

FINAL REPORT

Next-Generation Intensity-Duration-Frequency Curves Considering Spatiotemporal Nonstationarity in Climate, Intense Precipitation Events, and Snowmelt

SERDP Project RC-2546

February 2, 2020

Mark Wigmosta

Pacific Northwest National Laboratory

PNNL-30974

Page Intentionally Left Blank

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

Page Intentionally Left Blank

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 02-02-2021		2. REPORT TYPE SERDP Final Report		3. DATES COVERED (From - To) March 2016 - January 2021
4. TITLE AND SUBTITLE Next-Generation Intensity-Duration-Frequency Curves Considering Spatiotemporal Nonstationarity in Climate, Intense Precipitation Events, and Snowmelt		5a. CONTRACT NUMBER N/A		
		5b. GRANT NUMBER N/A		
		5c. PROGRAM ELEMENT NUMBER N/A		
6. AUTHOR(S) Mark Wigmosta Hongxiang Yan ~~Ning Sun~~ Ruby Leung		5d. PROJECT NUMBER RC-2546		
		5e. TASK NUMBER N/A		
		5f. WORK UNIT NUMBER N/A		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Pacific Northwest National Laboratory 902 Battelle Boulevard Richland, WA 99352 USA		8. PERFORMING ORGANIZATION REPORT NUMBER RC-2546		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Strategic Environmental Research and Development Program 4800 Mark Center Drive, Suite 16F16 Alexandria, VA 22350-3605		10. SPONSOR/MONITOR'S ACRONYM(S) SERDP		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) RC-2546		
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution A: Approved for public release; Distribution is unlimited				
13. SUPPLEMENTARY NOTES None				
14. ABSTRACT Current hydrologic design of infrastructure relies on historical precipitation-based intensity-duration-frequency (PREC-IDF) curves. The PREC-IDF curves ignore snow processes by assuming that precipitation is in the form of rain, immediately available for the rainfall-runoff process. In snow-dominated regions where snowmelt and rain-on-snow (ROS) events are the dominant flood-generating mechanisms, the PREC-IDF design approach can lead to substantial underestimation or overestimation of design floods and associated infrastructure. Given the potential impacts of climate change, the historical PREC-IDF curves further increase infrastructure development risks because structures designed to meet traditional criteria could be underdesigned under future climate conditions. The overarching research objective of this project is to develop a new hydrologic design tool—the next-generation IDF (NG-IDF) curve—that characterizes both snow process and nonstationary climate conditions, to better inform decisions regarding Department of Defense (DoD) long-lived infrastructure.				
15. SUBJECT TERMS NG-IDF Curves, Hydrologic Design, Design Flood, NG-IDF Curves, Hydrologic Design, Design Flood, DHSVM, SNOTEL, Snow-dominated Regions, CONUS, Canopy Gap, Snow Hydrology, Infrastructure Design, DoD Site, Climate Change, Uncertainty, Underdesign, Surface Water Design Manual Update				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	83
				19a. NAME OF RESPONSIBLE PERSON Mark Wigmosta
				19b. TELEPHONE NUMBER (include area code) 509 372-6238

Page Intentionally Left Blank

Table of Contents

List of Tables	iii
List of Figures.....	iv
List of Acronyms	vi
Abstract.....	1
1. Objectives	2
2. Background.....	4
2.1 Importance of Snow	4
2.2 Hydrologic Design in Snow-dominated Environments	5
2.3 Hydrologic Design under Non-stationarity	7
3. Materials and Methods	8
3.1 Objective 1: NG-IDF Proof-of-Concept using SNOTEL Observation.....	8
3.2 Objective 2: DHSVM Snow Modeling and Validation	11
3.3 Objective 3: Validation of NG-IDF Tool in Design Flood Estimates.....	14
3.4 Objective 4: NG-IDF Products under Historical Climate	18
3.5 Objective 5: Modeling Extreme Precipitation in both Historical and Future Climate...	18
3.6 Objective 6: Ensemble NG-IDF Products under Future Climate.....	20
4. Results and Discussion	21
4.1 NG-IDF Proof-of-Concept using SNOTEL Observation.....	21
4.2 DHSVM Snow Modeling and Validation	29
4.3 Validation of NG-IDF Tool in Design Flood Estimates	34
4.4 NG-IDF Products over CONUS under Historical Climate	39
4.5 Impacts of Atmospheric Rivers on Extreme Events	44

4.6	Project Climate Change Impacts on NG-IDF Products	49
5.	Conclusions and Implications for Future Research/Implementation	55
6.	Literature Cited.....	60
	Appendix A. Supporting Data.....	65
	Appendix A.1. BCQC SNOTEL Data Set.....	65
	Appendix B. List of Scientific/Technical Publications.....	67
	Appendix B.1. Articles in Peer-Reviewed Journals.....	67
	Appendix B.2. Technical Reports.....	68
	Appendix B.3. Conference Symposium Proceedings.....	68
	Appendix B.4. Conference or Symposium Abstracts	68
	Appendix B.5. Test Book or Book Chapters	70

List of Tables

Table 1.	Research objectives and associated hypotheses addressed in this project.....	4
Table 2.	Dominant drivers of extreme events by percent of events over CONUS for historical and future climate conditions.....	53
Table 3.	The basin mean annual maximum W depth for the 1-, 3-, 6-, 12-, 24-, 48-, and 72-hour durations.....	54
Table 4.	Variability/uncertainty of future changes in the NG-IDF curve values for 10- and 50-year return periods and durations of 1, 3, 6, 12, 24, 48, and 72 hours.....	55
Table 5.	Research objectives and summary of results	56

