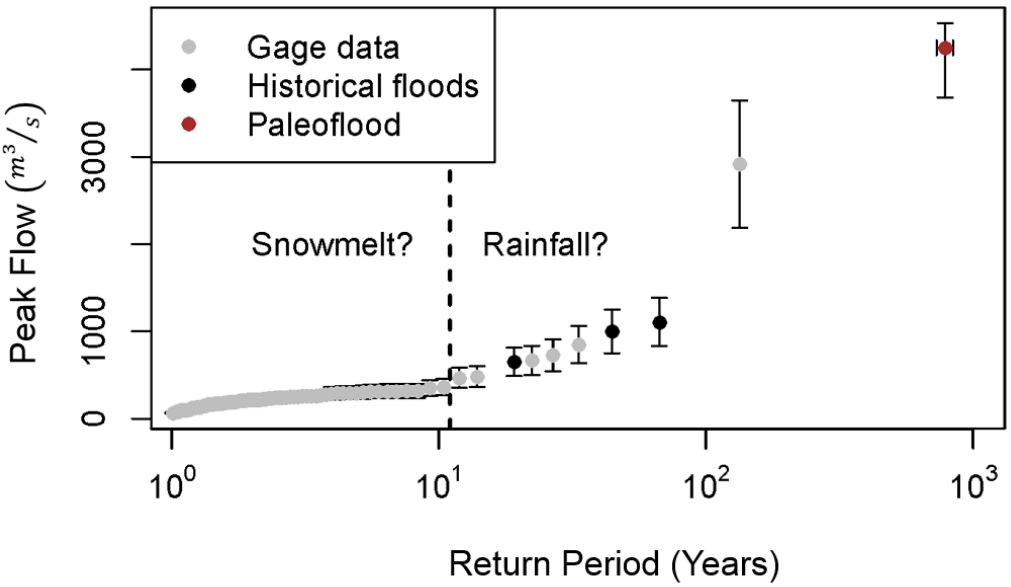

Figure 2.2: Return period plot of annual peak flows (USGS, 2018), historical floods (Campbell, 1922; Follansbee and Jones, 1922; Baker and Hafen, 1927; Hafen 1948; England *et al*., 2010), and the paleoflood (England *et al*., 2010) at Pueblo, Colorado. The hypothesized roughly 283 m^3s^{-1} transition from snowmelt to rainfall dominated flooding (England *et al*., 2010), corresponds to the roughly 11 year flood at Pueblo. The error bars represent the 95% confidence interval for flood magnitude and timing observations, estimated from generalized measurement error estimates in O'Connell *et al*. (2002).

I was unable to find complete uncertainty estimates for the USGS stream gage annual peak flows and for the historical floods. Additionally, while the paleoflood bound is estimated with upper and lower bounds (England *et al*., 2010), a probability distribution is not assigned to the estimated error. I estimate uncertainties for gaged annual peak flows and historical floods with uncertainty estimates for western United States stream gages and historical floods (O'Connell *et al*., 2002). Specifically, I assume normally distributed errors for annual peak flows and historical floods and assign a triangular probability distribution to the paleoflood bound based on the upper and lower bound discharge estimates and paleoflood age bound uncertainty distributions from O'Connell *et al*. (2002).

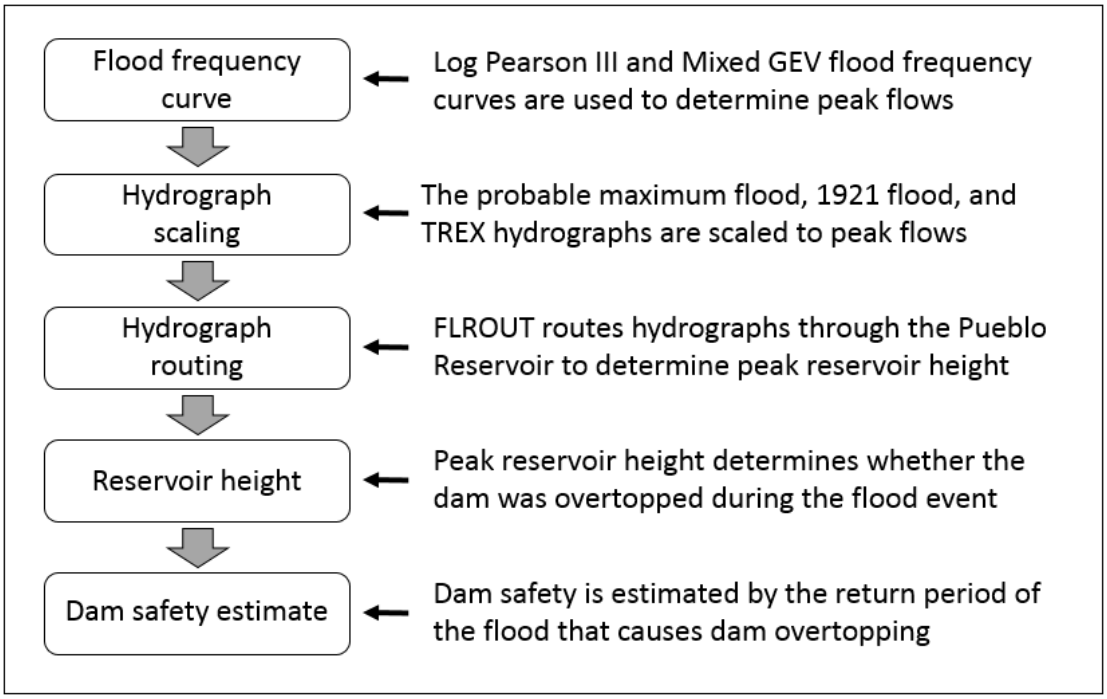

Figure 2.3: Flow chart illustrating the overall workflow of the methods in this analysis.

I use maximum likelihood estimates to fit the statistical flood frequency models to the annual peak flow, historical flood, and paleoflood data at Pueblo. I use adaptations of the standard

likelihood formula (1) to account for differences in timescales of annual, historical, and paleo observations (2) (Stedinger and Cohn, 1986), and for peak flow observation errors (3, 4) (O'Connell *et al*., 2002). Additionally, one can account for observation errors in the age of paleoflood events (5), however this method was not applied in this thesis (see caveats) (O'Connell *et al*., 2002).

The likelihood function for continuous annual peak flow data from the gage record is

$$L(\theta|x) = \prod f(x_i), \quad (1)$$

where L() is the likelihood, θ is the parameter set of the distribution, x is the annual peak flow magnitudes, and f() is the probability density function (pdf).

The likelihood function for censored data (i.e. historical or paleo floods) incorporates a binomial distribution to account for threshold censoring and normalizes to the total probability mass of the flood frequency distribution above the peak flow threshold value. The likelihood function for censored data is then

$$L(\theta|y) = \left\{\binom{h}{k} F(X_0)^{h-k}[1 - F(X_0)]^k\right\} \prod f(y_i)/[1 - F(X_0)], \quad (2)$$

where F() is the cumulative density function (cdf), f() is the pdf, X_0 is the peak flow threshold exceedance value, h is the record in years, k is the number of threshold exceedances, and y are the values of the threshold exceedances (i.e. recorded historical or paleo floods).

The likelihood functions are also adapted to incorporate peak flow measurement uncertainty based on the methods by O'Connell *et al*. (2002). The likelihood function for annual peak flow gage data with uncertainty calculates the probability density for a number of possible values of each data point and multiplies the probability density by the values' probability of being

the true value given an estimated distribution of the peak flow measurement error. The likelihood function is then

$$L(\theta|x) = \prod_{i=1}^{s}\left[\sum_j f_{ij}^{s} f(x_{ij})\right] \quad (3)$$

where L() is the likelihood, θ is the parameter set of the distribution, x is the annual maximum data, s is the length of the data, x_{ij} is a data point x_i with an error from the Gaussian error distribution, and f^s_{ij} is a discrete pdf of x_{ij} given the error distribution, i is the ith annual peak flow in the gage record, and j is the jth draw from the measurement error.

The likelihood function for historical floods and paleofloods follows the same formula as (3). Thus, the likelihood function is

$$L(\theta|y) = \prod_{i=1}^{v}\left[\sum_j f_{ij}^{y} L(\theta|y_{ij})\right], \quad (4)$$

where y is the historical flood or paleoflood data, v is the length of the data, y_{ij} is a data point y_i plus an error from the Gaussian error distribution, and f^s_{ij} is a discrete pdf of y_{ij} given the error distribution. Substitute (4) for the likelihood in (2),$\prod f(y_i)$, to account for both threshold exceedance and measurement error in historical flood and paleoflood data.

