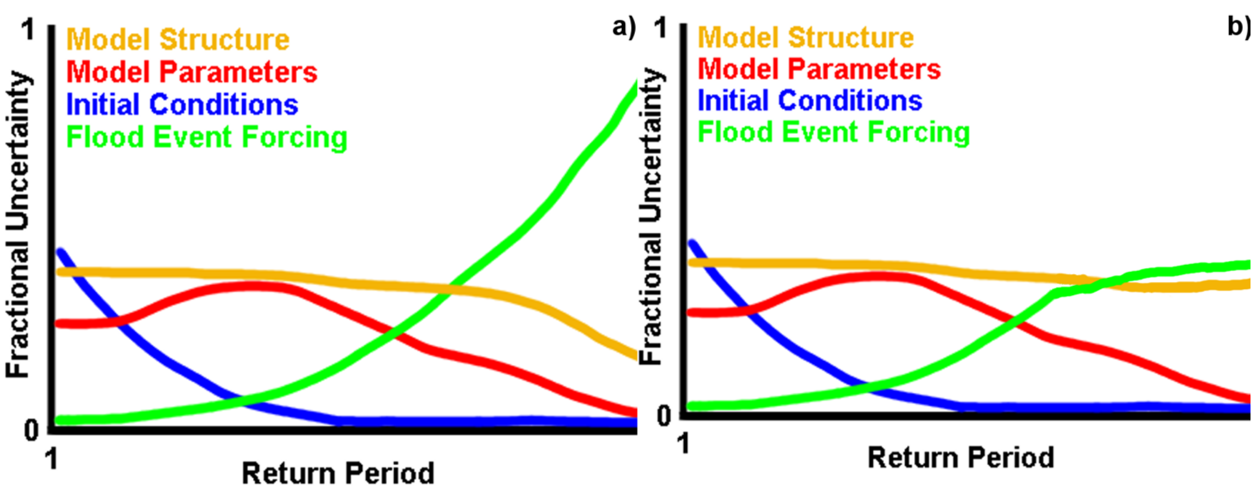
Identifying Sources of Uncertainty in Flood Frequency Analyses

Science and Technology Program
Research and Development Office
Final Report No. ST-2020-1794-1
Technical Memo No. ENV-2020-076



U.S. Department of the Interior

September 30, 2020

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-09-2020	Research	05/2017-09/2020

4. TITLE AND SUBTITLE
Identifying Sources of Uncertainty in Flood Frequency Analyses

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
1541 (S&T)

6. AUTHOR(S)
Amanda Stone, PE
Andrew Newman, Ph.D.
Kathleen Homan, Ph.D.

5d. PROJECT NUMBER
Final Report ST-2020-1794-1

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Bureau of Reclamation
U.S. Department of the Interior
Denver Federal Center
PO Box 25007, Denver, CO 80225-0007

National Center for Atmospheric Research
PO Box 3000
Boulder, CO 80307

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Science and Technology Program
Research and Development Office
Bureau of Reclamation
U.S. Department of the Interior
Denver Federal Center
PO Box 25007, Denver, CO 80225-0007

10. SPONSOR/MONITOR'S ACRONYM(S)
Reclamation

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Report ST-2020-1794-1

12. DISTRIBUTION/AVAILABILITY STATEMENT
Final Report may be downloaded from https://www.usbr.gov/research/projects/index.html

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Reclamation partnered with scientists at NCAR to assess sensitivity and sources of uncertainty in rainfall runoff modeling to support flood frequency analyses. Pieces of the modeling chain, including initial conditions, model parameters, precipitation inputs, and model structure were examined across a range of return intervals from 2-100,000 years. Two example watersheds representing different hydrologies in the 17 western states were used for this study. A stochastic hydrologic modeling workflow was developed using NCAR's FUSE modeling framework. Results indicated that modeling chain pieces have variable uncertainty contributions across return periods. Results found that precipitation inputs are most important for rare events while initial conditions are important for frequent events; however, uncertainties from model structure and structure-parameter interactions still play a proportionally important role in results. This highlights the importance of understanding flood generation processes and selecting appropriate models based on that understanding. In addition to these findings, a review of calibration metrics indicated that the KGE is a more robust metric than NSE for calibration of models for extreme events.

15. SUBJECT TERMS
Flood frequency, hydrologic model structures, stochastic methods, uncertainty analysis, model selection

16. SECURITY CLASSIFICATION OF:			19a. NAME OF RESPONSIBLE PERSON

			17.	18.	
a. REPORT U	b. ABSTRACT U	THIS PAGE U	LIMITATION OF ABSTRACT	NUMBER OF PAGES	19b. TELEPHONE NUMBER *(Include area code)*

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Mission Statements

The Department of the Interior (DOI) conserves and manages the Nation's natural resources and cultural heritage for the benefit and enjoyment of the American people, provides scientific and other information about natural resources and natural hazards to address societal challenges and create opportunities for the American people, and honors the Nation's trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities to help them prosper.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Disclaimer

Information in this report may not be used for advertising or promotional purposes. The data and findings should not be construed as an endorsement of any product or firm by the Bureau of Reclamation, Department of Interior, or Federal Government. The products evaluated in the report were evaluated for purposes specific to the Bureau of Reclamation mission. Reclamation gives no warranties or guarantees, expressed or implied, for the products evaluated in this report, including merchantability or fitness for a particular purpose.

Acknowledgements

The Science and Technology Program, Bureau of Reclamation, sponsored this research.

Identifying Sources of Uncertainty in Flood Frequency Analyses

Final Report No. ST-2020-1794-1
Technical Memo No. ENV-2020-076

prepared by

Amanda Stone, PE, Civil Engineer (Hydrologic)
Water Resources Engineering and Management Group
Technical Service Center
Bureau of Reclamation

Andrew Newman, Ph.D., Project Scientist
National Center for Atmospheric Research

Kathleen Holman, Ph.D., Meteorologist
Water Resources Engineering and Management Group
Technical Service Center
Bureau of Reclamation

Peer Review

Bureau of Reclamation
Research and Development Office
Science and Technology Program

Final Report ST-2020-1794-1
Technical Memo No. ENV-2020-076

Identifying Sources of Uncertainty in Flood Frequency Analyses

AMANDA STONE
Digitally signed by AMANDA STONE
Date: 2020.10.01 07:23:45 -06'00'

Prepared by: Amanda Stone, PE
Civil Engineer (Hydrologic), Bureau of Reclamation

Andrew Newman

Prepared by: Andrew Newman, Ph.D.
Project Scientist, National Center for Atmospheric Research

KATHLEEN HOLMAN
Digitally signed by KATHLEEN HOLMAN
Date: 2020.10.01 08:34:31 -06'00'

Prepared by: Kathleen Holman, Ph.D.
Meteorologist, Bureau of Reclamation

FRANK DWORAK
Digitally signed by FRANK DWORAK
Date: 2020.09.30 19:35:54 -06'00'

Peer Review by: Frank Dworak, PE
Civil Engineer (Hydrologic), Bureau of Reclamation

"This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Bureau of Reclamation. It does not represent and should not be construed to represent Reclamation's determination or policy."

Acronyms and Abbreviations

AEP	Annual exceedance probability
ANOVA	Analysis of variance
FF	Flood frequency
FUSE	Framework for Understanding Structural Errors
IC	Initial conditions
KGE	Kling-Gupta Efficiency
netCDF	Network common data format
NCAR	National Center for Atmospheric Research
NSE	Nash-Sutcliffe Efficiency
Reclamation	Bureau of Reclamation
RMSE	Root mean squared error
SCE	Shuffled Complex Evolution
TSC	Technical Service Center

Measurements

ft^3/s	cubic feet per second
km^2	square kilometers
mi^2	square miles
mm	millimeters

Contents

Executive Summary

Understanding flood risk for Bureau of Reclamation (Reclamation) facilities is important for both safety as well as to support design and operations. Hydrologic hazard curves and flood hydrographs are required to evaluate hydrologic risks for Reclamation facilities. There are numerous approaches to developing these curves, including statistical streamgage analysis, annual exceedance probability (AEP) neutral rainfall-runoff hydrologic model estimates, or more complex fully stochastic rainfall-runoff modeling. Most often, multiple methods are employed in these analyses to help understand uncertainty related to model results.

In the case of stochastic rainfall-runoff modeling, flood frequency (FF) estimates are produced using stochastic event simulations with randomly perturbed initial conditions (ICs), model parameters, and precipitation event forcing scenarios from defined precipitation frequency curves using one model structure. These pieces are referred to in this report as the "modeling chain." FF estimates from stochastic modeling are sensitive to the IC, model parameter, and precipitation event forcing perturbations portions of the modeling chain, but they are also sensitive to model structure (e.g. how the model calculates runoff), and the specification of meteorology before/after the flood event inputs. Model structure, or how the model calculates runoff, can vary widely. Model structure and event sequencing are not explored as part of project scoping process in the current Reclamation stochastic modeling system. The goal of the proposed research is to inform model choice, structure, and parameterization to more efficiently and accurately estimate flood loads for Reclamation infrastructure. Currently, there is not a substantial effort to choose model structure and focused calibration parameters based on hydrologic region, nor understand the relative contributions to FF uncertainty across the stochastic model components (initial conditions, model parameters and structure, precipitation event forcing and event sequencing).

The primary research question for this study is: What is the total uncertainty related to flood frequency analyses, and what aspects of the modeling chain in stochastic FF analysis have the most sensitivity across a range of return intervals spanning 2-100,000 years? Our baseline hypothesis was that for rare flood events (large return periods) the uncertainty related to the precipitation event dominates the total uncertainty of an FF estimate. We postulate that variability in FF estimates arises from the aforementioned factors and that there may be important contributions to FF uncertainty outside of precipitation event forcing for extreme events. We explore these key components of the modeling chain by: 1) using a multi-hydrologic model ensemble, 2) sampling model parameters across the model structures using calibrated model parameters, 3) sampling model initial conditions that are internally consistent for each model structure from continuous calibrated long-term simulations, 4) incorporating the statistical uncertainty of the precipitation event distributions, and 5) specifying two meteorological sequences to force a stochastic (ensemble) event simulation framework. We used the analysis of variance (ANOVA) methodology to examine relative contributions of uncertainty to FF estimates across the return periods of interest. While the study was focused on stochastic modeling, results regarding model sensitivity and uncertainty are directly applicable to all forms of rainfall-runoff modeling used for FF estimates.