List of Figures

Figure 1.	Flowchart of the six objectives to be accomplished over the life of the project.....	3
Figure 2.	An illustration of a rain-on-snow flooding event that occurs in mountainous regions.....	5
Figure 3.	Flowchart of using SNOTEL observation to examine the impact of snow processes on IDF curves and estimates of flood frequency for hydrologic design	9
Figure 4.	Grid representation of the canopy gap in the original and enhanced DHSVM	12
Figure 5.	Illustration of flowchart for regional snow parameters development.....	13
Figure 6.	Flowchart for NG-IDF tool validation using the DHSVM continuous simulation method. DHSVM was applied for both hillslope and basin scales for NG-IDF validation	15
Figure 7.	The experimental hillslope and the soil characteristics used in the DHSVM modeling	16
Figure 8.	The location of the small test basin that flows into the DoD Marine Corps Mountain Warfare Training Center	17
Figure 9.	Left panel: examples of three design cases in terms of IDF values.....	22
Figure 10.	PREC-IDF and NG-IDF curves at the selected SNOTEL (45.21, -117.19) in Cascades, Oregon.....	23
Figure 11.	Histogram of the actual associated return periods in NG-IDF curves for a 24-hour 100-year event in PREC-IDF curves, and the associated over/underestimation of risk of failure	24
Figure 12.	Relative errors between peak design flood estimates from the NG-IDF curves and PREC-IDF curves for the four selected return periods.....	26
Figure 13.	Histogram of the over/underestimation (%) of peak floods for four return periods over all SNOTELs in the western United States.....	27
Figure 14.	Regional trends at the 5% significance level for the snowpack, precipitation, temperature, and ROS events.....	29
Figure 15.	Enhanced DHSVM SWE modeling skill in complex forested environments including open, forest, and canopy gap conditions.....	29
Figure 16.	DHSVM SWE modeling skill with the best-fit parameter set for the 246 SNOTELs across the western United States	31
Figure 17.	Illustration of DHSVM-simulated and observed SWE for selected SNOTEL stations across different hydroclimatic regions with the value of NSE close to one.....	32
Figure 18.	Validation of DHSVM in SWE modeling skill with regional snow parameters	33
Figure 19.	Scatterplots of the normalized peak design floods from PREC-IDF curves and NG-IDF curves versus normalized design floods from DHSVM continuous simulation for the 246 SNOTEL stations across the western United States	34
Figure 20.	Relative peak design flood errors between the estimates from PREC-IDF curves and NG-IDF curves versus the DHSVM continuous simulations.....	36
Figure 21.	DHSVM calibration and validation results in streamflow and SWE for the Upper West Walker Basin.....	37
Figure 22.	Probabilistic W hyetograph for the small test basin	38

Figure 23.	Comparison of design flood estimates derived from DHSVM and NG-IDF technology with the use of different hyetographs.....	39
Figure 24.	The CONUS is divided into 5 spatial clusters based on the mean precipitation, air temperature, and wind speed during the general snow accumulation season from November through March over 1950–2013.....	40
Figure 25.	Four regional DHSVM snow parameters over the CONUS.....	41
Figure 26.	NG-IDF curve values for the selected duration and return periods over CONUS under historical climate condition from 1950–2013	42
Figure 27.	Dominant drivers of extreme events over CONUS under historical climate conditions from 1950–2013	43
Figure 28.	Ratio of NG-IDF curve values to PREC-IDF curve values for the selected durations and return periods over CONUS under historical climate conditions from 1950 to 2013.	44
Figure 29.	Correlation coefficients between AR intensity-duration-area and the monthly total of extreme daily precipitation amount.....	46
Figure 30.	Gilbert Skill Scores of predicting the number of extreme precipitation days using ARs.....	47
Figure 31.	The difference in surface water budget associated with AR and non-AR precipitation events	48
Figure 32.	Illustration of the surface hydrological response to non-AR and AR precipitation events with and without pre-existing snowpack.....	49
Figure 33.	Regional variability of future projections of extreme P and W events over CONUS using the data from the 10 selected statistically downscaled GCMs under RCP8.5.	50
Figure 34.	Ensemble mean ratio of the NG-IDF curve values to the PREC-IDF curve values for historical and future climate conditions over CONUS using the data for the 10 selected statistically downscaled GCMs under RCP8.5	51
Figure 35.	Dominant drivers of extreme events from historical to future periods over CONUS using the data for the 10 selected statistically downscaled GCMs under RCP8.5	52
Figure 36.	Basin-scale annual maximum W depth for historical and future climate conditions over the small test DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5.....	53
Figure 37.	NG-IDF curves for historical and future climate conditions over the DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5.	54
Figure 38.	Two examples of extreme W events for historical and future climate conditions over the small test DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5.....	55
Figure 39.	An example of a cyberinfrastructure for distributing probabilistic NG-IDF curves for a DoD site in Alaska with explicit and complete quantification of uncertainty for end users.....	60

List of Acronyms

AR – atmospheric River
AR6 – Sixth Assessment Report
ARI – average recurrence interval
ARF – Areal Reduction Factor
CDF – Cumulative Density Function
CMIP5 – Coupled Model Intercomparison Project Phase 5
CONUS – Contiguous United States
DHSVM – Distributed Hydrology Soil Vegetation Model
DoD – Department of Defense
ET – evapotranspiration
GCM – general circulation model
GEV – generalized extreme value
GHG – greenhouse gas
HPC – high-performance computing
HUC – Hydrologic Unit Code
IDA – Intensity-Duration-Area
IDF – Intensity-Duration-Frequency
IPCC – Intergovernmental Panel on Climate Change
KGE – Kling-Gupta efficiency
LID – Low Impact Development
MACA – Multivariate Adaptive Constructed Analog
MC – Monte Carlo
MCMWTC – Marine Corps Mountain Warfare Training Center
NARCCAP – North American Regional Climate Change Assessment Program
NG-IDF – Next-Generation Intensity-Duration-Frequency
NOAA – National Oceanic and Atmospheric Administration
NRCS – Natural Resources Conservation Service
NSE – Nash-Sutcliffe efficiency
P – Precipitation
PNNL – Pacific Northwest National Laboratory
PREC-IDF – Precipitation-based Intensity-Duration-Frequency
RCM – Regional Climate Model
RCP – Representative Concentration Pathway
RKT – regional Kendall test
ROS – Rain-on-Snow
SMMEW – Stormwater Management Manual for Eastern Washington
SNOTEL – Snow Telemetry
SWE – snow water equivalent
TR-55 – Technical Release 55
USGS – United States Geological Survey
UWWB – Upper West Walker Basin
W – water available for runoff
WRF – Weather Research and Forecasting

Keywords

NG-IDF Curves, Hydrologic Design, Design Flood, DHSVM, SNOTEL, Snow-dominated Regions, CONUS, Canopy Gap, Snow Hydrology, Infrastructure Design, DoD Site, Climate Change, Uncertainty, Underdesign, Surface Water Design Manual Update

Acknowledgements

The authors would like to acknowledge the support from the Strategic Environmental Research and Development Program of the U.S. Department of Defense.