Paleofloods contend with age uncertainty in addition to peak flow uncertainty. Paleoflood age and peak flow magnitude uncertainty may be accounted for simultaneously using the log likelihood function

$$\ln[L(\theta|T)] = \sum_{i=1}^{t}\left(\sum_j f_{ij}^{tn} n_{ij}\right) \ln\left[\sum_k f_{ik}^{td} \int_{X_{min}}^{tik} L(\theta|y)\right], \quad (5)$$

where T is the paleo event age data, t is the number of paleoflood event observations, f^{tn}_{ij} is the discrete probability of an age n_{ij}, n_{ij} is the age of a paleoflood event n_i plus an error from an error distribution, f^{td}_{ik} is the discrete pdf of the paleoflood bounds, t^{ik} are the ranges of the upper y limits for the paleoflood bounds, and y_{min} is the threshold exceedance value that dictates the

minimum observable paleoflood bound (O'Connell *et al*., 2002). However, errors in paleoflood age are considered secondary for this analysis, and neglected for computational reasons (See caveats).

I consider three single distributions common in flood frequency analysis (Table 2.1): The log normal (LN2), the Log Pearson III (LP3), and the Generalized Extreme Value (GEV) distributions. The LN2 has a long history in flood frequency analysis (Rao and Hamed, 2000). The LP3 is the current government standard in the United States for flood frequency analysis (England *et al*., 2018), while the GEV is popular elsewhere (COST, 2013).

The probability density function of the LN2 distribution is given by

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}, \quad (6)$$

where f() is the probability density function, x is the data, μ is the distribution log mean, and σ is the log standard deviation.

The probability density function of the LP3 distribution from Bulletin 17C (England *et al*., 2018) is given by

$$f(x|\tau,\alpha,\beta) = \frac{\left(\frac{x-\tau}{\beta}\right)^{\alpha-1} e^{\left(-\frac{x-\tau}{\beta}\right)}}{|\beta|\Gamma(\alpha)}, \quad (7)$$

where f() is the probability density function, x is the data, τ is the location parameter, and α is the shape parameter, and β is the scale parameter.

The probability density function of the GEV distribution is given by

$$f(x|\mu,\sigma,\zeta) = \left[\frac{1}{\sigma} t(x)^{\zeta+1} e^{-t(x)}\right], \quad (8)$$

when t() is

$$t\,(x) = \left(1 + \zeta\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\zeta}} \qquad \text{if } \zeta \neq 0$$

$$t\,(x) = e^{-\frac{x-\mu}{\sigma}} \qquad \text{if } \zeta = 0\ ,$$

where f() is the probability density function, x is the data, μ is the location parameter, σ is the scale parameter, and ζ is the shape parameter. I use the GEV equations as implemented in the fExtremes R package (Wuertz, 2013).

In addition to the single distributions, I consider three mixed distributions, the Two-Component Extreme Value, the Mixed (Two-Population) Generalized Extreme Value, and the Mixed Log Pearson III distributions (Table 2.1). The Two-Component Extreme Value distribution, derived from a compound Poisson process, was popularized for flood frequency analysis when Rossi *et al*. (1984) applied it to Italian annual peak flows that struggled with outliers. Rossi *et al*. (1984) assumes that these outliers are the product of a second, upper flood distribution.

The probability density function of the Two-Component Extreme Value distribution is given by

$$f\left(x|\Lambda_1,\ \theta_1, \Lambda_2, \theta_2\right) = e^{\left(-\Lambda_1 e^{-x/\theta_1} - \Lambda_2 e^{-x/\theta_2}\right)}\left(\Lambda_1/\theta_1\, e^{-x/\theta_1} + \Lambda_2/\theta_2\, e^{-x/\theta_2}\right), \qquad (9)$$

where f() is the probability density function, x is a mixture of two independent and identically distributed sets of data, Λ_1 and Λ_2 are the relative contributions of the two components, and θ_1 and θ_2 are the exponential random variables of the components.

The Mixed Generalized Extreme Value (Mixed GEV) distribution, also known as the Two-Population Generalized Extreme Value distribution was used by Raynal-Villasenor (2012)

to model annual peak flows in Mexico. The probability density function of the Mixed GEV is given by

$$f(x|\mu_1,\sigma_1,\zeta_1,\mu_2,\sigma_2,\zeta_2,\alpha) = \alpha\left[\frac{1}{\sigma}t_1(x)^{\zeta_1+1}e^{-t_1(x)}\right] + (1-\alpha)\left[\frac{1}{\sigma}t_2(x)^{\zeta_2+1}e^{-t_2(x)}\right] \qquad (10)$$

and t(x) is

$$t_i(x) = \left(1+\zeta_i\left(\frac{x-\mu_i}{\sigma_i}\right)\right)^{\frac{-1}{\zeta_i}} \qquad \text{if } \zeta_i \neq 0$$

$$t_i(x) = e^{-\frac{x-\mu_i}{\sigma_i}} \qquad \text{if } \zeta_i = 0,$$

where f() is the pdf, i is the distribution (i = 1, 2), x is a mixture of two independent and identically distributed sets of data, μ_1 and μ_2 are the location parameters for each GEV distribution, σ_1 and σ_2 are the scale parameters for each GEV distribution, ζ_1 and ζ_2 are the shape parameters for each distribution, and α is the relative contribution of the first distribution as a fraction of the whole distribution.

The Mixed Log Pearson III (Mixed LP3) follows the same process as the Mixed GEV. It consists of two LP3 distributions added together with a weighting parameter. This distribution is not unlike the Mixed GEV, but to our knowledge has not been applied in flood frequency analysis before this study.

$$f(x|\tau_1,\alpha_1,\beta_1,\tau_2,\alpha_2,\beta_2,\alpha) =$$

$$\alpha\left[\frac{\left(\frac{x-\tau_1}{\beta_1}\right)^{\alpha_1-1}e^{\left(-\frac{x-\tau_1}{\beta_1}\right)}}{|\beta_1|\Gamma(\alpha_1)}\right] + (1-\alpha)\left[\frac{\left(\frac{x-\tau_2}{\beta_2}\right)^{\alpha_2-1}e^{\left(-\frac{x-\tau_2}{\beta_2}\right)}}{|\beta_2|\Gamma(\alpha_2)}\right], \qquad (11)$$

where f() is the probability density function, x is a mixture of independent and identically distributed sets of data, τ_1 and τ_2 are the location parameters for each LP3 distribution, α_1 and α_2 are the shape parameters for each LP3 distribution, β_1 and β_2 are the scale parameters for each

distribution, and α is the relative contribution of the first distribution as a fraction of the whole distribution.

Table 2.1 Statistical model types considered in this study, numbers of parameters, and goodness-of-fit criteria. The table shows the number of parameters in each model, whether or not it is a mixed distribution model, the goodness-of-fits, and the names of the model. The models are compared with two goodness-of-fit criteria, the Bayesian Information Criterion (BIC), and the Akaike Information Criterion (AIC). Both goodness-of-fit criteria are designed to protect against overfitting, however the BIC penalizes over parameterization more heavily than the AIC. The mixed generalized extreme value distribution produces the best fit based on both goodness-of-fit criteria, while the Log Pearson III is the 2nd best fit by the BIC and the 5th best fit by the AIC.

Type	Parameters	BIC	AIC	Model
Single	2	1091.850	1086.965	Log Normal (LN2)
Single	3	1082.565	1075.237	Log Pearson III (LP3)
Single	3	1080.856	1072.528	Generalized Extreme Value (GEV)
Mixed	4	1066.097	1056.326	Two Component Extreme Value (TCEV)
Mixed	7	1086.054	1068.955	Mixed LP3
Mixed	7	1046.954	1029.856	Mixed GEV

I produce Maximum Likelihood Estimates (MLEs) for each statistical flood frequency distribution using the DEoptim package in R (Table 2.1) (Price, 2006). These MLEs use the annual peak flow, historical flood, and paleoflood bound measurements for Pueblo, Colorado, and account for estimated measurement uncertainty for each data type.

Mixed LP3 and Mixed GEV distributions have complex likelihood spaces due to strong parameter correlations. I compare two independent (starting from a different random seed) MLEs

for each distribution to assess the convergence of the algorithm to the global maximum. For the MLP3, the two MLEs differ by 0% at seven significant figures. For the MGEV, the two MLEs differ by 0.7002%. I calculate two more independent MLEs, show that predictions of the 100 and 1,000 year return period floods do not differ largely (Supplementary Fig. S3), and use the highest log likelihood MLE of the four estimates. I calculate the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for each model fit to determine goodness-of-fit for each distribution (Table 2.1). These goodness-of-fit criteria inform the assessment whether the model choice is subject to overfitting by penalizing models for both higher numbers of parameters and for lower likelihoods.