The Island Park Dam in Idaho and Altus Dam in Oklahoma watersheds are used as representative basins of mountainous snowmelt (Island Park) and semiarid high plains (Altus) hydrologies, respectively. These two types of hydrologies are similar to many Reclamation installations. This study then developed a stochastic hydrologic modeling workflow containing the Framework for Understanding Structural Errors (FUSE) hydrologic modeling framework, the Shuffled Complex Evolution (SCE) optimization algorithm, and precipitation frequency distributions from Reclamation. Additionally, we used the total probability theorem and the analysis of variance methods to compute the FF estimates and partition the partial uncertainty contributions across the workflow components, respectively.

Our results show that careful consideration of the various components of flood modeling should be undertaken as the above factors impact hydrologic model behavior and the final uncertainty estimates of FF studies. We reaffirm that calibration metrics truly only constrain model behavior for the portions of the hydrograph most related to the calibration metric (e.g. Mendoza et al. 2015, Mizukami et al. 2019). For streamflow-based calibration metrics, KGE is a robust metric that provides good model behavior across all components of the hydrograph and should be used over NSE if possible. Furthermore, metrics focusing on only peak flow events often may not capture performance for other parts of the hydrograph, such that those calibrated hydrologic models may have inferior performance for longer duration volume flood metrics.

We find that in general ICs are most important for frequent events and the precipitation frequency distribution specification is most important for extreme events. Varying the combinations of model structures results in scenarios where model structure is of similar importance to ICs and precipitation event forcing for frequent and extreme events respectively. Additionally, model parameters and model structure-parameter interactions can also have similar uncertainty contributions to ICs and precipitation event forcing for less constrained calibrations. Variations in model parameters are only important in Altus, where the available calibration data limited the ability for calibration to constrain model performance. The following key generalizable conclusions relevant to Reclamation have resulted from this work:

1) ICs and precipitation frequency distributions generally contribute the most uncertainty in the stochastic flood modeling chain for frequent and extreme events respectively.

2) Model structure can be equally as important given a diverse set of model responses, particularly for multi-day volume flood metrics. This highlights the need to understand basin flood generation processes and develop methods to select appropriate models. This includes examination of the AEP neutral assumption and selecting model process parameterizations that are most plausible for the study basin.

3) Model parameter and model structure-parameter interactions may be important if the sampled model parameter space is not well constrained by calibration.

4) The Kling-Gupta Efficiency (KGE) is a more robust metric than NSE (or RMSE) for calibration of models related to extreme events and volume integrated floods. This is because it is formulated in a manner that permits user understanding of how correlation, variance, and bias contribute to model performance, and is more flexible for application specific uses than NSE or RMSE.

1. Introduction

Understanding flood risk for Bureau of Reclamation (Reclamation) facilities is important for both safety as well as to support design and operations. Hydrologic hazard curves and flood hydrographs are required to evaluate hydrologic risks for Reclamation facilities. A hydrologic hazard curve is a curve that relates probability of occurrence to magnitude of a flood. There are numerous approaches to developing these curves, including statistical streamgage analysis, annual exceedance probability (AEP) neutral rainfall-runoff hydrologic model estimates (where the return period of the flood is equal to the return period of the precipitation), or more complex fully stochastic rainfall-runoff modeling. Most often, multiple methods are employed in these analyses to help understand uncertainty related to model results.

In the case of stochastic rainfall-runoff modeling, flood frequency (FF) estimates are produced using stochastic event simulations with randomly perturbed initial conditions (ICs), model parameters, and precipitation event forcing scenarios from defined precipitation frequency curves using a single model structure. These pieces are referred to in this report as the "modeling chain." FF estimates from stochastic modeling are sensitive to IC, model parameter, and precipitation event forcing portions of the modeling chain, but they are also sensitive to model structure (e.g. how the model calculates runoff), and the specification of meteorology before/after the flood event inputs. Model structure, or how the model calculates runoff, can vary widely. Model structures can be simple defined by a single loss methodology or can be more complex employing various methods to develop and melt snowpack and store and route subsurface flows. Additionally, the methods used to perturb model parameters and forcings do not improve understanding of which component contributes the most variance to an FF estimate. In this study, we systematically explored FF uncertainty to provide a better understanding of which components of the modeling chain cause sensitivity to FF estimates across example hydroclimatic regimes within the 17 Western States. In addition to the relatively well known need to quantify IC and precipitation input variability in FF estimates, recent research highlights the differences in model performance and responses for various event types given different model parameters, and model structures (e.g. Mendoza et al. 2015; Newman et al. 2015, 2017; Markstrom et al. 2016; Mizukami et al. 2019) across hydroclimates, motivating the inclusion of multiple model structures and more than one basin. While the focus of this study was on stochastic rainfall-runoff modeling, methods and implications can be applied to more simplistic rainfall-runoff modeling as well, such as AEP-neutral model estimates.

The goal of this study is to improve both the quality and efficiency of the hydrologic risk estimates for Reclamation infrastructure in the 17 Western States through improved understanding of model uncertainty, specifically understanding which components of the modelling chain contribute to the largest sensitivity in model results. The methodologies and results presented in this paper examine and provide insight into sensitivity of rainfall-runoff model inputs, parameters, ICs, and model structure for two representative hydrologies of the 17 States located west of the Mississippi River.

1.1 Research Question

The primary research question is: What is the total uncertainty related to flood frequency analyses, and what aspects of the modeling chain in stochastic FF analysis have the most sensitivity across a range of return intervals spanning 2-100,000 years?

Our baseline hypothesis is that for rare floods (floods with large return periods) the uncertainty related to the precipitation event forcing dominates the total uncertainty of a FF estimate as seen in Figure 1. We postulate that variability in FF estimates arises from the aforementioned factors: 1) initial conditions, 2) precipitation event forcing, 3) model parameters, 4) model structure, and that there may be other dominant factors contributing to FF uncertainty outside of precipitation event forcing for rare floods. We explore these key components of the modeling chain by: 1) using a multi-hydrologic model ensemble, 2) sampling model parameters across the model structures, 3) sampling model initial conditions that are internally consistent for each model structure from continuous calibrated long-term simulations, and 4) incorporating the statistical uncertainty of the precipitation event forcing distributions. We perform steps 1-4 across two meteorological sequences, one with only the precipitation event forcing, and one with random historical weather after the precipitation event to drive a stochastic (ensemble) event simulation framework. We use the analysis of variance (ANOVA) methodology to examine relative contributions of uncertainty to FF estimates across the return periods of interest for factors 1-4 for both meteorological sequences described by factor 5.

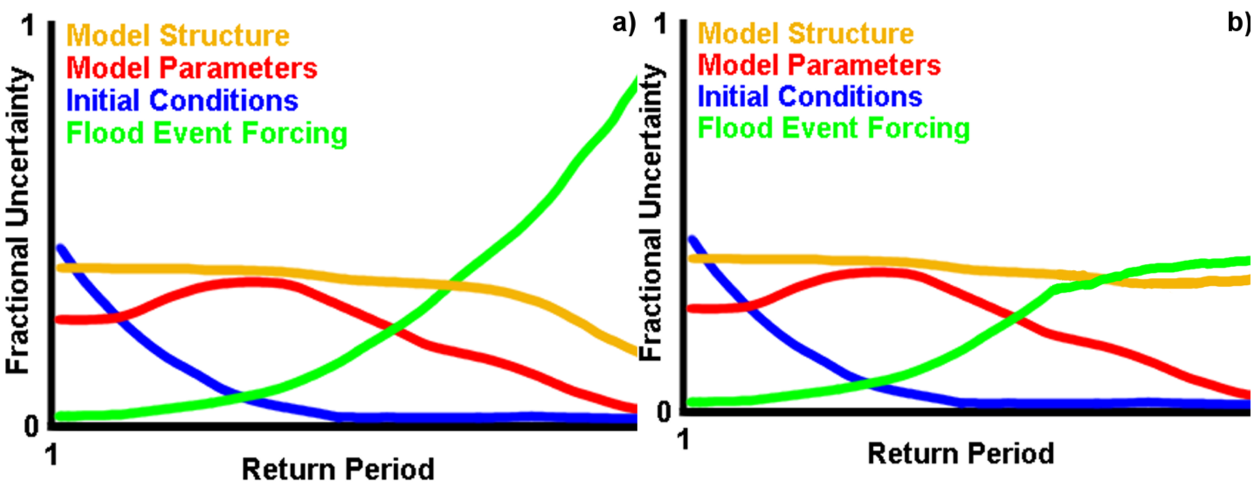


Figure 1. Conceptual contribution of relative uncertainty contribution from initial conditions (blue), model parameters (red), model structure (orange), and precipitation event forcing (green) across return periods (larger return periods towards right) for a) the base case and b) one possible alternative where model structure has similar importance to precipitation event forcing for extreme events.

1.2 Need, Benefit, and Urgency

Reclamation infrastructure requires estimates of probabilistic flood events for risk analysis, design and modification of existing structures (Brekke, 2011). Currently, hydrologic models are chosen at project onset from a small suite of existing models commonly used within the Technical Service

Center (TSC). Those models are then calibrated to existing flow observations through model parameter modification.

The goal of this study is to understand the relative contributions of model choice or structure, model parameters, model initial conditions, and precipitation event forcing to stochastic model flood frequency uncertainty which may lead to more efficient and accurate estimates of flood loadings for Reclamation infrastructure. The study uses two drainage basins representative of hydrologic regions relevant to Reclamation infrastructure. Currently, there is not a substantial effort to choose model structure based on hydrologic region, nor understand the relative contributions to FF uncertainty across the stochastic model components (initial conditions, model parameters and structure, precipitation event forcing and sequencing). As mentioned previously, while the focus of this study is on stochastic modeling, the results and conclusions can help inform model selection, parameterization, and calibration for non-stochastic rainfall runoff modeling efforts as well.