Abstract

Introduction and Objectives: Current hydrologic design of infrastructure relies on historical precipitation-based intensity-duration-frequency (PREC-IDF) curves. The PREC-IDF curves ignore snow processes by assuming that precipitation is in the form of rain, immediately available for the rainfall-runoff process. In snow-dominated regions where snowmelt and rain-on-snow (ROS) events are the dominant flood-generating mechanisms, the PREC-IDF design approach can lead to substantial underestimation or overestimation of design floods and associated infrastructure. Given the potential impacts of climate change, the historical PREC-IDF curves further increase infrastructure development risks because structures designed to meet traditional criteria could be underdesigned under future climate conditions. The overarching research objective of this project is to develop a new hydrologic design tool—the next-generation IDF (NG-IDF) curve—that characterizes both snow process and nonstationary climate conditions, to better inform decisions regarding Department of Defense (DoD) long-lived infrastructure.

Technical Approach: This project used physics-based modeling and data analysis to develop and demonstrate a scientifically defensible methodology for creating NG-IDF curves that directly consider intense precipitation events and snowmelt under conditions of climate nonstationarity. Under this project, we (1) *proposed* NG-IDF curves using a mass balance approach, (2) *demonstrated* NG-IDF curve benefits in estimating design floods using Snow Telemetry (SNOTEL) observations, (3) *enhanced* the Distributed Hydrology Soil and Vegetation Model (DHSVM) for modeling snow-vegetation interaction, (4) *developed* DHSVM regional snow parameters for large-domain hydrological applications, (5) *extended* NG-IDF curves from at-site observations to the Contiguous United States (CONUS) using well-validated DHSVM simulations, and (6) *explored* both ensemble statistically and dynamically downscaled general circulation model (GCM) data for constructing NG-IDF curves under nonstationary climate conditions.

Results: Our research has developed a new, robust hydrologic design tool—NG-IDF curves—to enhance hydrologic design and provide a consistent IDF design approach for both rainfall- and snow-dominated regions. Using SNOTEL observations, we have shown that PREC-IDF curves can substantially underestimate design floods by up to 324%, leading to a potentially high risk of infrastructure failure in snow dominated locations. A key challenge in wide adoption of the NG-IDF tool by practical engineers is the limited availability of observed snow data in space and over time. To overcome this limitation, we developed spatially continuous NG-IDF curves at 6 km resolution across the CONUS, using the well-validated DHSVM. We validated the NG-IDF tool at both hillslope and basin scales and demonstrated its accuracy and robustness in estimating design floods. We used both ensemble statistically and dynamically downscaled GCM data to develop NG-IDF curves for future climate conditions. We archived NG-IDF products for both historical and future climates, and they can be directly used as input to a single event hydrologic model to produce flood estimates.

Benefits: This project directly addressed DoD infrastructure planning and management decisions to update hydrologic design to assist in the management of design risks encountered under changing climate conditions. In particular, this project developed improved understanding of and

responses to changes in the timing and intensity of rainfall- and snowmelt-based runoff events, and provided a new, robust hydrologic design tool—NG-IDF curves—for potential use by practicing engineers and planners.

1. Objectives

Department of Defense (DoD) military installations and their supporting infrastructure cover nearly 30 million acres, and many of them are located in cold and snow-dominated environments prone to snowmelt or rain-on-snow (ROS) flooding events. A significant amount of infrastructure in the United States is currently designed using precipitation-based intensity-duration-frequency (PREC-IDF) curves, such as the National Oceanic and Atmospheric Administration's (NOAA's) Atlas 14 (Bonnin et al., 2011), which are developed based on historical precipitation observations and assume that the occurrence of precipitation is in the form of rainfall and immediately available for the rainfall-runoff process. In snow-dominated environments (e.g., the western United States), where most precipitation falls as snow in the cold season and melts during spring to early summer, the use of standard PREC-IDF curves can lead to substantial underestimation of design floods and a high risk of the failure of infrastructure, because it has been designed to only withstand flooding related to the rainfall-runoff process. Further, long-lived infrastructure design with the use of historical PREC-IDF curves has been done explicitly under the assumption of stationarity (e.g., hydrologic systems for a location behave consistently within a given time frame). Given climate and landscape changes, the probability of the occurrence of typical and extreme flood events most likely contains temporal trends. Failure to account for the impact of hydrologic trends substantially increases the risks associated with infrastructure designed and developed without considering this impact, because structures designed to meet traditional criteria (e.g., 10-year 1-hour storm or 100-year 6-hour storm) could be underdesigned. Within the DoD, Low Impact Development (LID) goals have been established to manage site development and design for stormwater runoff drainage and flood protection. To avoid the excess capital and operational costs associated with LID and related design conditions, new methods are needed to account for the snow process and historical temporal trends, as well as those caused by future altered climates.

The overarching research objective of this project is to develop a new, robust hydrologic design tool—the next-generation IDF (NG-IDF) curve—that characterizes both the snow process and nonstationary climate, to better inform decisions regarding DoD long-lived infrastructure. In particular, given the lack of coordinated and consistent surface water design methods for snow-dominated regions of the United States (Yan et al., 2019b), we aim to provide a consistent IDF-based design approach—the NG-IDF curves—for both rain- and snow-dominated regions. The NG-IDF tool is expected to benefit both practical engineers and planners along with DoD and jurisdictions responsible for developing unified facility criteria and approving subsequent infrastructure designs.

To achieve our primary research objective of better informing decisions regarding DoD long-lived infrastructure, we outlined six objectives to be accomplished over the life of the project. Figure 1 presents the flowchart for the six objectives that are described in detail below. The accomplishments under each objective provided the basis for the work under the next objective.