Dam safety depends on both the ability to attenuate flood peaks through storing large volumes of water, and to pass high peak flows through the emergency spillway. As a result, flood frequency analysis is combined with hydrograph scaling to produce dam safety assessments (Fig. 2.3). Hydrograph scaling creates a flood time series by scaling a representative flood hydrograph to a peak flow (Appendix B). The method creates floods of constant durations, and volumes that scale with the peak flow. The scaled flood hydrographs are then routed through the reservoir to determine reservoir elevation. The return period of the peak flow that causes the reservoir to overtop is then calculated from the flood frequency curve. This return period is the dam safety for overtopping failure (Fig. 2.3). While a number of flood, earthquake, and static failure scenarios may be considered (England *et al*., 2006). For the simplicity of this didactic analysis, I consider only dam overtopping.

Uncertainty in the shape of the flood hydrograph is important to consider when performing hydrograph scaling. The duration of the flood and the volume delivered over the flood duration determines the flood peak attenuation capacity of the reservoir, and the magnitude of the

peak flow through the emergency spillway. I account for hydrograph uncertainty at Pueblo, Colorado, by using three potential extreme flood hydrographs drawn from different sources of data. The flood hydrographs each last three days, but differ in shape, resulting in different total flood volumes for the same peak flow (Fig. 2.4).

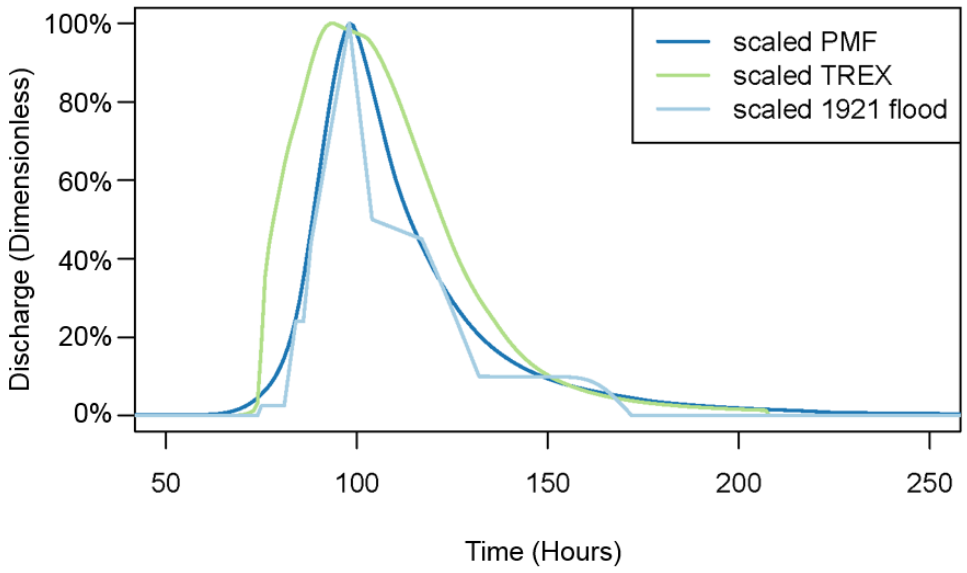

Figure 2.4: Differing shapes of each considered flood hydrograph as a function of time. Each possible flood hydrograph was scaled to the same reference peak flow.

The 1921 "great flood" hydrograph (England *et al*., 2006) represents the highest peak flow on record. This hydrograph has the lowest volume to peak flow ratio of the three hydrographs (Fig. 2.4, Supplementary Fig. S4), meaning that for the same peak flow, the reservoir must accommodate a smaller volume of water. The TREX hydrograph is the direct output of the TREX model from stochastic storm transpositioning, and represents the physically modeled reaction of the system to extreme rainfall (England *et al*., 2014). The TREX hydrograph has the largest volume to peak flow ratio of the three hydrographs, and thus delivers the largest

volume for the same peak flow (Supplementary Fig. S4). The PMF hydrograph is a regionally generalized hydrograph meant to be generally representative across multiple watersheds (Bullard and Leverson, 1991). The volume to peak flow ratio is between the 1921 and TREX flood hydrographs and represents the middle estimate.

I route the three possible flood hydrographs through the Pueblo Reservoir to account for hydrograph uncertainty in the dam safety assessment. The hydrographs are routed with the United States Bureau of Reclamation FLROUT flood routing software (England *et al*., 2006) (Appendix B). Hydrograph scaling and routing determined the peak flows required for each flood hydrograph to overtop the Pueblo Dam (Fig. 2.3). I calculate the return periods of the overtopping peak flows for each statistical flood frequency model using the associated flood frequency curve. The return period of the overtopping peak flow determines the return period of dam overtopping and thus the safety of the dam (Fig. 2.3). Accounting for hydrograph uncertainty changes what would be a line of best estimate reservoir elevation for each flood frequency distribution into a cone of uncertainty (Fig. 2.6).

I classify the dam as “meets regulation”, “uncertain”, or “does not meet regulation” for each distribution based on the return periods of the overtopping floods (Fig. 2.3; Fig. 2.5; Fig. 2.6). England *et al*. (2006) estimates between 131 and 376 deaths from a catastrophic failure of the Pueblo Dam due to overtopping. Thus the “uncertain” safety return period is between 131,000 and 376,000 years, and the “meets regulations” return period beyond 376,000 years, while the “does not meet regulations” return period is less than 131,000 years. England *et al.*’s (2014) assessment of the safety rounds the “meets regulation” return period up to 400,000 years, corresponding to a 1 in 400,000 yearly failure chance. In this study, however, I consider the full range of uncertainty in regulations, from 131,000 to 376,000 year return periods.

Results

The mixed distribution statistical flood frequency models performed better by the Akaike Information Criteria (AIC). The results are more mixed for the Bayesian Information Criteria (BIC). The Mixed Generalized Extreme Value (Mixed GEV) distribution, however, performed best by both goodness-of-fit criteria (Table 2.1).

Both goodness-of-fit criteria are designed to protect against overfitting, however the Bayesian Information Criteria (BIC) penalizes over parameterization more heavily than the Akaike Information Criteria (AIC). The Mixed Generalized Extreme Value (Mixed GEV) distribution produces the best fit based on both goodness-of-fit criteria, while the Log Pearson III is the 5th best fit by the BIC and the 3rd best fit by the AIC (Table 2.1).

The Mixed GEV presents the best fit for the data based on the goodness-of-fit criteria. Our findings are hence consistent with our first hypothesis that a mixed distribution better represents the peak flow data at Pueblo, Colorado (Table 2.1).

The Mixed GEV model predicts larger greater than 112 year return period events than the current LP3 approach (England *et al*., 2010) with the greatest differences at the longer return periods. The increase in predicted greater than 112 year return period peak flows by accounting for the hypothesized mixed distribution at Pueblo, Colorado, confirms our second hypothesis. Note, however that our Mixed GEV model fit still underestimates the size of The Great Flood of 1921 and the paleoflood bound. Nevertheless, the Mixed GEV model produces substantially smaller underestimation than the LP3 (Fig. 2.5).

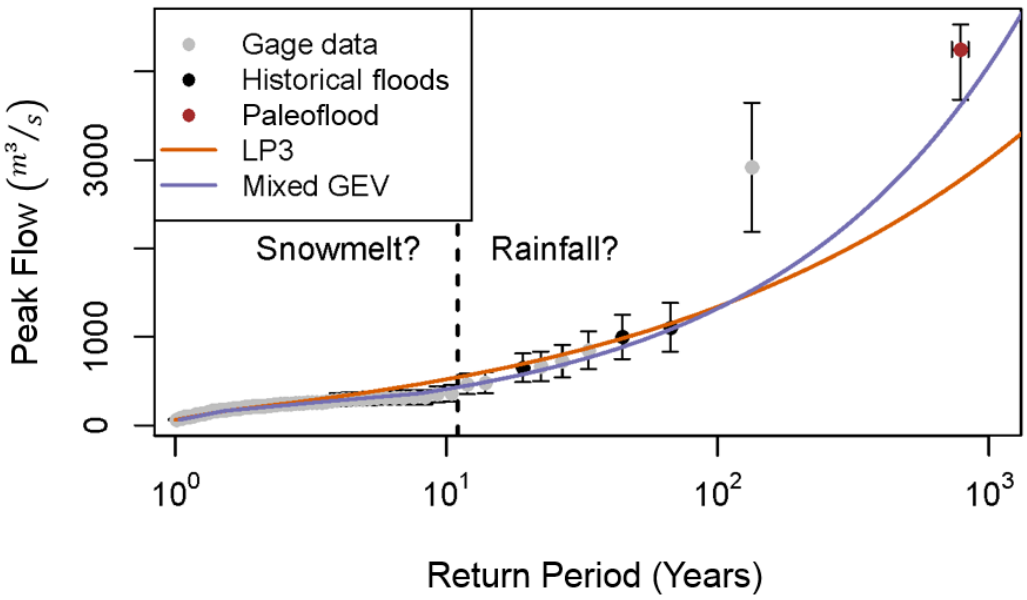

Figure 2.5: Return period plot of the Log Pearson III (LP3) and Mixed Generalized Extreme Value (Mixed GEV) distributions for Pueblo, Colorado. The y-axis shows the peak flow in cubic meters per second, and the x-axis shows the peak flow return period in years. The Mixed GEV distribution predicts larger peak flows for the most extreme events (greater than 112 year return period).