1.3 Study Team

The team for this study includes members from both Reclamation's Technical Service Center and the National Center for Atmospheric Research (NCAR). The lead for TSC was Amanda Stone of the Water Resources Engineering and Management Group. The NCAR lead was Andrew Newman. Additional key team members included Katie Holman of the TSC, and NCAR post-doctoral researcher Manabendra Saharia now at Indian Institute of Technology, Delhi.

2. Study Basins

The Island Park Dam in Idaho and Altus Dam in Oklahoma watersheds are used as representative basins of mountainous snowmelt (Island Park) and semiarid high plains (Altus) hydrologies, respectively. These two types of hydrologies are similar to many Reclamation projects. Island Park (Figure 2) is located on Henry's Fork River approximately 35 miles north of Ashton, Idaho. Island Park Dam impounds a reservoir with a total capacity of 135,500 acre-feet (active 135,200 acre-feet). Water stored at Island Park is used in Madison and Fremont counties in Idaho for irrigation. The Island Park watershed is roughly 501 mi^2 and includes steep mountain slopes along portions of the watershed boundary to nearly level slopes around Henrys Lake. Soils for the watershed range from low permeability clays in the west to permeable volcanic sand in the east. There are areas within the watershed which are heavily forested and other areas which are barren. Elevations within the drainage area range from 6302 feet at the crest of the spillway to 10,600 feet at Sheep Point along the northern boundary of the watershed (Reclamation 2016).

Island Park has a strong seasonal cycle of precipitation, soil moisture, and streamflow with a majority of the watershed precipitation occurring as snow in October through May in the higher elevations. This results in a seasonal snowpack, maximized in late spring which then melts through the summer, maximizing soil moisture and streamflow during late spring and early summer as well.

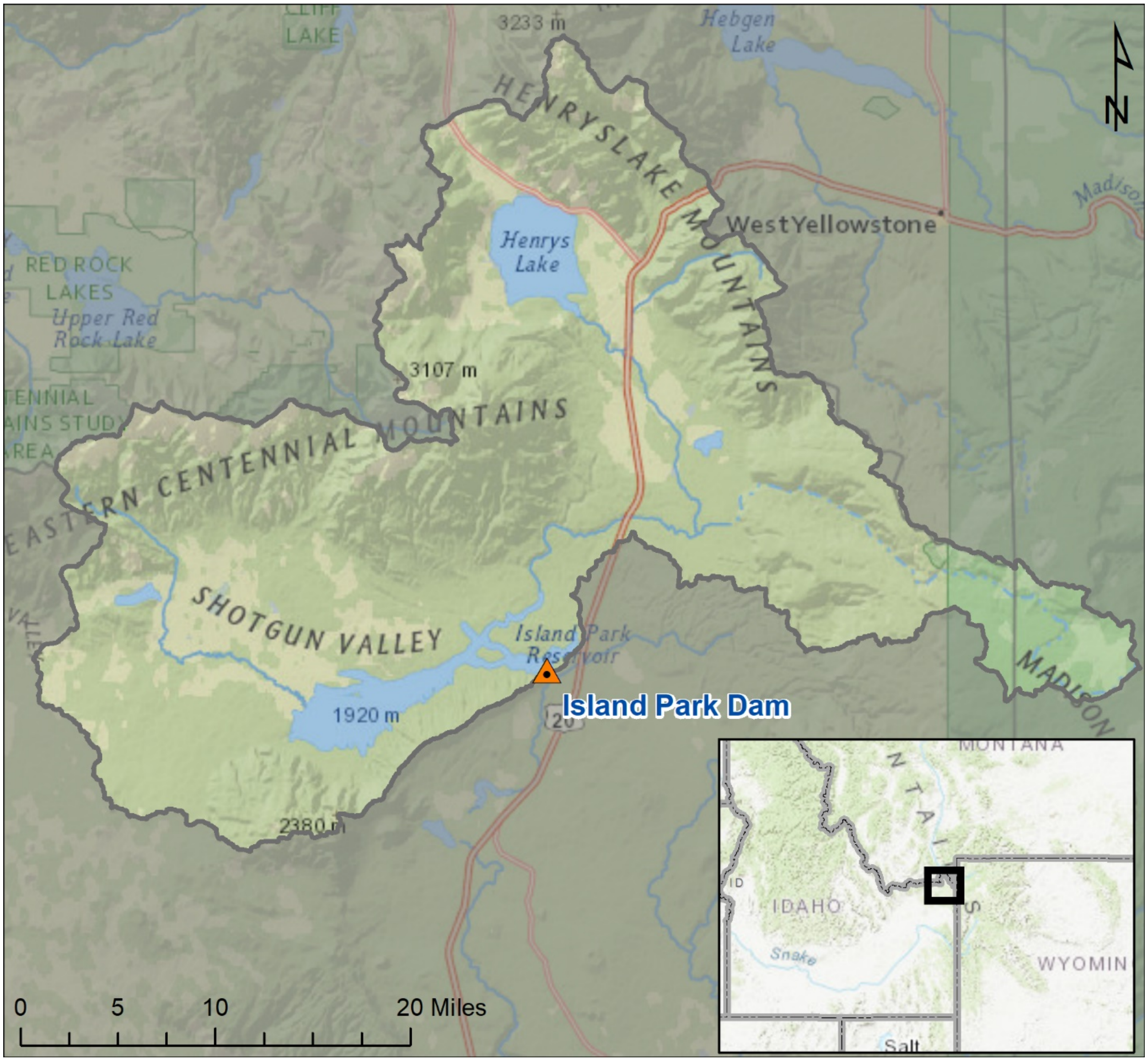


Figure 2. Island Park watershed location

Altus Dam is on the North Fork Red River about 17 miles north of the city of Altus, OK. The purposes of the dam and reservoir are to provide irrigation storage for about 48,000 acres of project lands in southwestern Oklahoma, flood control on the North Fork of the Red River, an augmented municipal water supply for the city of Altus, fish and wildlife conservation benefits, and recreation. The watershed extends from Altus Dam in Oklahoma westward to Amarillo, Texas (Figure 3). The watershed consists of generally rolling terrain with medium to coarse textured soils. This area contains many topographic features known as playa lakes (closed basins with a low area in the center that may see water storage following heavy rainfall) and thus the total contributing area is smaller than the total area of the watershed. We used the Reclamation estimated contributing area of 1951 mi^2. Much of the basin above Altus Dam is devoted to agriculture with a majority of the land cover consisting of cultivated crops, pasture, and hay production. The drainage basin contains no large forested areas, but there are treed riparian zones along the watercourses and trees in cultivated shelterbelts (Reclamation 2012).

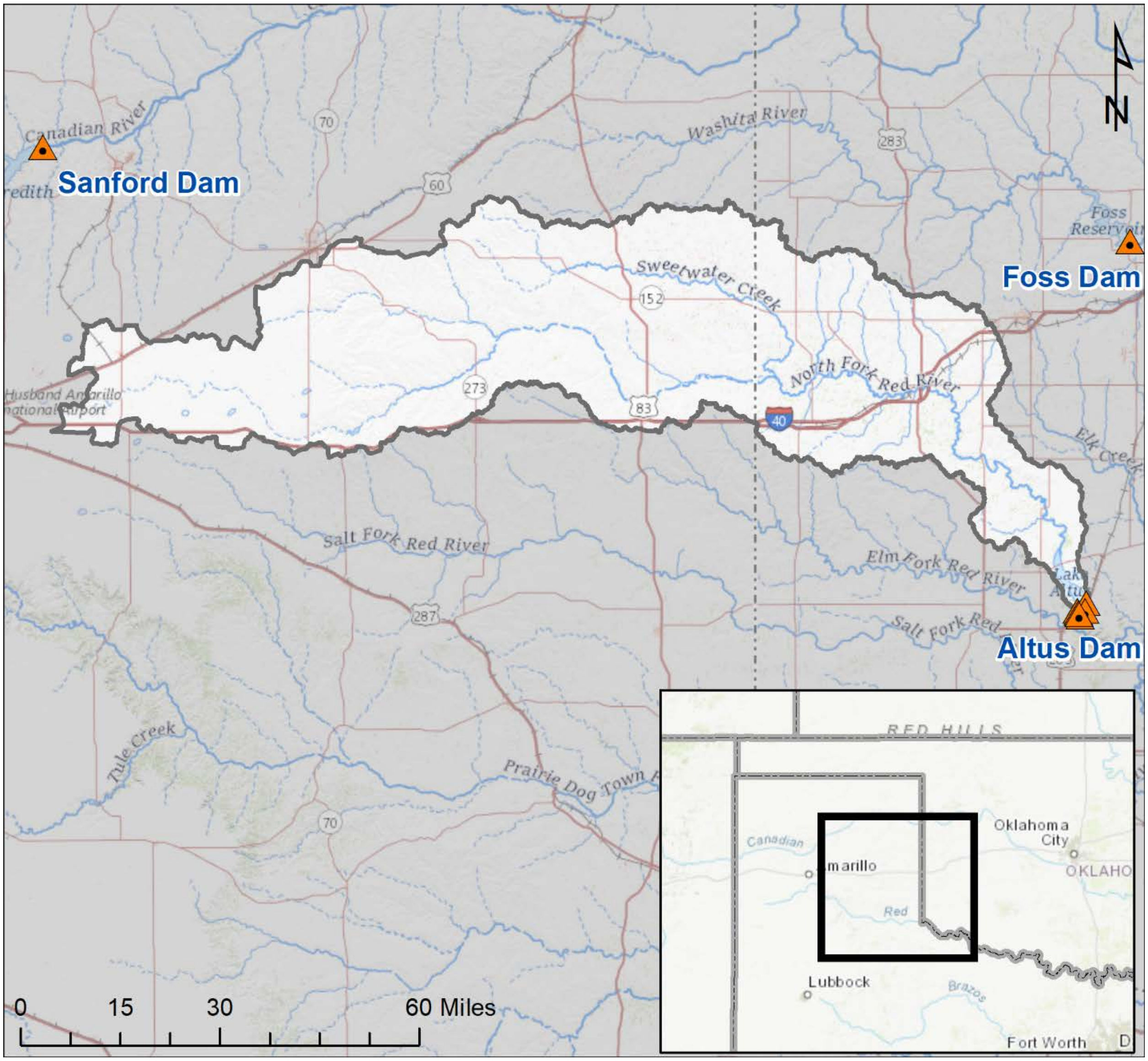


Figure 3. Altus Dam watershed location

Altus Dam is a semi-arid basin that also has a seasonal cycle to precipitation with most occurring in winter through summer, primarily as rainfall. The spring and summer rainfall events are primarily convective in nature with sometimes very intense rainfall rates and high total accumulations over short periods of time, that may coincide with peak basin soil moisture in the spring.