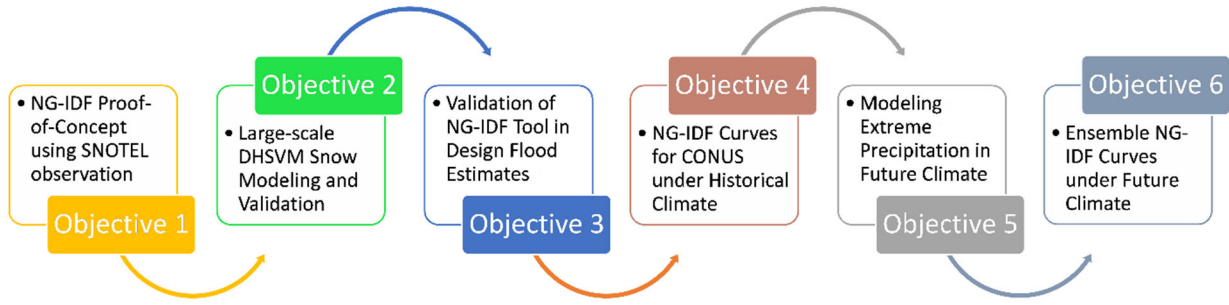


Figure 1. Flowchart of the six objectives to be accomplished over the life of the project.

- Objective 1: Propose NG-IDF curves and examine the limitations of the use of standard PREC-IDF curves in design flood estimates in snow-dominated regions using Snow Telemetry (SNOTEL) measurements as a proof-of-concept.
- Objective 2: Advance snow-vegetation interaction modeling in complex forest environments and develop regional snow parameters for large domain hydrological applications, laying out the scientific basis for developing NG-IDF curves across the Contiguous United States (CONUS).
- Objective 3: Evaluate the NG-IDF tool in design flood estimates based on the well-validated Distributed Hydrology Soil Vegetation Model (DHSVM) continuous simulation at both hillslope and basin scales (containing a DoD site) to investigate the accuracy and robustness of NG-IDF curves in design flood estimates using a standard single event-based rainfall-runoff model.
- Objective 4: Develop NG-IDF products under historical climate conditions across CONUS at 6 km spatial resolution using the well-validated DHSVM to provide NG-IDF products that can be used in standard engineering design.
- Objective 5: Model extreme precipitation under historical and future climate conditions using the Weather Research and Forecasting (WRF) model for dynamical downscaling applied to a multi-model general circulation model (GCM) ensemble to project future climate and provide future meteorological data for developing NG-IDF curves under nonstationarity.
- Objective 6: Develop ensemble NG-IDF curves with uncertainty quantification for future climate conditions under nonstationarity, using both WRF dynamically downscaling data and available statistically downscaling data.

Table 1 summarizes the objectives and related hypotheses for research Objectives 1–6 throughout the life of the project.

Table 1. Research objectives and associated hypotheses addressed during this project.

Objective	Description	Task
1	NG-IDF Proof-of-Concept using SNOTEL Observation	I. Compare estimates of extreme water available for runoff between NG-IDF and PREC-IDF curves in snow-dominated regions. II. With the use of a single event rainfall-runoff model, compare estimates of design floods between NG-IDF and PREC-IDF curves in snow-dominated regions. III. Determine if observed extreme water available for runoff events shows temporal trends under a warming climate.
2	DHSVM Snow Modeling and Validation	I. Enhance DHSVM to improve representation of snow-canopy interactions, including open, forest, and gaps. II. Develop and test regional snow parameters for DHSVM.
3	Validation of NG-IDF Tool in Design Flood Estimates	I. Compare estimates of flood peaks based on NG-IDF curves with estimates based DHSVM continuous simulation at both hillslope and basin scales. II. Examine the sensitivity of design flood peaks to the selection of the hyetograph of water available for runoff.
4	NG-IDF Products under Historical Climate	I. Determine the dominant flood-generating mechanisms across the CONUS. II. Examine the spatial variability of NG-IDF curves across different climate regions.
5	Modeling Extreme Precipitation in Future Climate	I. Examine the impacts of climate change on storm characteristics including storm frequency and intensity, storm size and storm track.
6	Ensemble NG-IDF Products under Future Climate	I. Examine the impacts on design storm events of the use of historical NG-IDF curves under a changing climate. II. Examine the uncertainty in NG-IDF curves under future climate conditions due to GCM uncertainty.

2. Background

2.1 Importance of Snow

Globally, about 1.2 billion people (1/6 of the world's population) rely on snowmelt water for agriculture and human consumption (Barnett et al., 2005). In the western United States, mountain snowmelt contributes more than 50% of the total annual runoff and more than 60 million people rely on snowmelt for electrical power and agriculture (Bales et al., 2006). As a result, many civil infrastructure systems, including the DoD's (e.g., military training base, energy, transportation, water resources, medical services) to support U.S. national security

interests, are located in cold environments prone to snowmelt or ROS flooding events. Another recent example is the California ROS flood in February 2017 that led to the Oroville Dam Crisis, which resulted in the evacuation of about 190,000 people. Figure 2 illustrates a typical flood event caused by ROS in mountainous regions.