Switching from the method currently recommended by regulation to the statistically better fitting model changes the dam safety classification. Specifically, LP3 estimates overtopping return periods of 123,000 to 664,000 years, spanning the range of current United States Bureau of Reclamation regulation return periods, 131,000 to 376,000 years (England *et al*., 2006). However, Mixed GEV estimates are approximately one order of magnitude shorter, 25,000 to 77,000 years, and do not meet current United States Bureau of Reclamation regulations in any of the three hydrograph scenarios (Fig. 2.6).

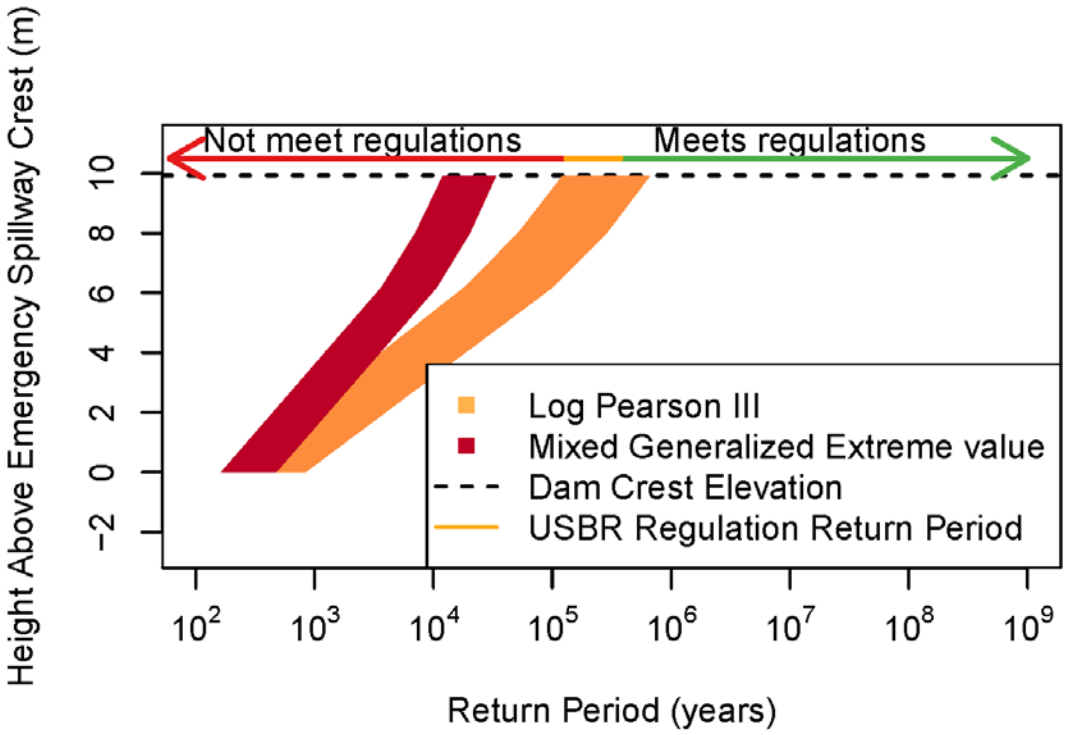

Figure 2.6. Return period plot of reservoir water surface elevation. The y-axis shows the reservoir elevation in meters over the base of the emergency spillway. The x-axis shows the return period in years. The bands of uncertainty show uncertainty in reservoir elevation for a given peak flow due to hydrograph choice (also see Supplementary Fig. S8). The Mixed Generalized Extreme Value distribution (Mixed GEV) predicts lower than regulation return periods of overtopping for the Pueblo Dam, while the Log Pearson III (LP3) predicts a range of overtopping return periods spanning the range of regulation return period uncertainty.

Flood hydrograph uncertainty presents roughly half an order of magnitude of uncertainty in dam overtopping return periods. This is secondary to model structural uncertainty, which accounts for approximately one order of magnitude in dam overtopping return periods (Fig. 2.6). Still, this large source of uncertainty is critical for decisions about dam safety (Fig. 2.6).

Caveats

This didactic case study focuses on method development, model testing, and making a case for further research. This raises a number of caveats, and the results are not to be used to inform actual risk assessments and decision making. For example, I explore the use of mixed distribution statistical flood frequency analysis as an improvement on the relatively fast and easy procedures of flood frequency analysis. This is only one method of dam safety assessment and is not necessarily the optimal method for the safety assessment of the Pueblo Dam (Swain *et al*., 2006). Rather, the study presents the capabilities of mixed distribution flood frequency models to address mixed annual peak flows without prior knowledge of the physical causes of annual peak flows. Additionally, I argue for further consideration of mixed distributions based on the potentially large impacts on assessed flood infrastructure safety.

I focus on Pueblo dam as a didactic case study because it is a well-studied area with an impactful safety question. The assessment of dam safety performed in the study is used for illustrative purposes. I present an argument for further investigation of mixed distributions, particularly the Mixed GEV, not a finished set of methods for flood frequency analysis and dam safety assessment.

Statistical flood frequency analysis is an important aspect of dam and levee safety assessments, however extrapolation to return periods of 376,000 years from 81 years of gage data, three historical floods, and one approximately 785 year return period paleoflood bound is subject to extreme statistical and model structural uncertainty. For this reason the Bureau of Reclamation recommends physically based modeling approaches for such long return periods

(Swain *et al*., 2006; England *et al*., 2014). However, physically based methods are also subject to large uncertainties in watershed and storm behavior (England *et al*., 2014).

I use the statistical method to identify mixed distributions in an area with two known physical mechanisms of floods and a hypothesized mixed distribution. More testing is necessary before it is applied without a strong physical justification, if it should ever be applied without a strong physical justification.

Hydrograph uncertainty is considered by taking three available flood hydrographs determined by three independent methods and assuming each are equally likely, and considering no other hydrograph shapes. Further quantification of flood hydrograph variability and uncertainty is strongly recommended by this investigation.

I encountered difficult implementing the paleoflood age uncertainty likelihood function from O'Connell *et al*. (2002) in R, and assumed the roughly 7% error corresponding to the 95% confidence interval in paleoflood age was secondary to the roughly 10% error in paleoflood magnitude corresponding to the same confidence interval. Inclusion of paleoflood age error may affect confidence intervals at large return periods, but I believe has sufficiently little effect on MLEs for this study.

Discussion

This thesis illustrates the importance of considering the physical context of statistical models and accounting for mixed distributions where they are present. It also proposes a method for using robust goodness-of-fit metrics to identify mixed distributions with statistical models.

Accounting for mixed distributions results in better statistics and potentially large changes in assessed safety. This has implications for flood risk at any locations where multiple physical drivers cause floods.

In this analysis I consider the case of a Pueblo, Colorado, where snowmelt tends to dominate yearly peak flows, but the largest peak flows, most relevant to flooding and dam safety, tend to be rainfall dominated. I show how accounting for the mixed distribution changes the safety of the dam in a didactic example. However, mixed distributions may be present and unaccounted for in many other regions and types of flood prone areas. Mixed distribution are already well documented in the Front Range (England *et al*., 2010, Jarrett and Costa, 1988). Additionally, mixed distributions are posited in mountainous Italian watersheds due to multiple rainstorm distributions (Rossi *et al*., 1984), and there is the potential for mixed distributions where tropical cyclones dominate extreme rainfall such as the East Coast of the United States (Knight and Davis, 2009). This thesis has potentially broad implications for flood risk assessments across many regions and flood causes.

I develop methods for identifying mixed distributions and illustrate the importance of accounting for mixed distribution in flood risk analysis, however, this thesis also illustrates the difficulties numerous difficulties of predicting the size of a roughly 100,000 to 400,000 year return period flood and building to it based on the very limited data available. The best fit statistical method suggests that the dam does not meet current safety regulations for overtopping return periods, while the stochastic storm tranpositioning and rainfall runoff model approach concludes the dam does meet safety regulations. Additionally, both assessments are associated with great and currently unquantified uncertainties.

In the face of extreme statistical and model structural uncertainty, an alternative method may be to adopt adaptive safety standards focused on minimizing loss of life in the case of a disaster. These standards could include improved early warning systems, efficient evacuation plans, or encouraging non-housing development in the potential flood zone. These methods may be a more desirable options for the community given deep uncertainty and finite resources. This thesis does not address potential tradeoffs between dam safety and increased flood survivability, economic or social, however, these are open questions.

Conclusions

I use statistical flood frequency models and goodness-of-fit criteria that account for overfitting to assess the presence of a mixed distribution of annual peak flows at the Pueblo Dam. I confirm the presence of mixed distributions of annual peak flows with this method. While caution and further testing are clearly needed, this method may be used to test the existence of mixed distributions, and address them, at other locations where current methods are inadequate. These methods may help to solve problems raise in (England *et al*., 2018) by contributing to a better understanding of how to identify and deal with mixed distributions.