The flood metric of interest at Island Park Dam is the 14-day total runoff volume, similar to Reclamation's 2016 Island Park Dam flood study, while the flood metric of intertest for Altus Dam is the daily peak flow similar to Reclamation's 2012 Altus Dam study.

3. Methods

3.1 Modeling Workflow

This project developed a stochastic hydrologic modeling workflow. This workflow contains the Framework for Understanding Structural Errors (FUSE) hydrologic modeling framework, the Shuffled Complex Evolution (SCE) optimization algorithm, and precipitation frequency distributions from Reclamation. Additionally, we have used the total probability theorem (Nathan, et al., 2003; Micovic, et al., 2015) and the analysis of variance methods to compute the FF estimates and partition the partial uncertainty contributions across the workflow components respectively.

For each basin, hydrologic models are configured and calibrated using an ensemble of historical meteorology. Then, long-term continuous simulations are made to generate spun-up initial conditions for event simulations. Event simulations are then performed across hydrologic models, model parameters, initial conditions, and precipitation frequency distribution estimates for two event sequence possibilities. For each precipitation frequency distribution, we split the probability density function into 50 bins and sample 25 events per bin and perform 2500 model simulations for each possible model-parameter-IC-precipitation frequency combination. This follows the total probability theorem methodology used at Reclamation in their stochastic flood modeling. In total, there are 10 hydrologic models, 11 parameter sets, 4 initial condition sets, and 11 precipitation frequency estimates for Island Park Dam (3 for Altus Dam) for a total of 12.1 million event simulations for Island Park Dam (referred to as Island Park) and 3.3 million event simulations for Altus Dam (referred to as Altus). The different precipitation frequency estimates come from the fact that this project leveraged previously completed studies for these data. We do not believe this will significantly impact the results as the ANOVA analysis takes these sampling differences into account.

3.1.1 Hydrologic Model Framework

The FUSE hydrologic modeling system is a freely available, flexible modular modeling framework written in FORTRAN that allows for the development and testing of many conceptual hydrologic models in a controlled environment. It incorporates multiple parameterizations for many hydrologic fluxes (or processes) at the individual flux level with each equation formulated as a function of the model state, each in a separate code module. This allows the numerical solver to be separated from the flux parameterizations so that every FUSE configuration has the exact same numerics. FUSE also incorporates a conceptual temperature index snowmodel and elevation bands with user specified precipitation and temperature lapse rates to represent seasonal snowpack and changes in meteorology with elevation. Control at the individual flux level with constant numerics is key to understanding how changes in process representation correlates with modeled system behavior. See Clark et al. (2008) for more details regarding FUSE.

FUSE uses several ASCII configuration files where the user can specify the model decisions for process representation, numerical solver parameters, model calibration options, input data, etc. One set of files specifies a unique hydrologic model and simulation configuration. FUSE contains the

SCE optimization algorithm (Duan et al. 1992) for calibration of any hydrologic model the user specifies. SCE is a robust global optimization algorithm that is widely used across the operational and research communities. FUSE uses the network common data format (netCDF) for all input and output data streams (forcing meteorology, any available observations for calibration, calibration information, model simulation states and fluxes) with the same file formats regardless of hydrologic model configuration. Overall, the design of the FUSE system allows for easy configuration, calibration, and simulation of multiple hydrologic models for long term continuous simulations or short event simulations.

FUSE is first used to mimic three widely used hydrologic models: Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) model (Bennett 1998), the Variable Infiltration Capacity (VIC) model (Liang et al. 1994), and the SACramento-Soil Moisture Accounting (SAC-SMA) model (e.g. Anderson 2002) (Table 1). This provides a relatable base set of models to operational groups within and external to Reclamation. Note that the FUSE instantiations of the models only mimic the actual models cited, FUSE does not use the same numerical solver, some process simplifications are made (particularly for VIC where we simplify evapotranspiration), and FUSE does not contain the same coding errors as the original models (see Clark et al. 2008 for FUSE details). We then develop seven other hydrologic models by varying particular processes from the three base models for a total of ten hydrologic models that can be used to compute FF estimates for both basins.

Table 1. FUSE hydrologic processes (far left column) and the various selected process representations for the ten hydrologic models.

FUSE Config.	HECHMS	VIC	SACSMA	MODEL4	MODEL5	MODEL6	MODEL7	MODEL8	MODEL9	MODEL10
rainfall error	multiplc_e	multiplc_e	multiplc_e	multiplc_e	multiplc_e	multiplc_e	multiplc_e	multiplc_e	multiplc_e	multiplc_e
upper-layer architecture	tension1_1	onestate_1	tension1_1	tension2_1	onestate_1	tension2_1	onestate_1	tension1_1	onestate_1	tension1_1
lower-layer architecture and baseflow	unlimfrc_2	fixedsiz_2	tens2pll_2	unlimfrc_2	unlimfrc_2	unlimpow_2	tens2pll_2	tens2pll_2	tens2pll_2	unlimfrc_2
surface runoff	arno_x_vic	arno_x_vic	prms_varnt	arno_x_vic	arno_x_vic	prms_varnt	prms_varnt	prms_varnt	prms_varnt	arno_x_vic
percolation	perc_f2sat	perc_w2sat	perc_lower	perc_f2sat	perc_f2sat	perc_lower	perc_lower	perc_f2sat	perc_w2sat	perc_lower
evaporation	sequential	rootweight	sequential	sequential	sequential	sequential	sequential	sequential	rootweight	sequential
interflow	intflwnone	intflwnone	intflwsome	intflwnone	intflwnone	intflwsome	intflwsome	intflwnone	intflwnone	intflwsome
time delay in runoff	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma	rout_ gamma
snow model	temp_index	temp_index	temp_index	temp_index	temp_index	temp_index	temp_index	temp_index	temp_index	temp_index

3.1.2 Model Calibration

All 10 hydrologic models for both basins were calibrated using the SCE optimization algorithm. Reclamation reconstructed daily inflows for Island Park were used, while Reclamation annual peak flow data was used for Altus due to lack of better available data for calibration at the time of this study. The impact of these different calibration data for the basins will be discussed in Section 4.

The meteorological forcing data consisted of a 100-member ensemble of gridded precipitation and temperature at 6 km resolution which followed the methods described in Newman et al. (2015). Observations of precipitation and temperature and the process of projecting point measurements to grids across sometimes complex terrain are inherently uncertain. This ensemble dataset was designed to estimate those uncertainties and provide many plausible historical traces for hydrologic model applications. Each individual member was used to calibrate each hydrologic model, resulting in a 100-member ensemble of calibrated model parameters for each model for each basin. For Island Park, the hydrologic models were spun up for water years (WY) 1970-1979 and calibrated on WY 1980-2014 (35 WYs), while Altus was spun up for WY 1980-1984 and calibrated on WY 1985-2011 (27 WYs). The specific calibration objective function is discussed in Section 4.1 as an exploration of different calibration objective functions and their performance for flood specific metrics and overall hydrologic model performance.

3.1.3 Model Parameter Specification

The 100 parameter sets available for each model and basin were subsampled for the final FF event simulations. Because Island Park had more available data for calibration, , the final calibrated model performance was very similar across the 100 members for all 10 hydrologic models. Therefore, 11 parameter sets were sampled using the 1st, 10th, 20th, 30th, 40th, 50th, 60th, 70th 80th, 90th and 99th percentiles of calibrated performance.

For Altus, the calibrated model behavior was less constrained due to the much smaller amount of calibration data available. Therefore, the 10 best calibrated parameter sets for each hydrologic model were used, which constrained possible model behavior, but still not to the same level as Island Park (see Section 4.1).

3.1.4 Initial Condition Specification

Continuous simulations using the subsampled parameter sets where then performed and full model states were output each day for the full calibration periods for each hydrologic model and basin. These states were sampled to determine the initial conditions (ICs) for the event simulations. Sampling initial states from continuous simulations has the advantage of providing ICs that are consistent with the specific hydrologic model and parameter set. Applying random perturbations to an IC may result in unrealistic states and subsequent simulation results.

For Island Park, the focus was on ICs from April through June that had minimal (> 10 mm) snow water equivalent snowpack to represent flood events near the end of the snowmelt season around peak climatological soil moisture storage. For Altus, the focus was on late winter through mid-summer ICs (February-July) when both soil moisture and precipitation event intensity and volumes are around their climatological maximums. For both basins and all models, four ICs were sampled in the top 10 percent, 90th, 94th, 97th, and 99th percentiles of total column soil moisture within all validation years and months.

3.1.5 Precipitation Frequency Estimates

Regional frequency analysis (RFA) is a useful method for extending the period of record in environmental datasets by means of a "space-for-time" substitution where additional information in space supplements the lack of information in time. The basic assumption of RFA is that extreme events recorded at stations located within a predetermined homogeneous region can be described by the same probability distribution. By scaling the data by the respective at-site mean (ASM), the user assumes that a single probability distribution is valid for every location within the homogeneous region, while the magnitude can vary spatially.