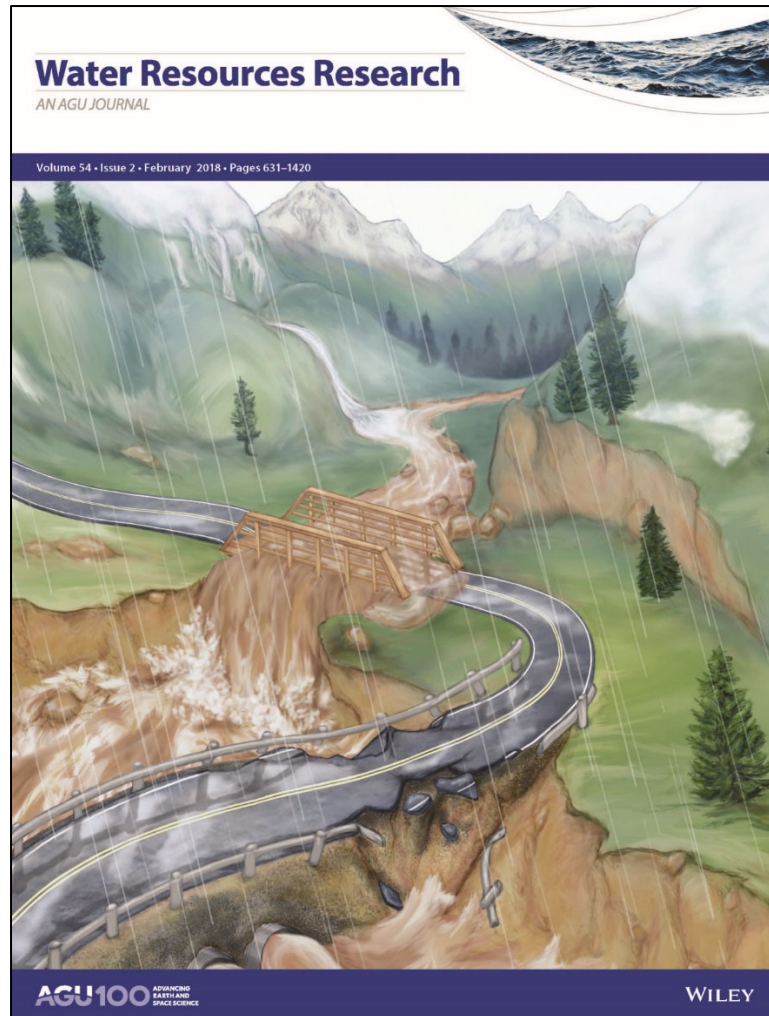


Figure 2. An illustration of a rain-on-snow flooding event that occurs in mountainous regions. Figure designed by Rose Perry at Pacific Northwest National Laboratory (PNNL) was featured on the American Geophysical Union-Water Resources Research journal front cover (see Yan et al., 2018).

2.2 Hydrologic Design in Snow-dominated Environments

The use of PREC-IDF curves in hydrologic infrastructure design has been a standard practice for more than 50 years (Chow, 1964). However, this method can lead to overdesign or underdesign in snow-dominated locations and both results come with costs (underdesign is more dangerous). These costs could be avoided if more sophisticated (yet economically viable and easily implementable) methods such as those we propose were used instead.

PREC-IDF curves are developed by fitting extreme precipitation data with an extreme value statistical distribution. A precipitation depth with a fixed duration and return period or average

recurrence interval (ARI), suggested or required in local surface water design manuals, can then be obtained from the PREC-IDF curves; for example, a 24-hour, 100-year return period or ARI design flood corresponding to a 1% exceedance probability. The associated design flood magnitude or volume is then estimated using a single event-based rainfall-runoff model, such as the Natural Resources Conservation Service (NRCS) Technical Release 55 (TR-55) (Cronshey et al., 1986). The impact of current or proposed land cover and soil condition is represented by the user-assigned curve number, which is also a function of assumed antecedent rainfall. This type of design approach implicitly assumes that all precipitation is in the form of rainfall that is immediately available to begin the rainfall-runoff process. As such, it misrepresents precipitation that falls as snow and is detained in the snowpack without generating immediate runoff. On the other hand, it also ignores antecedent snowpack and snowmelt that may exceed or supplement rainfall-based runoff. In the United States, NOAA Atlas 14 (Bonnin et al., 2011) provides nationwide PREC-IDF curves, which are commonly used by water resources and storm management practitioners. The appeal of these publicly available products for engineers who have limited resources is easy to understand, but some PREC-IDF products are outdated and were generated in 1973 (e.g., Washington State).

In the United States systematic, coordinated, and consistent local surface water design manuals for snow-dominated regions are generally lacking (Yan et al., 2019b). In most cases, the local surface design manuals use the “blind approach” of simply using the PREC-IDF curves and ignoring the snowmelt process. For example, Snohomish County in Washington State extends from the Puget Sound to the Cascade Mountains and features hydroclimatic conditions that range from rain-dominant to snow-dominant. The Snohomish County Drainage Manual recommends the use of NOAA PREC-IDF curves for infrastructure design of facilities such as wetpool treatment facilities. In other cases, local surface water design manuals provide a “tuning factor approach” of adding a snowmelt tuning factor to the PREC-IDF values. For example, Chelan County in Washington State, which includes snow-dominated watersheds on the eastern Cascade Mountains, adopts the tuning factor approach suggested in the Stormwater Management Manual for Eastern Washington (SMMEW) published in 2004. Data to support this approach are provided for nine cities and are based on the assumption that the observed December–February average daily snow depth will melt during a 72-hour ROS event. Reductions are made to estimate melt for a 24-hour time period, which is then added to the 24-hour design storm precipitation estimate for that location. The Washington State Department of Ecology updated the SMMEW in 2019, but the section of addressing snowmelt was untouched.

In snow-dominated regions, the use of the simple blind approach may result in the high risk of infrastructure failure if snowmelt or ROS intensity exceeds precipitation intensity. The snowmelt tuning factor approach can only be used for limited locations (i.e., in cases for which there are no data for ungauged basins) and suffers from several implicit assumptions (e.g., the timing and intensity of melt), which may lead to large errors in the snowmelt calculation. Over the past several decades, considerable research has been conducted in the hydrologic science community using physics-based hydrologic models to simulate the snow process and overcome the deficiency of PREC-IDF design in snow-dominated regions. However, as Yan et al. (2018) pointed out, a large portion of infrastructure in cold environments is still designed using the traditional PREC-IDF technology, including the NOAA PREC-IDF curves. There are two main reasons for this: (1) the use of a sophisticated physics-based hydrologic modeling approach can

be cost-prohibitive in the design of smaller infrastructure, such as highway culverts or facility storm water systems; and (2) the adherence to local surface water design manuals that may require or recommend the use of NOAA PREC-IDF curves with a selected single-event rainfall-runoff model. As a result, there is an emerging need for a consistent and scientifically defensible approach for hydrologic design in snow-dominated regions, which can benefit both practicing engineers and local jurisdictions responsible for developing design manuals and approving subsequent infrastructure designs. Ideally, this approach should be adapted to the current standard design procedure and require minimum extra training for practicing engineers.