I determine the mixed GEV is the best model for representing mixed distribution peak flows and for detecting mixed distributions of annual peak flows. I recommend this model for further investigation. However, the seven parameter nature of the distribution makes fitting it by maximum likelihoods difficult.

Ignoring mixed distributions of peak flows where they are present can lead to serious underestimation of extreme events, as exemplified in this case-study. The LP3 distribution fit

shows the dam may be safe, uncertain, or unsafe, depending on the choice of flood hydrograph. However, in all flood hydrograph cases, the Mixed GEV predicts the dam does not meet current regulations based on United States Bureau of Reclamation safety regulations and the loss of life assessment (England *et al*., 2006).

The LP3 distribution underestimates greater than roughly 110 year return period floods as compared to the Mixed GEV. Underestimations are larger at larger return periods. The 100 year flood calculated by the LP3 fit is a roughly 100 year flood based on the Mixed GEV fit, however, the 500 year flood calculated by the LP3 fit is a roughly 330 year flood based on the Mixed GEV fit, and the 1000 year flood calculated by the LP3 fit is a roughly 540 year flood based on the Mixed GEV fit.

This study uses both AIC and BIC to assess goodness-of-fit. The metrics are derived from different assumptions and calculated with different equations, but are both widely used and accepted as model selection criteria that properly protect against overfitting. Due to the asymptotic nature of these metrics and the small sample size, I strongly recommend caution when these metrics do not clearly agree on a best fit distribution.

Acknowledgements

I would like to thank the members of the Keller research group for thoughtful discussions about this work. Special thanks to Benjamin Seiyon Lee for vital mentorship in statistics and programming, to Caitlin Spence for personal and professional mentorship and valuable insights into the world of hydrological engineering, to John F England, Jr. for sharing the TREX hydrographs and for incredibly helpful recommended readings, and to Randy Miller for

reviewing and testing the reproducibility of the analysis' code. This study was partially co-supported by the Department of Energy sponsored Program on Coupled Human Earth Systems (PCHES) and the Penn State Center for Climate Risk management. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entities. All coauthors contributed to the interpretation of the results, writing, and revision of the manuscript

Code Availability and Disclaimer

All results, model codes, analysis codes, data and model outputs used for analysis are freely available from https://github.com/Joelroopeckart/mixed-distributions and are distributed under the GNU general public license. The datasets, software tools, results, and other resources in this thesis and the cited website are provided as-is without warranty of any kind, express or implied. In no event shall the authors or copyright holders be liable for any claim, damages or other liability in connection with the use of these resources.

References

Baker, J.H., Hafen, L.R. 1927. *History of Colorado Volume II*. Linderman Co., Denver, CO: 429–867.

Bullard, KL. Leverson, V. 1991. "Pueblo Dam, Fryingpan-Arkansas Project Probable Maximum Flood (PMF) Study." *Bureau of Reclamation*.

Campbell, M.R. 1922. Guidebook of the western United States, Part E. The Denver and Rio Grande Western Route: U.S. Geological Survey Bulletin, 707. *United States Government Printing Office*, Washington, DC: 266.

(COST) the European Cooperation in Science and Technology, and the Centre for Ecology and Hydrology. 2013. "WG4 : Flood Frequency Estimation Methods and Environmental Change." *Center for Ecology and Hydrology*.

Ellingwood, Bruce, Ross B Corotis, John Boland, and Nicholas P Jones. 1993. "Assessing Cost of Dam Failure." *Journal of Water Resources Planning and Management* 119 (1): 64–82. https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%290733-9496%281993%29119%3A1%2864%29.

England, JF. Jr., Jeanne E Klawon, Ralph E Klinger, and Travis R Bauer. 2006. "Flood Hazard Study Pueblo Dam, Colorado." *United States Bureau of Reclamation.*

England, JF. Jr. 2011. "Flood Frequency and Design Flood Estimation Procedures in the United States: Progress and Challenges." *Australian Journal of Water Resources* 15 (1): 33–47. doi:10.1080/13241583.2011.11465388.

England, JF. Jr. 2010. "Geomorphology Paleohydrologic Bounds and Extreme Flood Frequency of the Upper Arkansas River, Colorado, USA." *Geomorphology* 124 (1–2). Elsevier B.V.: 1–16. doi:10.1016/j.geomorph.2010.07.021.

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr. 2018. Guidelines for determining flood flow frequency—Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. B5: 148, https://doi.org/10.3133/tm4B5.

Follansbee, R., Jones, E.E. 1922. "Arkansas River flood of June 3–5, 1921." United States Geological Survey Water-Supply Paper 487. *United States Geological Survey*: 44.

Foster, Mark, Robin Fell, and Matt Spannagle. 2000. "The Statistics of Embankment Dam Failures and Accidents." *Canadian Geotechnical Journal* 37 (5): 1000–1024. doi:10.1139/t00-030.

Graham, Wayne J. 2009. "Major U.S. Dam Failures: Their Cause, Resultant Losses, and Impact on Dam Safety Programs and Engineering Practice." *World Environmental Water Resources Congress 2009: Great Rivers History*: 52–60. http://www.nrcresearchpress.com/doi/abs/10.1139/t00-030.

Hafen, L.R. 1948. *Colorado and Its People. A Narrative and Topical History of the Centennial State, Volume I*. Lewis Historical Publishing, NY: 644.

Jackson, Christopher. 2016. flexsurv: A Platform for Parametric Survival Modeling in R. *Journal of Statistical Software* 70 (8): 1-33. doi:10.18637/jss.v070.i08

Jarrett, R.D., and Costa, J.E., 1988, "Evaluation of the flood hydrology in the Colorado Front Range using precipitation, streamflow, and paleoflood data for the Big Thompson River Basin: U.S. Geological Survey Water-Resources Investigations Report 87–4117." *United States Geological Survey*: 37, accessed May 13, 2018, at https://pubs.er.usgs.gov/publication/ wri874117.

Kenneth V. Price, Rainer M. Storn and Jouni A. Lampinen. 2006. "Differential Evolution - A Practical Approach to Global Optimization." Berlin Heidelberg: Springer-Verlag. ISBN 3540209506.

Knight, David B., and Robert E. Davis. 2009. “Contribution of Tropical Cyclones to Extreme Rainfall Events in the Southeastern United States.” *Journal of Geophysical Research Atmospheres* 114 (23): 1–17. doi:10.1029/2009JD012511.

O’Connell, Daniel R. H. 2002. “Bayesian Flood Frequency Analysis with Paleohydrologic Bound Data.” Water Resources Research 38 (5). doi:10.1029/2000WR000028.

Oddo, Perry C., Ben S. Lee, Gregory G. Garner, Vivek Srikrishnan, Patrick M. Reed, Chris E. Forest, and Klaus Keller. 2017. “Deep Uncertainties in Sea-Level Rise and Storm Surge Projections: Implications for Coastal Flood Risk Management.” *Risk Analysis*. doi:10.1111/risa.12888.

Prentice, R. L. 1975. “Discrimination among Some Parametric Models.” *Biometrika* 62 (3): 607–614. doi:10.1093/biomet/62.3.607.

Raynal-Villasenor, Jose a. 2012. “Maximum Likelihood Parameter Estimators for the Two Population GEV Distribution.” *International Journal of Research and Reviews in Applied Sciences* 11 (3): 350–357.

Rossi, Fabio, Mauro Fiorentino, and Pasquale Versace. 1984. “Two-Component Extreme Value Distribution for Flood Frequency Analysis.” *Water Resources Research* 20 (7): 847–856. doi:10.1029/WR020i007p00847.

Stedinger, Jery R., and Timothy A. Cohn. 1986. “Flood Frequency Analysis With Historical and Paleoflood Information.” *Water Resources Research* 22 (5): 785–793. doi:10.1029/WR022i005p00785.

Swain, R.E., J.F. England, K.L. Bullard, and D.A. Raff. 2006. “Guidelines for Evaluating Hydrologic Hazards.” *The United States Bureau of Reclamation*. https://www.engr.colostate.edu/~pierre/ce_old/classes/ce717/Hydrologic_Hazard_Guidelines_final.pdf.

Wong, Tony E., Alexandra Klufas, Vivek Srikrishnan, and Klaus Keller. 2018. “Neglecting Model Structural Uncertainty Underestimates Upper Tails of Flood Risk.” *Environmental Research Letters. in press* https://doi.org/10.1088/1748-9326/aacb3d

Wuertz, Diethelm, Tobias Setz, and Yohan Chalabi. 2013. fExtremes: Rmetrics – Extreme Financial Market Data. R package version 3010.81. https://CRAN.R-project.org/package=fExtremes

Chapter 3

Research Opportunities and Needs

Advancements made and to be made

This thesis explores applications and methods for addressing mixed distributions in flood frequency analysis, and adds to current literature in identifying and addressing mixed distributions of annual peak flows. However, this is far from an exhaustive study. Many open challenges remain. This chapter discusses two broad directions for future work, uncertainty quantification and further testing and model comparison.