The L-moments method (Hosking and Wallis 1997) is one example of a regional frequency method. The basis of the L-moments algorithm is that linear combinations of moments can be used to estimate model parameters for extreme value distributions. The moments of interest (also referred to as L-statistics) include L-CV, L-skewness, and L-kurtosis and are computed for every site utilized in an analysis. Regional moments are developed using weighted averages of the site-specific moments, where the weight is proportional to period of record. The regional L-moments are then used to define the regional growth curve (RGC), a dimensionless quantile function that represents the cumulative distribution function of the frequency distribution valid for all sites located within the HR. Site-specific precipitation-frequency estimates ($Q_i(F)$; Equation 1) are developed by scaling the RGC ($q(F)$) by a site-specific ASM (μ_i), allowing the magnitudes of precipitation-frequency estimates to vary spatially across the region of interest.

Equation 1

$$Q_i(F) = \mu_i q(F)$$

Reclamation (2015) developed median and uncertainty precipitation-frequency curves for the Island Park watershed using a regional L-moments approach combined with Latin hypercube resampling procedures. More specifically, the authors used annual maximum two-day precipitation totals from 45 stations in a homogeneous region surrounding the Island Park watershed to estimate parameters of the four-parameter Kappa distribution. The authors used Latin-hypercube sampling methods in R via the "qnorm" function to perform Monte Carlo sampling to create 300 parameter sets using variations in five parameters: at-site mean, regional L-Cv, regional L-skew REF _Ref51156307 \h , regional L-kurtosis, and areal-reduction factor. Results from this analysis include 11 frequency distributions REF _Ref51065407 \h , 14th, 23rd, 32nd, 41st, 50th, 59th, 68th, 77th, 85th, and 95th percentiles. Kappa parameters from Reclamation (2015) are reproduced in Table 2.

Table 2. Parameters used to define the four-parameter Kappa distribution. Table reproduced from Table 4.5 in Reclamation (2015).

Sim	Percentile	xi	alpha	K	H	Basin Mean
1	95th	0.8059	0.02842	-0.068	0.1374	1.66
2	85th	0.8083	0.2827	-0.0635	0.1235	1.64
3	77th	0.8108	0.2812	-0.0590	0.1095	1.63
4	68th	0.8132	0.2798	-0.0546	0.0956	1.61
5	59th	0.8157	0.2783	-0.0501	0.0816	1.6
6	50th	0.818	0.2768	-0.0456	0.0676	1.58
7	41st	0.8188	0.2768	-0.0395	0.0634	1.57
8	32nd	0.8195	0.2768	-0.0334	0.0592	1.55
9	23rd	0.8203	0.2767	-0.0272	0.0549	1.54
10	14th	0.821	0.2767	-0.0211	0.0507	1.52
11	5th	0.8217	0.2767	-0.0430	0.0463	1.51

Similarly, Reclamation (2012) developed precipitation-frequency estimates including median and uncertainty bounds for the Altus watershed using a regional L-moments approach combined with Latin hypercube sampling procedures. The authors focused on annual maximum one-day (or 24-hour) precipitation totals recorded at 482 stations with at least five years of data and used Latin hypercube sampling to produce 150 parameter sets based on variations in the same five parameters listed above: at-site mean, regional L-Cv, regional L-skewness, regional L-kurtosis, and areal-reduction factor. The report provides all precipitation-frequency estimates in the form of fourth-order polynomials, with coefficients reproduced in Table 3.

Table 3. Polynomial coefficients (fourth order) that describe the lower, median, and upper precipitation-frequency curves for Altus. Table reproduced from Table 5.7 in Reclamation (2012).

	A0	A1	A2	A3	A4
Lower Estimate (5%)	0.906821	0.359010	0.031004	0.009728	-0.000563
Median Estimate (50%)	0.999012	0.391658	0.033909	0.013662	-0.000692
Upper Estimate (95%)	1.082307	0.426903	0.04651	0.017021	-0.000828

3.1.6 Event Sequencing

The current event simulation methodology used by Reclamation specifies a precipitation event forcing followed by no precipitation for the remaining simulation time. Other agencies use historical meteorology after the specified flood event input, so for this project we examine both dry and historical meteorology event sequencing after the flood event input. Future work should examine event sequencing in much greater detail, particularly to quantify the impacts of possible future circulation changes on FF estimates and uncertainty.

3.2 Analysis Method

As noted above, the total probability theorem is used to compute modeled basin runoff at return periods of 2, 5, 10, 20, 50 100, 500, 1,000, 5,000, 10,000, 50,000, and 100,000 years from the stochastic simulations for all model, parameter, IC, and precipitation distribution combinations, for both event sequences. An ANOVA analysis is then performed on the runoff values for all the return periods for both event sequences and basins. The ANOVA framework is a relatively simple, computationally frugal, way to estimate individual component contributions to the total variance (or uncertainty) of a variable such as runoff probability. ANOVA is notably relatively robust to violations in the underlying assumptions. The ANOVA framework can also estimate the uncertainty contributions of the interactions between input factors, for example the uncertainty contribution from model-parameter or IC-precipitation event forcing factor interactions. By estimating the fractional (relative) uncertainty contributions of each factor and all two factor interactions the pieces of the modeling workflow which contribute to FF uncertainty can be provided at many return periods of interest to Reclamation.

4. Results

4.1 Calibration Metrics

Hydrologic model calibration is an integral part of the Reclamation stochastic flood modeling methodology. Thus, multiple different calibration metrics were explored at Island Park using streamflow observations to provide calibration metric selection guidance to Reclamation. Two objective function types were used, root mean squared error (RMSE), which is directly related to Nash-Sutcliffe Efficiency (NSE), and the Kling-Gupta Efficiency (KGE). It can be shown that RMSE/NSE is made up of three component contributions to the total value: correlation (r), variability (α), and bias (β). KGE is a metric that contains these same components but is reformulated to weight each component equally by default and allow the user to easily understand their individual contributions to the total KGE value (Gupta et al. 2009) and is shown in Equation 2.

Equation 2

$$\mathrm{KGE} = 1 - \mathrm{ED_s}$$

$$\mathrm{ED_s} = \sqrt{[s_r \cdot (r-1)]^2 + [s_\alpha \cdot (\alpha - 1)]^2 + \left[s_\beta \cdot (\beta - 1)\right]^2}$$

where $\mathrm{ED_s}$ is the scaled Euclidian distance from the ideal point and s_r, s_α, and s_β are scale factors to adjust the weighting of the correlation, variability and bias terms (set to 1 typically). The KGE is also beneficial to use because the scale factors can be adjusted to emphasize the different components of KGE. For this study KGE was examined with the scale factors set to unity and increasing s_α from 1 to 5 to emphasize model flow variance in an effort to better capture flood peaks. Finally, RMSE and KGE calibrations were examined using daily streamflow, three-day smoothed streamflow (as a reconstructed flow noise reduction exploration which will not be discussed further) and peak flow values at yearly intervals.

Figure 4 highlights the results for the different calibration metrics for the validation period annual peak flows at Island Park. For annual peak flows, the interval metrics that specifically calibrated to peak flows perform the best with the two interval KGE options having better performance than interval RMSE. However, the interval KGE metrics result in model overprediction of the largest annual maximums. All calibrations using daily flow metrics under predict the largest annual maximums with daily RMSE resulting in severe underestimation.

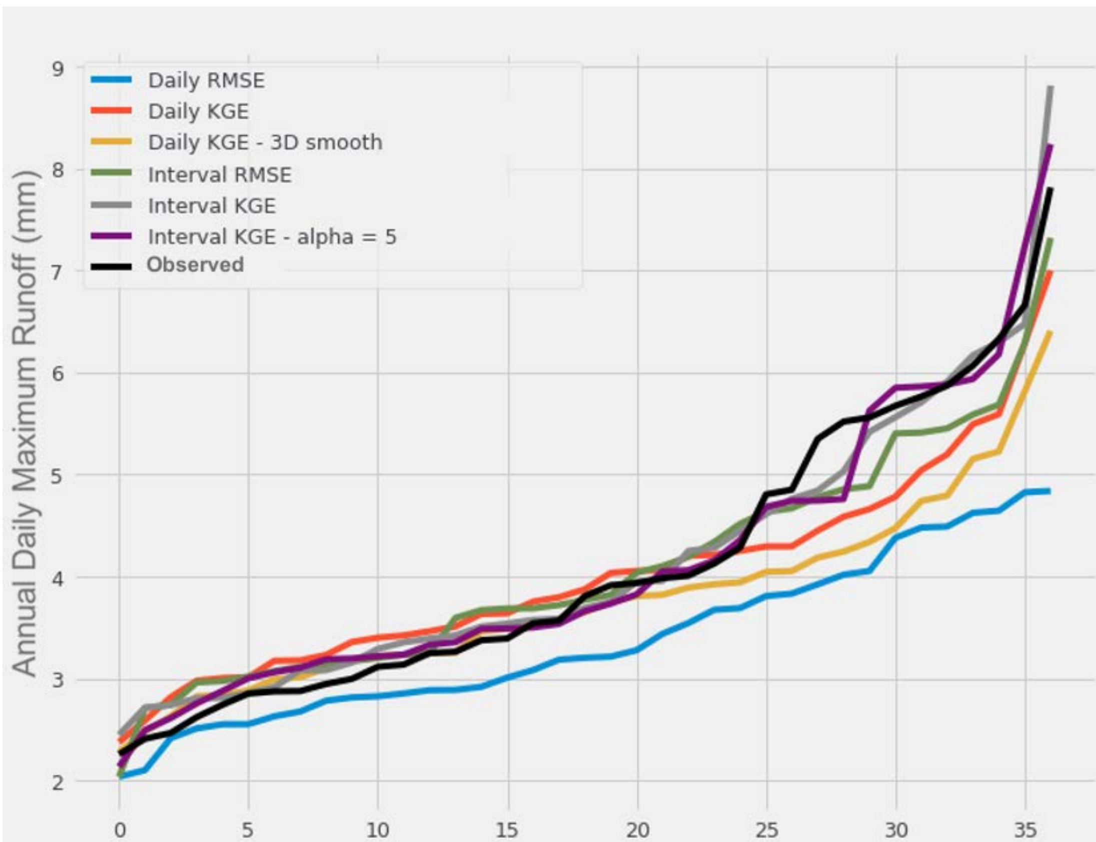


Figure 4. Sorted annual daily maximum runoff for Island Park for the 35 validation WYs for multiple calibration metrics.