2.3 Hydrologic Design under Non-stationarity

Global climate is projected to warm as a result of increasing concentrations of greenhouse gases (GHGs). Since the pre-industrial era, the global mean temperature has increased by 0.75°C, and regional variations have been dominated by land-sea and meridional contrasts. Over land, mountains are expected to experience greater warming than lowlands because of the moist adiabatic structure of the atmosphere (the so-called lapse-rate effect) and snow albedo feedback. As the water-holding capacity of the atmosphere increases by roughly 7% per degree Celsius of warming, precipitation is expected to increase in conjunction with this warming. Climate models consistently simulated a lower rate of increase in global mean precipitation of about 2% per degree Celsius of warming, compared to extreme precipitation, which increases at a rate from 5% to 10% for each degree Celsius of warming globally (Sugiyama et al., 2010).

The use of PREC-IDF curves in current infrastructure design is based on historical precipitation measurements. These curves are based on the concept of temporal stationarity, which assumes that the probability of extreme precipitation events occurring is not expected to change significantly during the lifespan of the designed infrastructure (e.g., 50 years). However, as discussed earlier, climate change is expected to alter the intensity, duration, and frequency of precipitation extremes over time. For example, considering that a DoD's infrastructure has a lifespan of at least 30 years, the planned infrastructure will still be used in 2050 when changes in extreme precipitation events will have far more significant impacts than they do today (Hou et al., 2019; Ren et al., 2019). Thus, DoD requests that climate nonstationarity be considered in long-lived infrastructure design to avoid potential underdesign and future damages (ESTCP, 2018).

Inclusion of snow processes in hydrologic design approaches vastly increases the complexity of the problem over traditional rainfall-only-based approaches. Changes in the frequency, intensity, and duration of snowmelt and/or ROS events respond to local changes in precipitation, air temperature, incoming short- and longwave radiation, wind, relative humidity, and vapor pressure, as well as antecedent snowpack conditions including pack surface albedo, temperature, and snow water equivalent (SWE) (Yan et al., 2019a, 2019b). Vegetation canopy interception and release of snowfall are significantly more complicated than they are for rainfall. Overstory vegetation may significantly affect below-canopy water and energy balance and directly affect the magnitude and timing of snow accumulation and melt. Thus, relative to traditional event-based rainfall-runoff design, the incorporation of nonstationarity in snow processes requires consideration of the complete water and energy balance, a more detailed representation of canopy processes, and explicit representation of antecedent conditions.

3. Materials and Methods

Technical methods and data for this project are described in detail and separately for each of the six objectives below.

3.1 Objective 1: NG-IDF Proof-of-Concept Using SNOTEL Observation

Unlike PREC-IDF curves, which are developed using precipitation data without discriminating the precipitation phase, the proposed NG-IDF curves are developed based on the water available for runoff (W), which characterizes the actual water reaching the land surface that results from the combined effects of rainfall, snowmelt, and/or ROS. The W is obtained through mass balance as

$$W = P - \Delta SWE \quad (1)$$

where P is precipitation and ΔSWE is the change in snowpack SWE. The advantage of this approach is that it uses the change in SWE instead of an air temperature threshold to differentiate between rainfall and snowfall. The drawbacks of the air temperature threshold approach have been discussed in a large body of literature by the hydrologic science community (Fassnacht and Records, 2015; McCabe et al., 2007), yet NOAA Atlas 14 (Perica et al., 2013) still uses a single temperature threshold (1.1°C) method to determine precipitation phase.

In this project, we first used SNOTEL measurements as a proof-of-concept for demonstrating the NG-IDF tool in hydrologic design, because at-site observations provide the most accurate SWE data, and even the most sophisticated hydrologic models cannot replace the observation record (Milly et al., 2008). In particular, we used SNOTEL P and SWE measurements to construct both NG-IDF and PREC-IDF curves and then compared their design flood estimates using the NRCS TR-55 single-event rainfall-runoff model at the SNOTEL point scale. Figure 3 presents a flowchart indicating how SNOTEL observations were used to compare the performance of PREC-IDF and NG-IDF curves in design flood estimates.