Uncertainty quantification

This thesis uses maximum likelihood estimates (MLEs) to fit each flood frequency distribution. MLEs are best estimates and do not quantify associated parameter uncertainties. Further work is needed to establish uncertainty bounds for the mixed distributions. This could

potentially be accomplished using Markov Chain Monte Carlo simulations with sampling methods designed to handle highly correlated parameter structures associated with mixed models. Preliminary work suggests both are difficult and require further work in statistical and computational implementation on this problem.

This thesis assumes the three potential flood hydrographs (England *et al*., 2014, 2006; Bullard and Leverson, 1991) are all equally likely and scale reasonably to flood peak magnitudes unobserved in the known record (Swain *et al*., 2006). This assumption of point estimated equi-probable flood hydrographs are made under unknown uncertainty.

Peak flow data uncertainty is estimated based on generalized large flood uncertainty estimates in O'Connell *et al*. (2002) for stream gages and historical and paleofloods in the western United States. The generalized uncertainties may not reflect peak flow measurement errors in the study area. I recommend further investigation of annual peak flow, historical flood, and paleoflood uncertainties on both broad and local scales.

This study only tests the three single distributions and three mixed distributions discussed in table 2.1. Peak over threshold approaches (Lang *et al*., 1999; Bezak *et al*., 2014) and other statistical flood frequency distributions, such as the Wakeby distribution (Houghton, 1978) or generalized logistic distribution (Ahmed *et al*., 1988) are not considered.

This thesis assumes the Pueblo Dam can only fail by overtopping. However, other failure modes are possible (Foster *et al*., 2000; England *et al*., 2002). The dam could potential experience structural failure prior to overtopping due high water levels behind the dam. These failure

probabilities are not quantified or included in this study and constitute a continuing challenge in comprehensive dam safety assessments.

Further testing and model comparison

This thesis focuses on method development and testing using a real world problem for demonstration purposes. I make a case for further investigation into (i) using statistical distributions to gain insight into peak flow data and the hydrological system, (ii) using mixed distribution statistical models where mixed distributions of peak flows cannot be addressed by current composite methods (England *et al*., 2018), and (iii) the importance of hydrograph uncertainty on dam safety assessments.

Bulletin 17C (England *et al*., 2018) outlines methods for handling potentially influential low flows (PILFs). A mixed distribution can be thought of as a case of PILFs, where the lower distribution is made up of PILFs. There are important questions about how these methods might handle a mixed distribution case differently from a mixed flood frequency distribution.

Further study is necessary to fully investigate the advantages and disadvantages of mixed flood frequency distributions over alternative methods. For further studies, synthetic datasets may be helpful in creating a known testing environment for comparing the accuracy of eliminating PILFs and fitting a single distribution vs fitting a mixed distribution to the entire dataset. More investigation of the robustness of PILFs vs mixed distributions is important. While these methods have been investigated independently (England *et al*., 2018; Jarrett and Costa, 1988, Rossi *et al*., 1984; Raynal-Villasenor, 2012) they have not been comprehensively compared.

References

Ahmad, M.I., C.D. Sinclair, and A. Werritty. 1988. “Log-Logistic Flood Frequency Analysis.” *Journal of Hydrology* 98: 205–24. doi:10.1016/0022-1694(88)90015-7.

Bezak, Nejc, Mitja Brilly, and Mojca Šraj. 2014. “Comparison between the Peaks-over-Threshold Method and the Annual Maximum Method for Flood Frequency Analysis.” *Hydrological Sciences Journal* 59 (5): 959–77. doi:10.1080/02626667.2013.831174.

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr., 2018, Guidelines for determining flood flow frequency—Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p., https://doi.org/10.3133/tm4B5.

Foster, Mark, Robin Fell, and Matt Spannagle. 2000. “The Statistics of Embankment Dam Failures and Accidents.” Canadian Geotechnical Journal 37 (5): 1000–1024. doi:10.1139/t00-030.

Houghton, John C. 1978. “Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows.” *Water Resources Research* 14 (6): 0–4.

Jarrett, R.D., and Costa, J.E., 1988, Evaluation of the flood hydrology in the Colorado Front Range using precipitation, streamflow, and paleoflood data for the Big Thompson River Basin: U.S. Geological Survey Water-Resources Investigations Report 87–4117, 37 p., accessed May 13, 2018, at https://pubs.er.usgs.gov/publication/ wri874117.

Lang, M, TBMJ Ouarda, and B Bobée. 1999. “Towards Operational Guidelines for Over-Threshold Modeling.” *Journal of Hydrology* 225: 103–17.

O’Connell, Daniel R. H. 2002. “Bayesian Flood Frequency Analysis with Paleohydrologic Bound Data.” Water Resources Research 38 (5). doi:10.1029/2000WR000028.

Raynal-Villasenor, Jose a. 2012. “Maximum Likelihood Parameter Estimators for the Two-Population GEV Distribution.” *International Journal of Research and Reviews in Applied Sciences* 11 (3): 350–57.

Rossi, Fabio, Mauro Fiorentino, and Pasquale Versace. 1984. “Two-Component Extreme Value Distribution for Flood Frequency Analysis.” *Water Resources Research* 20 (7): 847–56. doi:10.1029/WR020i007p00847.

Swain, R.E., J.F. England, K.L. Bullard, and D.A. Raff. 2006. “Guidelines for Evaluating Hydrologic Hazards.” June: 1–91.

Appendix A

Supplementary figures

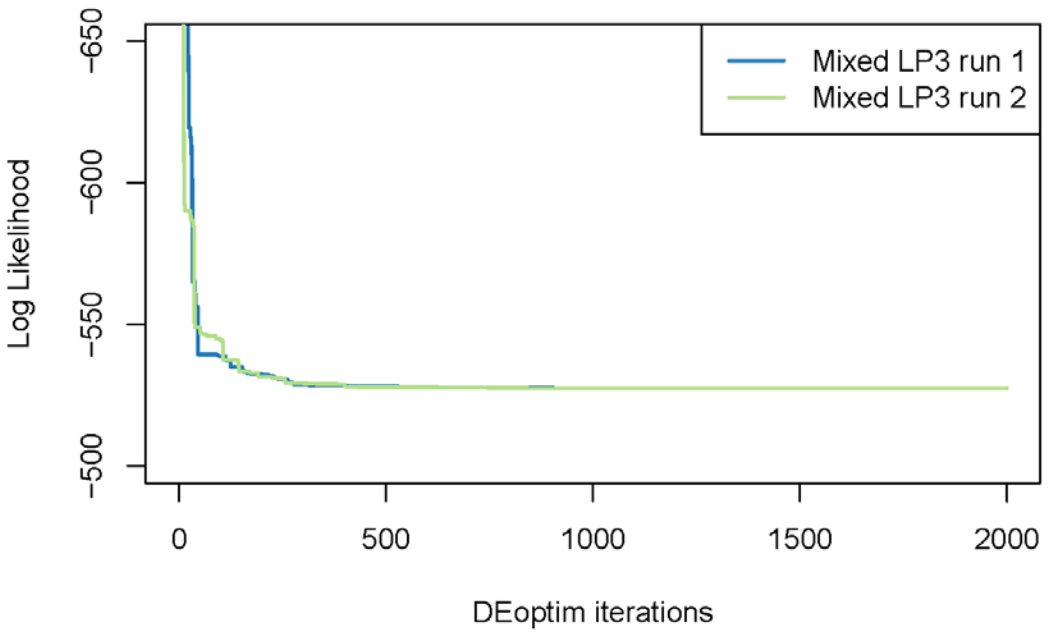

Figure S1: Log likelihood agreement between two independent MLE using DEoptim. Each independent MLE was calculated with 2,000 DEoptim iterations. The independent runs converge to the same log likelihood within 1,000 iterations, MLE parameter values differ by less than 0.0001%.