When examining daily flow time series, the daily metrics outperform the interval metrics as seen in Figure 5. This is a somewhat expected result as the interval metrics contain no time information (correlation) on the daily scale. Again, the daily KGE metric based calibrations outperform the daily RMSE based calibration, where the daily RMSE based calibration underestimates the flow variance. The interval metric-based calibrations represent the peak flows well (with some overrepresentation) but have large differences in their recession curves with overestimation of flow in the days and weeks immediately following high flow events. This erroneous recession curve representation would result in very different volume-based floods versus daily metric-based calibrations.

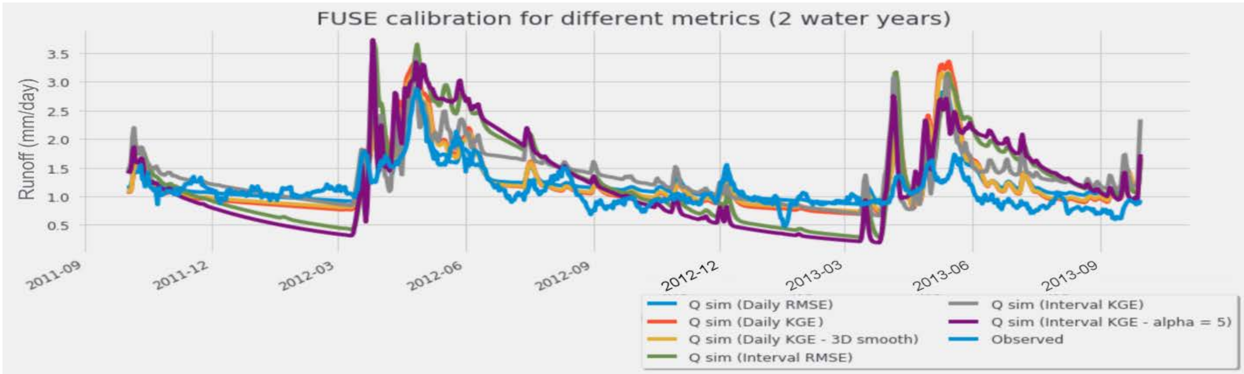


Figure 5. Island Park runoff for two example water years using multiple calibration metrics.

Given the above calibration characteristics and the available calibration data at Island Park (daily flow) and Altus (annual peak flow), daily KGE was selected as the calibration metric for Island Park and interval KGE as the calibration metric for Altus. Daily KGE provides the best all-around simulation when considering daily peak flows as well as volume integrations over days to weeks at Island Park. For Altus, calibrating to yearly peak flows using KGE provided a better overall peak flow calibration than RMSE, likely due to the reformulated weighting of bias and variance as compared to RMSE. These results agree with Mizukami et al. (2019), which examined some of the same calibration metrics using multiple hydrologic models and hundreds of basins across the contiguous United States. They found that KGE outperforms RMSE (or NSE) based calibrations and that peak flow metrics do outperform KGE for peak flow simulation but result in much degraded daily model performance with sometimes severe modeled flow biases.

Figure 6 highlights the final distribution of the calibrated KGE for all ten models for Island Park (a) and one representative example model for Altus (b). Note that Island Park model behavior is much more constrained than Altus (different x-axis ranges from left to right panels). These differences informed the model parameter sampling strategies (Section 3.1.3).

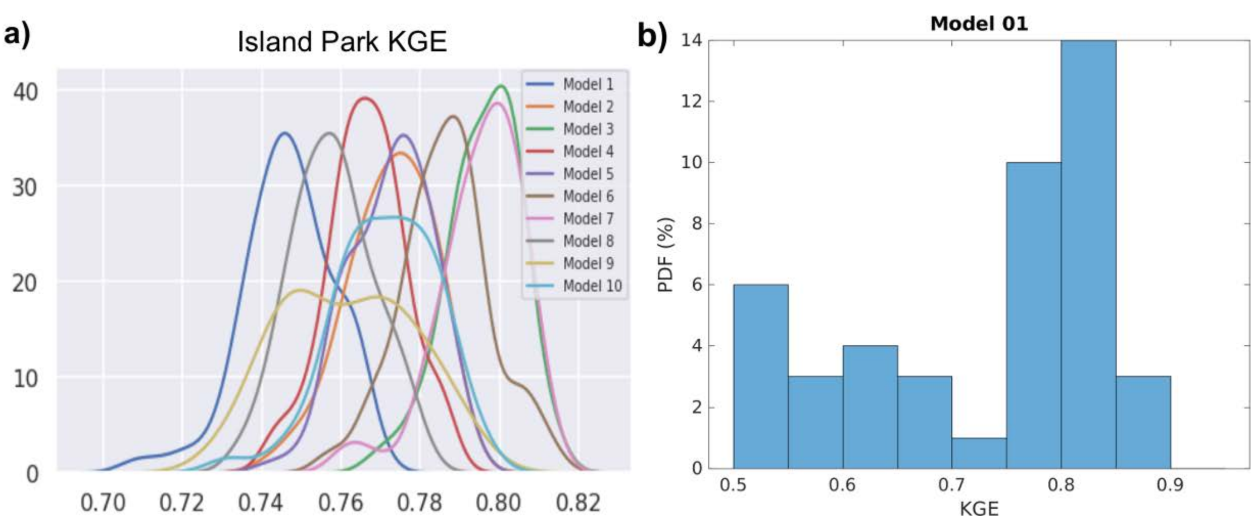


Figure 6. a) Island Park daily flow calibrated KGE distributions for all 10 models and b) Altus yearly peak flow calibrated KGE distribution for an example model.

4.2 Event Simulations

Example FF curves are shown in Figure 7 for Island Park (a) and Altus (b). These are taken from the full set of component combinations for one Model (Model #1), one parameter set (50^{th} percentile of model performance), one IC percentile (99th), for all available precipitation frequency distributions. This behavior is consistent across all combinations, where frequent events have smaller flood flows than less frequent events and precipitation frequency curves specifying larger events across the distribution result in larger floods. This provides a sanity check that our stochastic modeling system is behaving correctly by showing that the system produces increasing flood magnitudes with larger return periods.

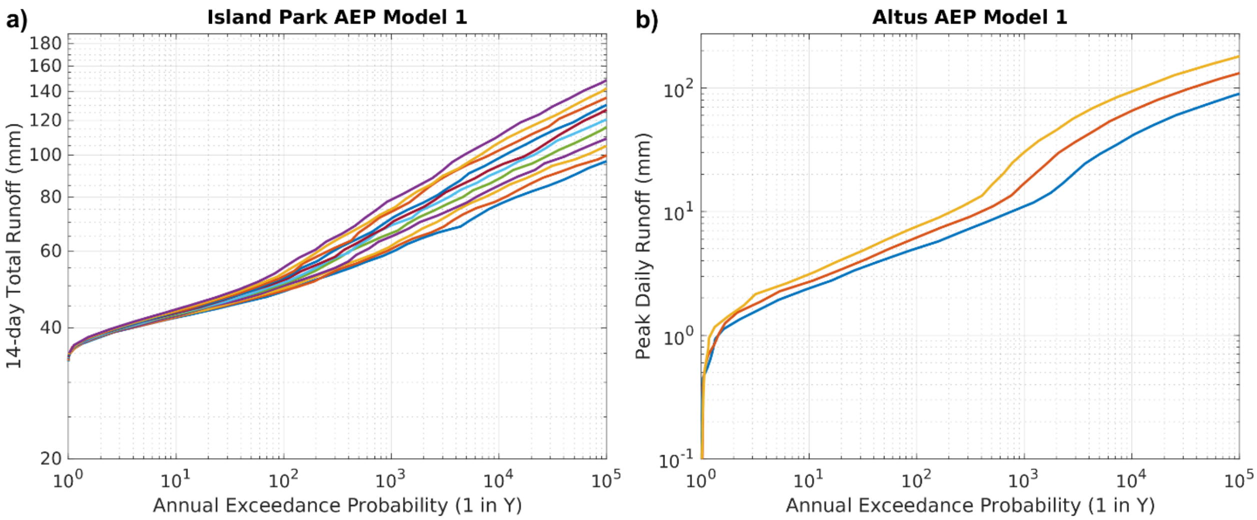


Figure 7. Example FF curves for Model #1, 50th percentile model performance parameters, 99th percentile IC, and all available precipitation frequency distributions for a) Island Park and b) Altus.

To give a sense of the relative spread across the different simulated combinations, normalized FF values are shown in Figure 8. For both basins, there is increasing relative spread with return period, while Island Park (Figure 8a) has more relative spread than Altus at return periods less than a few hundred years (Figure 8b). Specifically, for Island Park, events larger than around 10,000 years have a total relative spread of about a factor of 2.5 while events less than around 100 years have a relative spread of about a factor of 1-1.5. For Altus, infrequent events have little spread with variance growing rapidly for events larger than 1,000 years with events larger than 10,000 years having a range of about a factor of 5-7, or nearly an order of magnitude.