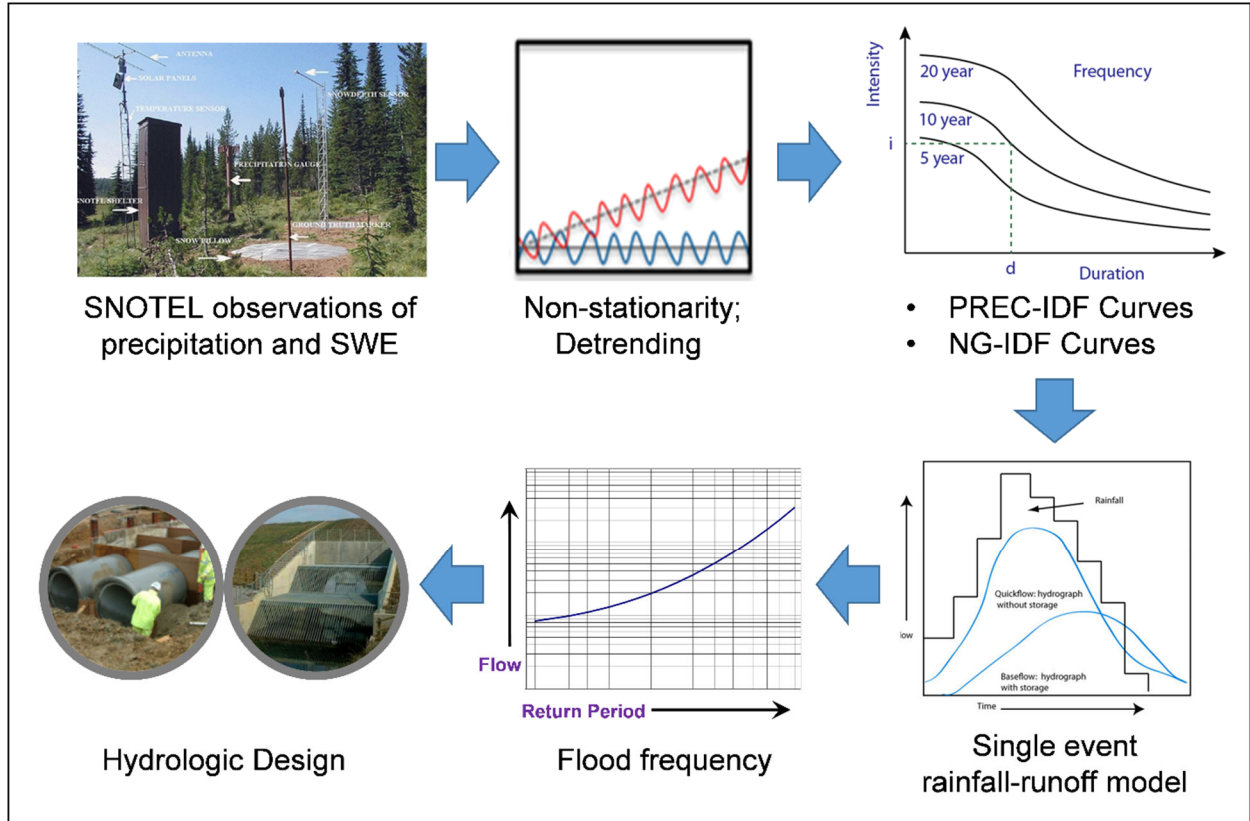


Figure 3. Flowchart of using SNOTEL observation to examine the impact of snow processes on IDF curves and estimates of flood frequency for hydrologic design.

We acquired long-term daily meteorological (precipitation and air temperature) and SWE measurements from 805 active SNOTEL stations across the western United States. We developed a rigorous three-stage SNOTEL quality control filter (Yan et al., 2018) to identify missing data, erroneous values, outliers, and inconsistent data. After completing the quality control process, screened SNOTEL data were subsequently bias corrected for snowfall undercatch and temperature bias (Sun et al., 2019). The resulting data set is referred to as bias-corrected quality-controlled (BCQC) SNOTEL data and can be accessed at <https://dhsvm.pnnl.gov/>. For the details about the BCQC methodology, readers are referred to Yan et al. (2018) and Sun et al. (2019).

Using the BCQC SNOTEL data, both PREC-IDF and NG-IDF curves were developed for SNOTEL stations that each had at least 30 years of high-quality P and W records. For each selected SNOTEL, we constructed two data sets consisting of the annual maximum for P and W (estimated from eqn. 1) for durations from 1 to 30 days. Before conducting a frequency analysis, we first tested stationary assumptions for the annual maximum series of P and W using the nonparametric Mann-Kendall test (Kendall, 1975; Mann, 1945). Where trends were statistically significant ($p < 5\%$), we detrended the annual maximum time series using Sen's slope (Sen, 1968) while maintaining the time series average. After detrending, following the NOAA Atlas 14 approach (Bonnin et al., 2011), the generalized extreme value distribution (GEV) was fitted to each annual maximum data set using L-moments statistics (Hosking and Wallis, 1997). The

PREC-IDF and NG-IDF curves were then developed for the selected exceedance probabilities. For example, the 0.02 exceedance probability corresponds to extreme events that have a return period or ARI of 50 years. The direct comparisons between the PREC-IDF and NG-IDF curves illustrated differences between extreme precipitation and extreme water available for runoff throughout the western United States.

In the TR-55 modeling, the NRCS curve number approach is used to convert a selected design storm depth (from IDF curves) into a runoff depth by taking into account watershed characteristics and initial moisture conditions. The NRCS unit hydrograph method is then used to convert the runoff depth to a design flood hydrograph where peak flood magnitude and flood volume can be obtained. In addition to determining the IDF curves, it is also necessary to determine the design storm hyetograph (time distribution of P or W) and watershed characteristic (e.g., land cover, soil type). The variety of design storm hyetographs proposed in the literature include the uniform hyetograph proposed in the rational method, the triangular hyetograph, the Chicago hyetograph, and the NRCS hyetographs (McCuen, 1998). Of these hyetographs, the uniform hyetograph is the most widely used, probably because of its simplicity. However, all of the standard design storm hyetographs were developed only for rainfall and no hyetograph has been considered for W. Hence, for this objective of the NG-IDF proof-of-concept, we chose the uniform hyetograph to model the rainfall-runoff process for both PREC-IDF and NG-IDF curves. The W hyetograph shape and its impacts on design flood estimates are discussed in Section 3.3. A synthetic watershed unit was developed and used in the TR-55 event-based model at each SNOTEL to isolate the impact of climatic conditions on IDF design and peak flow estimates. The observed P and W were assumed to be uniformly distributed over the whole watershed unit and therefore no areal reduction factor is necessary. The synthetic watershed with a unit area of one had homogeneous bare soil cover (sandy loam) approximating SNOTEL open area conditions, leading to the same curve number in the rainfall-runoff modeling at all SNOTEL sites. The time of concentration for the synthetic watershed unit was set to 24 hours to be consistent with the temporal resolution (daily) of the SNOTEL observation and the corresponding IDF curves. To estimate the potential peak flood, the antecedent moisture condition was set to the wet condition across all selected SNOTEL stations and the critical design duration approach was used; specifically, the duration (varying from 1 to 30 days) that produced the largest flood peak was considered as the critical design duration.

Based on the BCQC SNOTEL data, we also provided a regional trend analyses for SWE and extreme W events across the mountainous western United States. These results are important and can provide the directly observed trends or snow change signals in the historical period, because even the most sophisticated hydrologic models cannot replace the observation record. The regional Kendall test (RKT) (Helsel and Frans, 2006) was used in this project to estimate the temporal trends for a region by grouping data from different SNOTEL sites within the region. Compared to the commonly used simple linear regression, the RKT can relax the Gaussian error assumption and is more resistant to outliers. Further, the RKT trades “space for time” and thus provides stronger statistical power for detecting trends. This is particularly important for short record trend detection, such as the BCQC SNOTEL records (Yan et al., 2019c). In our RKT analysis, the western United States was clustered into 11 regions based on the boundaries of the Commission for Environmental Cooperation Ecological Regions of North America, Level III (CEC, 2009).