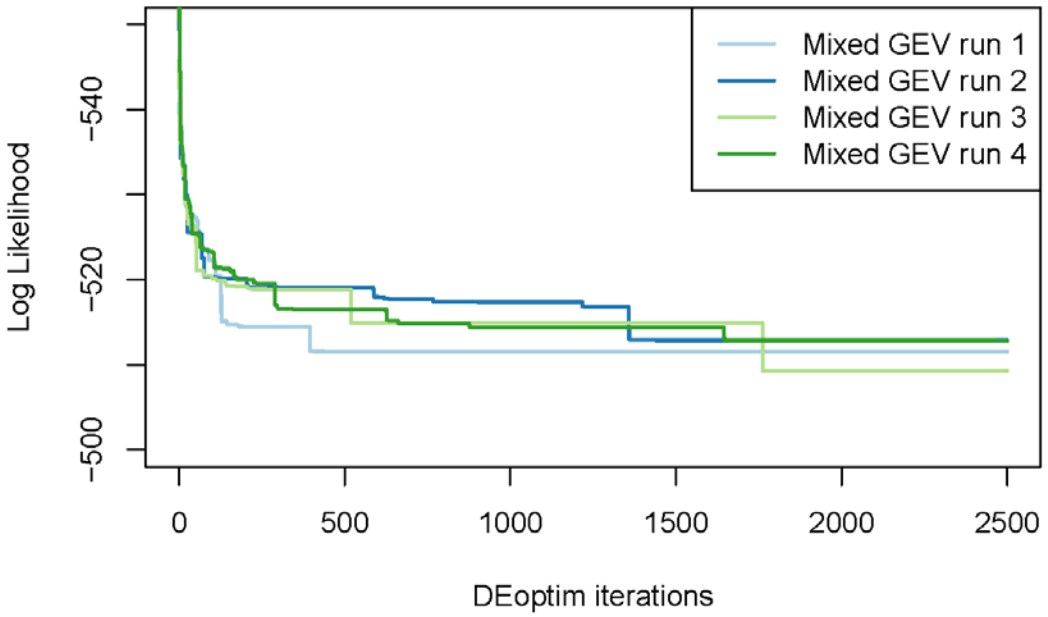

Figure S2: Log likelihood agreement and disagreement between four independent MLEs using DEoptim. Each MLE is determined by 2,500 DEoptim iterations. The independent runs do not converge to a global maximum. For this analysis I use the highest log likelihood, run one. More iterations or a better optimizer for this problem could potentially produce improved MLEs.

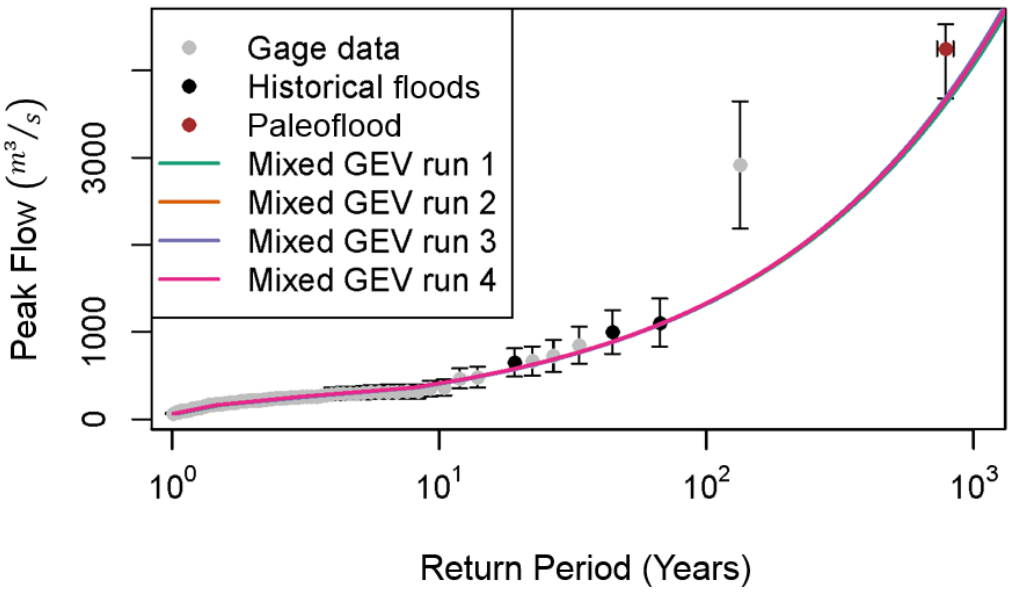

Figure S3: Return period plot comparing the four independent Mixed GEV MLE fits. The four independent MLEs calculated by DEoptim are virtually identical in the magnitude of events up to 1,000 year return periods.

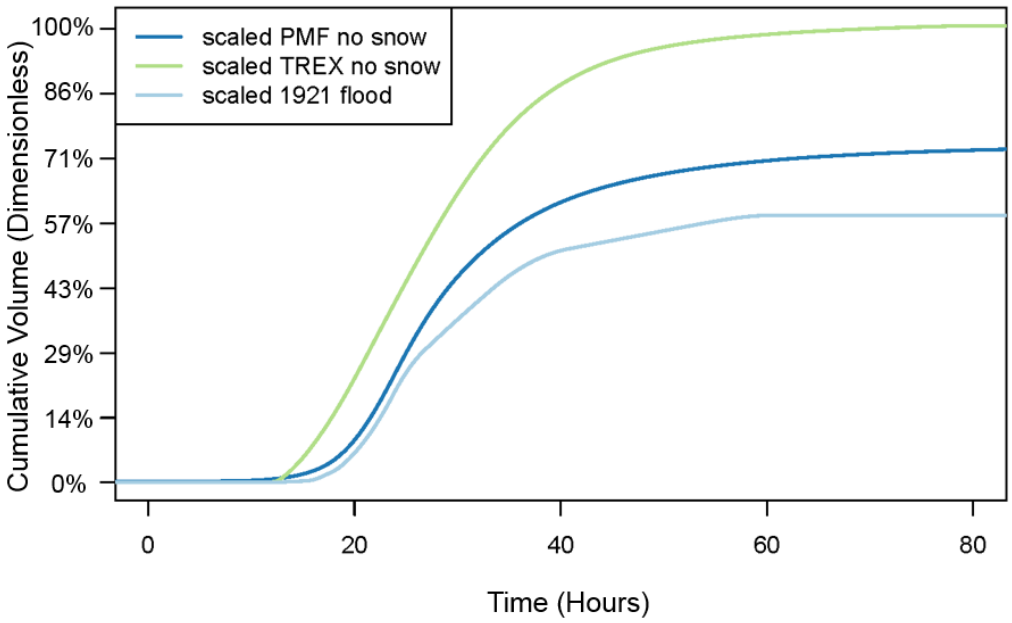

Figure S4: Cumulative volume of water delivered over the duration of each considered flood hydrograph as a function of time. Each considered flood hydrograph was scaled to the same reference peak flow.

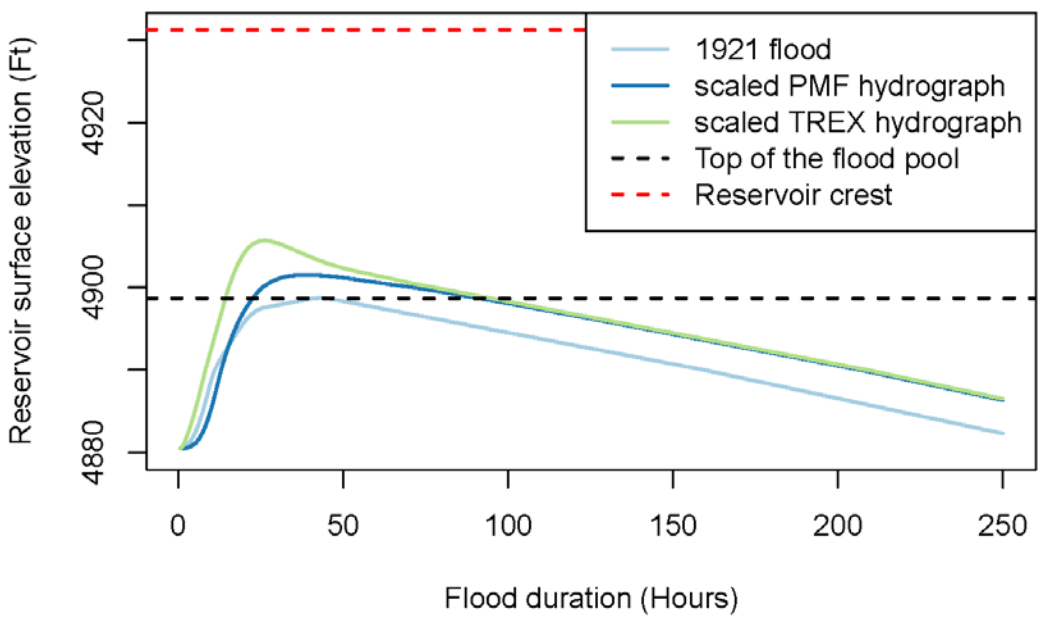

Figure S5: Reservoir water surface elevation response to each considered flood hydrograph scaled to the 1921 flood peak flow. Differences in cumulative water delivered by each flood hydrograph produce differing reservoir surface elevation responses, and different levels of hazard. The flood pool consists of the extra available storage in the reservoir allotted to flood control. The emergency spillway is at the top of the flood pool. The reservoir crest represents the maximum possible water surface elevation in the reservoir before overtopping. I used FLROUT to route the flood hydrographs.