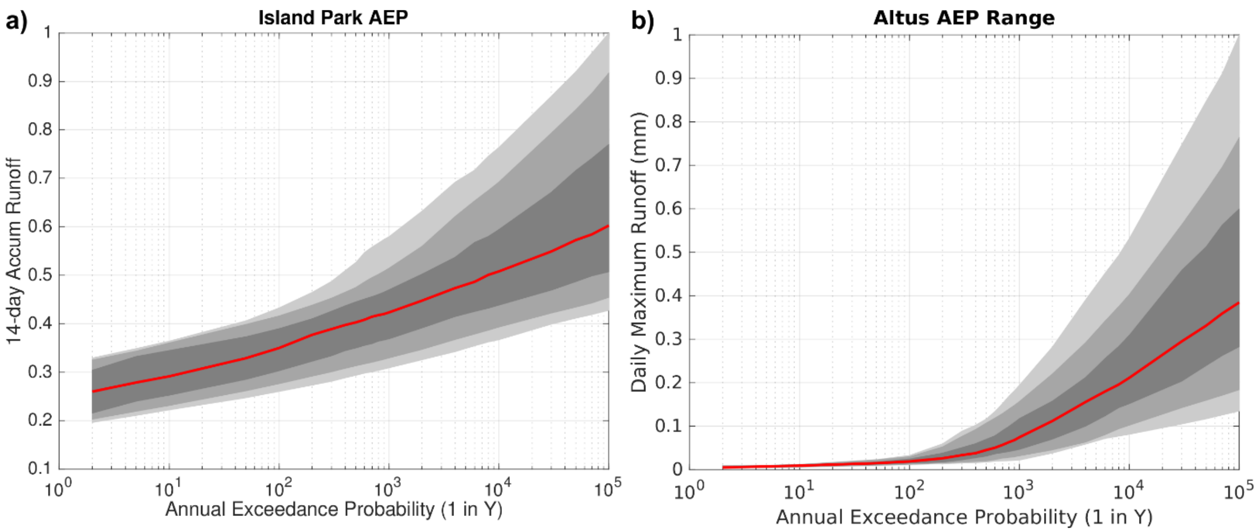


Figure 8. Normalized (by maximum possible flood runoff) FF curves with the median in red, and the interquartile range (25th-75th percentiles) in dark gray, 10th-90th percentile spread in medium gray, and the minimum to maximum spread in light gray for a) Island Park and b) Altus.

4.3 Uncertainty Contributions

The ANOVA analysis was performed following Section 3.3 using the full complement of FF estimates for both basins and precipitation event forcing sequences. All fractional uncertainty contributions are normalized by the total variance in the FF estimate for each return period such that if a component has a fractional uncertainty of 0.5 that component contributes half of the total variance for that return period. The plots represent the 2, 5, 10, 50, 100, 1,000, 10,000, 50,000, and 100,000-year return periods. For Figures Figure 9 through Figure 12, the dry event sequence is always in panel a) and the historical meteorology event sequence is always in panel b), and the color coding follows Figure 1. Interaction terms are a blend of the two primary components (e.g. model structure-model parameter interactions are red-orange).

4.3.1 Island Park

Figure 9 presents the fractional uncertainty contributions for Island Park using the three base models: HEC-HMS, VIC, and SAC-SMA. For this model set, ICs and the precipitation frequency distribution specification dominate for frequent and extreme events, respectively. Model structure is the second most important contributor at most extreme return periods, but it still contributes around 3 times less variance than precipitation frequency distributions for 50,000-100,000 year events. Moving from dry days after the event precipitation to historical meteorology increases the importance of initial conditions across all return periods (compare Figure 9a to Figure 9b). This is somewhat counter intuitive but may be related to the fact that soil states can strongly influence recession curve characteristics and additional non-extreme precipitation event forcing is either stored or released within the 14-day volume integration depending on the IC for these models.

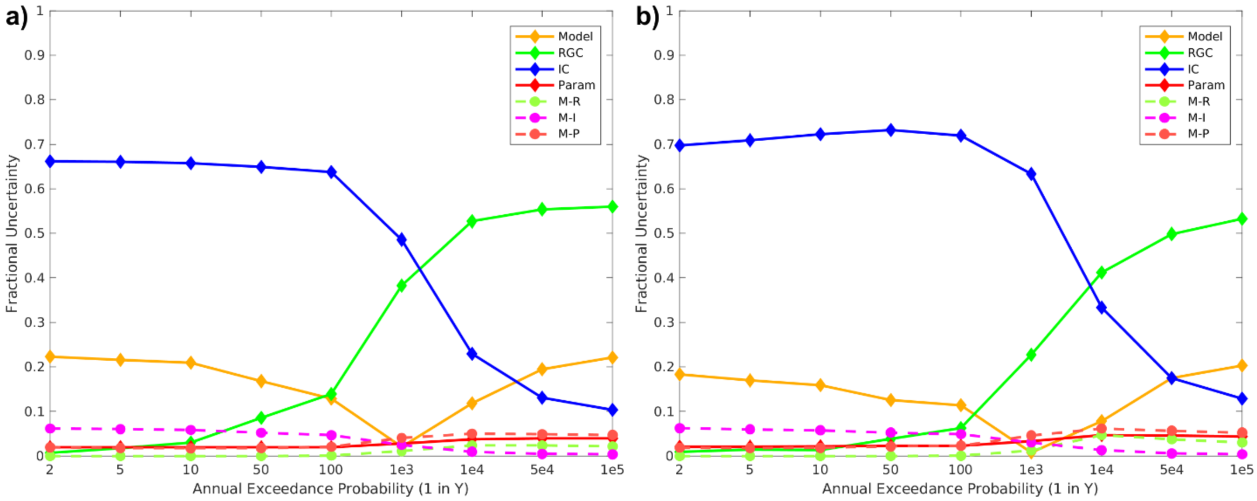


Figure 9. Island Park fractional uncertainty contributions using the three base models: HEC-HMS, VIC, SAC-SMA, for the a) dry event sequence and b) historical meteorology event sequence.

However, using a different combination of the ten possible model structures results in a slightly different conclusion. The set of simulations presented in Figure 10 represents the set of three hydrologic models that generates the largest flood responses to larger precipitation event forcing . Overall, the precipitation frequency distribution specification is still the most important at extreme events, and ICs are most important for very frequent events, but model structure contributes a larger fraction of the total uncertainty across all return periods and is often of similar magnitude to either ICs or precipitation frequency distribution changes (Figure 10). Here we see that moving from dry to historical event sequences increases the importance of model structure (compare Figure 10a to Figure 10b). This is because these three model structures have more variation between each other given additional precipitation input than the variability in runoff changes due to ICs. Differences in surface runoff versus subsurface storage and slower baseflow appear to be driving the model structure variability and is discussed more in Section 5.

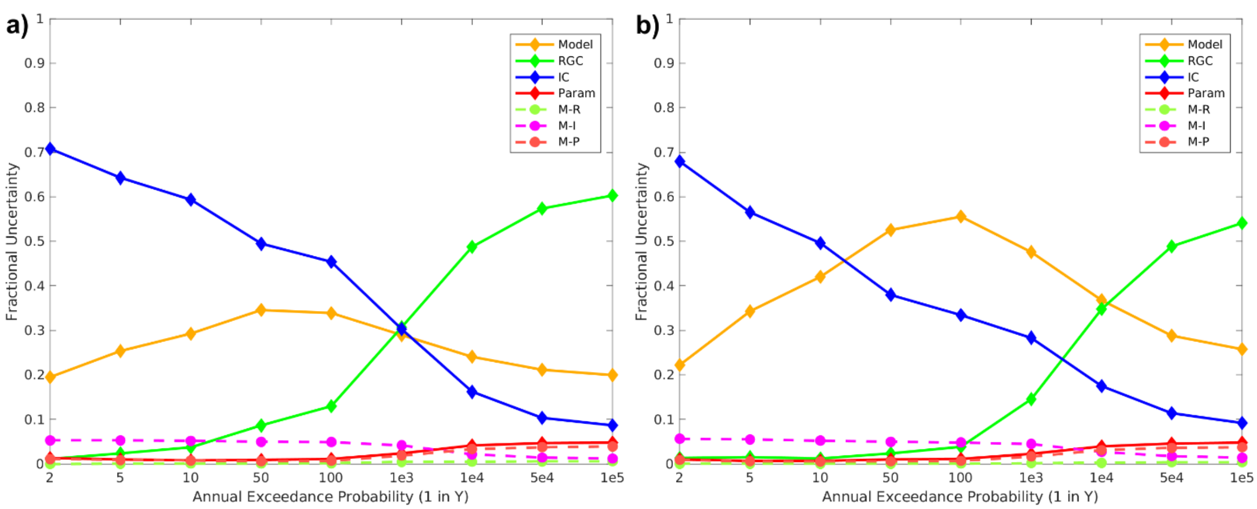


Figure 10. Island Park fractional uncertainty contributions for the three most responsive model structures: HEC-HMS (Model #1), HEC variant (Model #4), and a SAC-SMA/HEC-HMS combination (Model #6), for the a) dry event sequence and b) historical meteorology event sequence.

4.3.2 Altus

The ANOVA results for Altus using the three base models show a similar picture as for Island Park. ICs contribute the most variance for frequent events (less than a few hundred years) and the precipitation frequency distributions are the most important for larger events (Figure 11). However, for Altus the precipitation frequency distributions are even more important than at Island Park as they contribute over 70% of the total variance for 50,000-100,000-year events as compared to around 50% at Island Park. Moving from dry to historical meteorology does not change the picture significantly at Altus (compare Figure 11a to Figure 11b), which is expected as the flood metric is the single day maximum flow and generally single day maximum flow is directly related to the extreme precipitation flood event input and not subsequent small events.

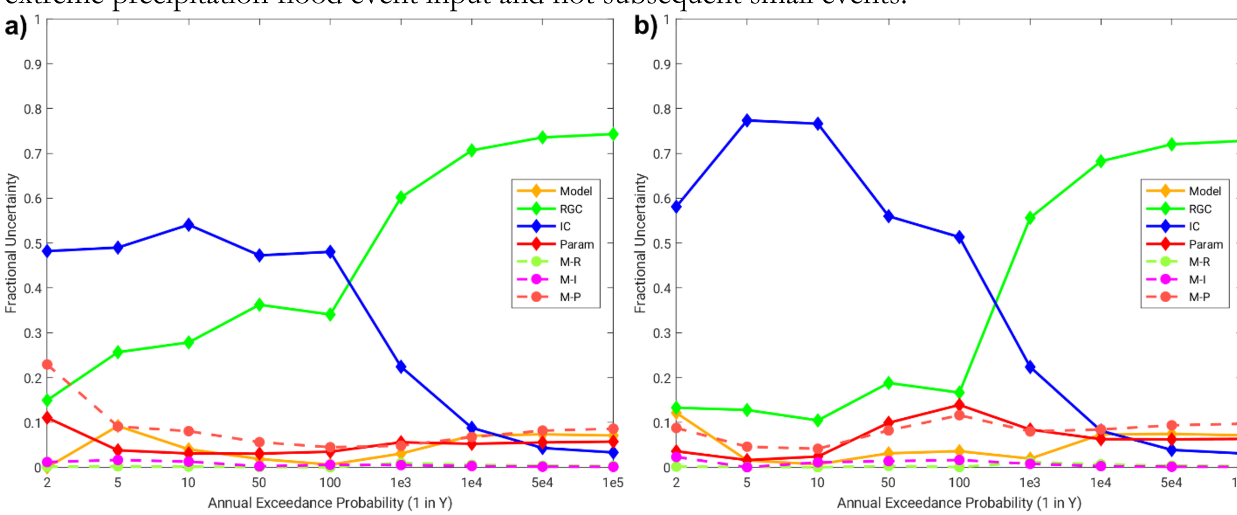


Figure 11. Altus fractional uncertainty contributions using the three base models: HEC-HMS, VIC, SAC-SMA, for the a) dry event sequence and b) historical meteorology event sequence.

Further examination of multiple model combinations at Altus revealed that in nearly all cases the uncertainty contributions in Figure 11 generally hold true (not shown). In the most extreme case, using only the two most disparate model responses, SAC-SMA (Model #3) and the SAC-SMA/HEC-HMS combination (Model #6) models results in substantial increase in importance of model parameters and model parameter – model structure interactions (Figure 12). These two model parameter related components contribute around 30% of the total variance for infrequent and extreme events with return periods greater than 1,000 years. Again, moving from dry to historical meteorology does not substantially change the message here as expected (compare Figure 12a to Figure 12b).

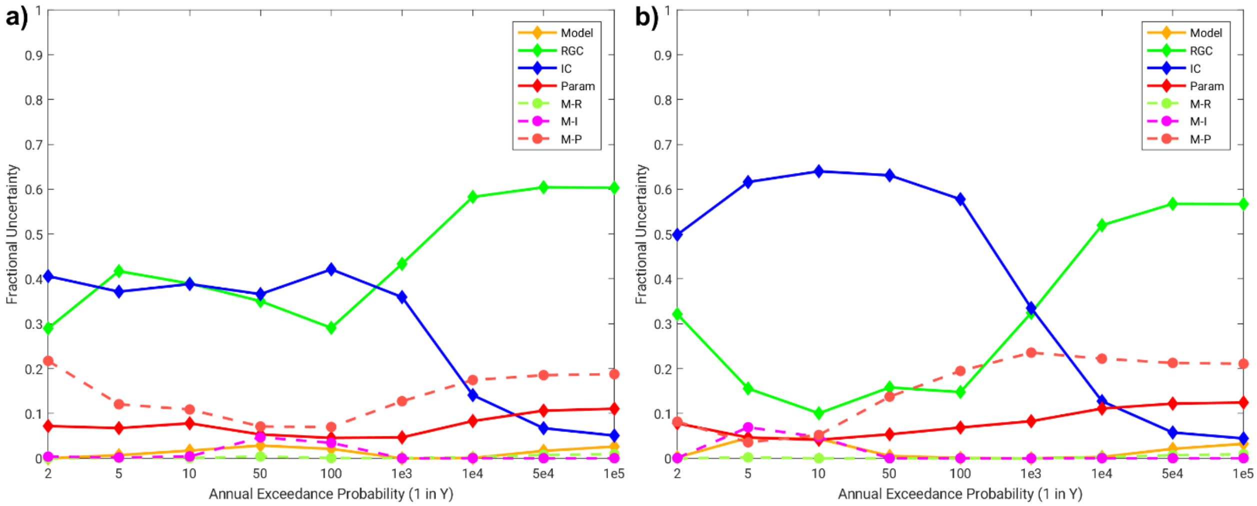


Figure 12. Altus fractional uncertainty contributions for the two most disparate flood responses: SAC-SMA (Model #3), and a SAC-SMA/HEC-HMS combination (Model #6), for the a) dry event sequence and b) historical meteorology event sequence.

5. Discussion

Based on the results of this study careful consideration of the various components of stochastic flood modeling should be undertaken as workflow and methodological decisions impact hydrologic model behavior and the final uncertainty estimates of a FF study. We reaffirm that calibration metrics truly only constrain model behavior for components of the hydrograph most related to the calibration metric (e.g. Mendoza et al. 2015, Mizukami et al. 2019). For streamflow-based calibration, KGE is a robust metric that provides balanced model behavior across all components of the hydrograph because of its formulation (Eq. 2) and should be used over NSE if possible. Furthermore, calibration metrics focusing on high flow only generally result in degraded model performance for other parts of the hydrograph such as the recession curve. In this case, the calibrated hydrologic models may have inferior performance for longer duration volume flood metrics because of substantial biases introduced during calibration that was not designed to constrain flow volumes.

Across the ANOVA uncertainty analysis, in general ICs contribute the most variance for frequent events and the precipitation frequency distribution specification contribute the most variance for extreme events. However, varying the combinations of model structures shows that model structure or model parameters and model structure-parameter interactions have important but still secondary contributions as ICs for frequent or precipitation event forcing for extreme events. According to the study results, the model parameter variations were only important in Altus, where the available calibration data limited the ability for calibration to constrain model performance. This should be taken into account when scoping projects with little calibration data available.

A key difference in the physical response of model structures driving the different flood responses appears to be related to the interplay of surface versus subsurface runoff generation. Models with high event-based runoff ratios activate surface runoff more readily and have smaller subsurface storages, while models with lower event runoff ratios allow for more infiltration and larger subsurface storage. This appears to have a more dramatic impact at Island Park as the responsive models generate larger volumes while the other models essentially store the precipitation input and release it over longer periods of time. This point should be the focus of additional study and provides one physical process comparison to identify the appropriate model structures for a given basin.

While the focus of this study was on stochastic rainfall-runoff modeling for FF studies, there are potentially broader implications to hydrologic modeling for Reclamation. Hydrologic rainfall-runoff modeling is used for a variety of purposes at Reclamation, including planning, design, or restoration often focused on more frequent floods up to extreme events for risk analysis. Stochastic rainfall-runoff modeling is data and labor intensive. Less intensive methods are frequently used, most commonly AEP-neutral assumptions of precipitation return period being equal to flood return period. Even in those studies, model selection, parameterization, initial conditions, calibration, and forcing still play an important role in model outcome. The focus of this study on a range of return periods rather than just extreme floods was intentional to help inform a broader range of studies beyond those focused on risk for large dams. Understanding of uncertainty in rainfall-runoff modeling, whether stochastic or not, is important for flood studies. The results of this study can help guide model selection and development and provide a better understanding of uncertainty in a variety of flood studies.

5.1 Key Findings

The following key generalizable conclusions relevant to Reclamation have resulted from this work:

1) ICs and precipitation frequency distributions a generally contribute the most uncertainty in the stochastic flood modeling chain for frequent and extreme events respectively.

2) Model structure can be equally as important given a diverse set of model responses, particularly for multi-day volume flood metrics. This highlights the need to understand basin flood generation processes and develop methods to select appropriate models. This includes examination of the AEP neutral assumption and selecting model process parameterizations that are most plausible for the study basin.

3) Model parameter and model structure-parameter interactions may be important if the sampled model parameter space is not well constrained by calibration.

4) The Kling-Gupta Efficiency (KGE) is a more robust metric than NSE (or RMSE) for calibration of models related to extreme events and volume integrated floods, is formulated in a manner that permits user understanding of how correlation, variance, and bias contribute to model performance, and is more flexible for application specific uses.

Acknowledgements

We would like to thank Martyn Clark and Andrew Wood of NCAR, Nans Addor of Exeter University, and David Raff, Ken Nowak, and Levi Brekke of Reclamation for their helpful science and methodology discussions throughout this project. We would also like to thank Frank Dworak of Reclamation for his helpful peer review comments that improved this report.

References

Anderson, E. A. (2002). Calibration of conceptual hydrologic models for use in river forecasting. Office of Hydrologic Development, US National Weather Service, Silver Spring, MD.

Bennett, T. H. (1998). Development and application of a continuous soil moisture accounting algorithm for the Hydrologic Engineering Center Hydrologic Modeling System (HEC-HUMS). University of California, Davis.

Brekke, L. D. (2011). Addressing climate change in long-term water resources planning and management: User needs for improving tools and information. DIANE Publishing.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., ... & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44(12).

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. Journal of optimization theory and applications, 76(3), 501-521.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.

Liang, X., Wood, E. F., & Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. Global and Planetary Change, 13(1-4), 195-206.

Markstrom, S. L., Hay, L. E., & Clark, M. P. (2016). Towards simplification of hydrologic modeling: identification of dominant processes. Hydrology & Earth System Sciences, 20(11).

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., ... & Arnold, J. R. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. Journal of Hydrometeorology, 16(2), 762-780.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models.

Micovic, Z., Hartford, D., Schaefer, M., & Barker, B. (2015). Flood hazard for dam safety – where should the focus be? 25th ICOLD Congress, Stavanger, Norway, June 17-19, 2015.

Nathan, R., Weinmann, E., & Hill, P. (2003). Use of Monte Carlo simulation to estimate the expected probability of large to extreme floods. The Institute of Engineers Australia, 28th International Hydrology and Water Resources Symposium, Wollongong, NSW, November 10-14, 2003

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., ... & Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. Journal of Hydrometeorology, 16(6), 2481-2500.

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B., & Nearing, G. (2017). Benchmarking of a physically based hydrologic model. Journal of Hydrometeorology, 18(8), 2215-2225.

Reclamation (Bureau of Reclamation), 2012. Altus Dam Hydrologic Hazard and Reservoir Routing for Corrective Action Study. W.C. Austin Project, OK. Great Plains Region. 230 pp.

Reclamation (Bureau of Reclamation) 2015. Island Park Dam Meteorology for Application in Hydrologic Hazard Analysis. Minidoka Project, ID. Pacific Northwest Region. 88 pp.