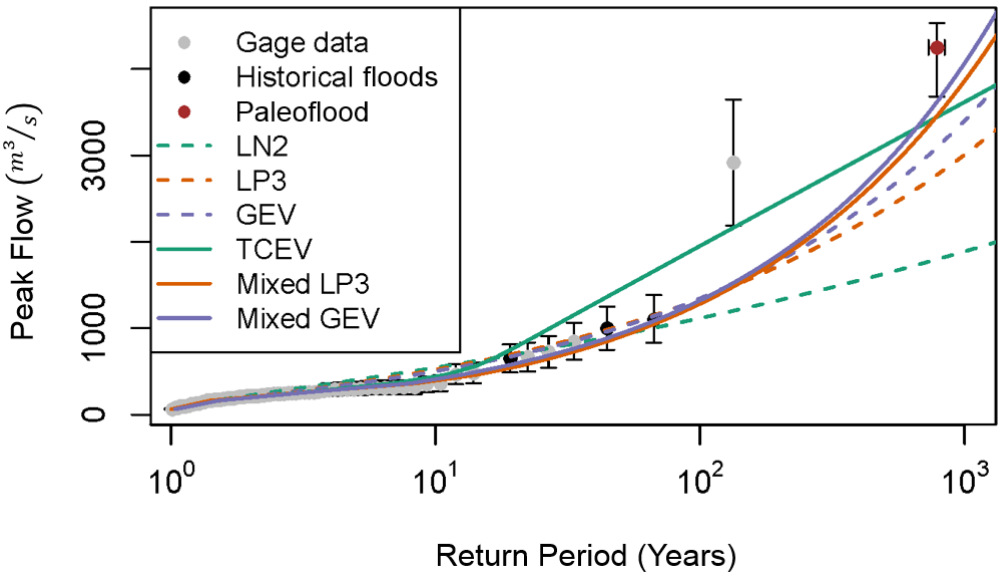

Figure S6: Return period plot of each considered flood frequency model fit and the data at Pueblo, Colorado.

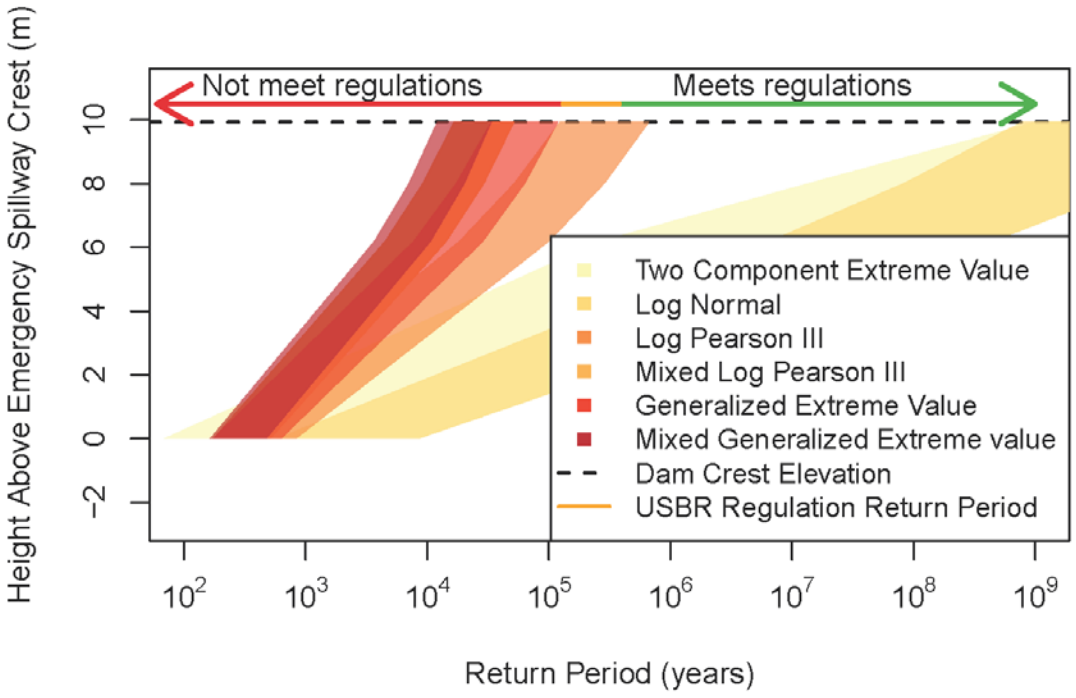

Figure S7: Return period plot of peak reservoir elevations for each of the considered flood frequency models.

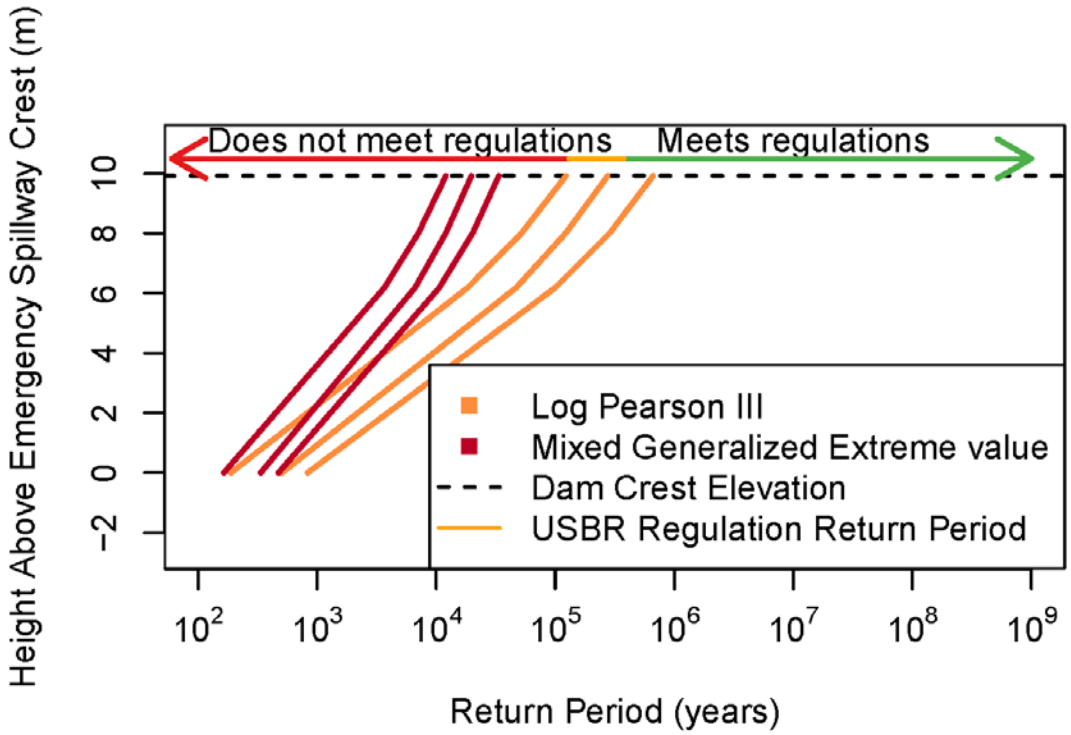

Figure S8: Return period plot of reservoir water surface elevation. The y-axis shows the reservoir elevation in meters over the base of the emergency spillway. The x-axis shows the return period in years. The lines for a given peak flow and hydrograph shape. Three possible hydrographs were considered for each distribution, resulting in three possible reservoir elevation to return period curves for each.

Appendix B

Hydrograph scaling

The hydrograph scaling in this study follows the methods described in Swain *et al*. (2006). Three suitable flood discharge hydrographs are obtained from three independent sources of information, a physical rainfall-runoff model (England *et al*., 2014), a regionally representative PMF hydrograph (Bullard and Leverson, 1991), and the flood of record (England *et al*., 2010). These hydrographs are time series of roughly durations of half hour average discharges in units of m^3s^{-1}, with the exception of the 1921 flood, which is linearly interpolated to half hour increments from roughly four hour increments due to poor temporal sampling during the flood. The hydrographs remain constant in duration and discharge is scaled linearly in so that the maximum half hour averaged discharge matches the peak flow the hydrograph is scaled to.

Reservoir elevations are determined by FLROUT, the Bureau of Reclamation flood routing program (England *et al*., 2006). The program uses rating curves of reservoir volume-elevation and elevation-discharge relationships to determine elevation and outflows given flood hydrograph inflows for each half hour time-step.

For the dam safety assessment, overtopping peak flows are back calculated by scaling each of the three flood hydrographs to the smallest peak flow that causes dam overtopping, i.e. when the peak reservoir elevation exceeds the dam height. The return periods for each of the three peak flows are then calculated for each flood frequency model. The method produces three estimates of dam overtopping return period for each flood frequency model. These three estimates, based on hydrograph uncertainty, are considered to be the range of possible dam overtopping return periods for each flood frequency model. In the results I consider only the Log Pearson Type III and the Mixed Generalized Extreme Value distribution results. The results from all six models are included in Appendix A.