3.2 Objective 2: DHSVM Snow Modeling and Validation

Despite the development of NG-IDF curves using SNOTEL measurements under Objective 1, practical application of NG-IDF curves is limited by snow data availability. The SNOTEL measurements are limited in both spatial coverage and temporal extent, and it is unlikely that the location of infrastructure is coincident with these SNOTEL sites. Given these circumstances, Hamlet (2018) advocated that a well-validated hydrologic model is key for extending snow data in space and time to make NG-IDF curves useful in engineering applications, and the NG-IDF curves produced under Objective 1 are an important source for model validation.

Hamlet (2018): *“To my mind, it is here that Yan et al. (2018) make their most substantial contribution by providing a new data set in mountain environments, based on observations, that can be used to validate and improve off-line hydrologic models and climate model land-surface schemes that include snow processes. Indirectly, these new data sets also allow the hydrologic modeling community to evaluate meteorological forcing data sets developed for mountain environments, which are a key source of uncertainty in simulating hydrologic extremes, especially at moderate elevations close to the freezing elevation.”*

In an extensive review of 30 hydrological models, Beckers et al. (2009) showed that the DHSVM (Wigmosta et al., 1994) is the best model for hydrologic modeling in forest environments over complex mountainous terrain because the detailed process representation of topographic and canopy control of the energy and mass exchange occurs in a spatially distributed manner. DHSVM simulates the effects of soil, vegetation, and topography on the movement of water at and near the land surface. The model consists of a two-layer canopy representation for evapotranspiration, a two-layer energy-balance model for snow accumulation and melt, a multilayer unsaturated soil model, and a saturated subsurface flow model. The canopy snow model explicitly represents the combined canopy processes that govern snow interception, sublimation, mass release, and melt. Snow accumulation and melt below the canopy are simulated using a two-layer mass- and energy-balance model that explicitly incorporates the effects of topography and vegetation cover on the energy and mass exchange at the snow surface. In the absence of a snowpack, available water on the ground is simply rainfall. Surface land-cover and soil properties are assigned to each digital elevation model grid cell or pixel. In each model pixel, the land-surface (vegetation and soil) and meteorological conditions are prescribed and an independent one-dimensional (vertical) coupled energy and water balance is calculated for each pixel. Moisture can then move laterally on the surface or through the subsurface and be intercepted by roads and streams. Vegetation characteristics and meteorological forcing can be manipulated to evaluate changes in land cover and changes in climate.

In this project, we enhanced the DHSVM capability for modeling complex snow-vegetation interactions and developed and validated regional snow parameters for large-domain hydrological application. These advancements laid the foundation for developing NG-IDF curves in complex forest environments, reducing uncertainties in modeling of snow-related hydrologic events, and providing NG-IDF curves for ungauged sites.

1) Enhanced DHSVM Model for Snow Modeling in Canopy Gap

The original DHSVM partitions a forested grid cell either as canopy-covered or open area, neglects the interactions between the two conditions (i.e., the surrounding forest does not shade the open areas), and treats the whole grid cell as a uniform single snowpack (Figure 4). In this project, we enhanced the DHSVM's capability to feature subgrid representation of forest-snow dynamics in canopy gaps. The development was made within the framework of the original DHSVM. Figure 4 illustrates the grid representation of the canopy gap module. In our enhanced DHSVM structure, any model grid cell assigned to have a canopy gap condition is partitioned into two parts: gap and surrounding forest. The model treats the two parts as independent snowpack governed by separate mass and energy input. The enhanced DHSVM explicitly accounts for the impact of the surrounding forest (e.g., wind attenuation and shading) on the gap energy balance and generates spatially varied irradiance with an idealized cylindrical gap geometry. For details about the canopy gap model structure, the readers are referred to Sun et al. (2018).

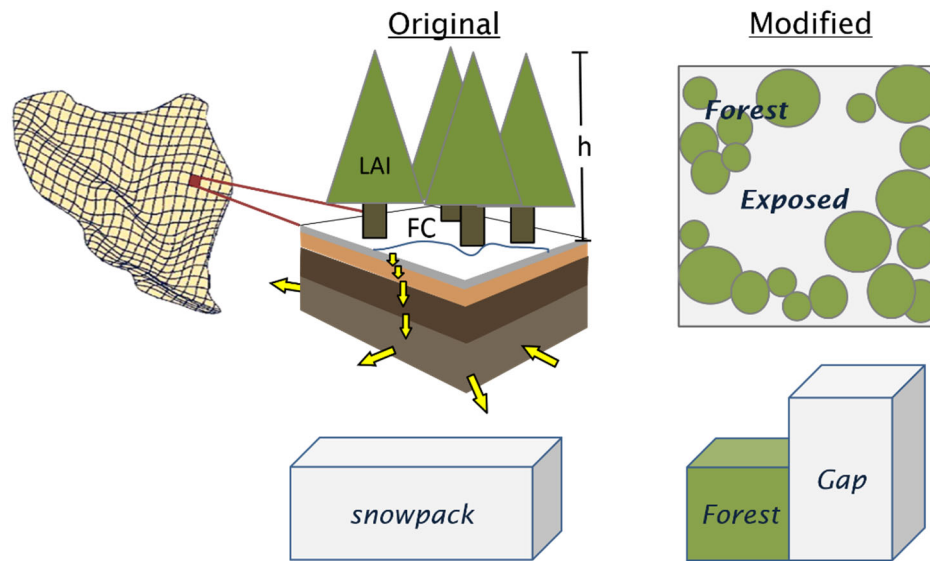


Figure 4. Grid representation of the canopy gap in the original and enhanced DHSVM. Forest canopy is characterized by leaf area index, canopy height, and fractional coverage (see Sun et al., 2018).

After the gap model development, we validated the model at three sites in Idaho that include an open reference site, a densely forested reference site, and a forest gap experimental site located at in the Flat Creek unit of the University of Idaho Experimental Forest. Subhourly meteorological and SWE data were continuously collected at the three sites from mid-February through early May in 2008 (Sun et al., 2018). This model validation can ensure the enhanced DHSVM is able to capture radiation and snowpack dynamics at a fine subhourly scale under distinct canopy covers and lay the foundation for developing NG-IDF curves in complex forested environments.

2) Regional Snow Parameters for Large-Domain Hydrological Application

For modeling snow in large river basins or regions, a major challenge is developing the parameterization of snow processes with reasonable predictive skill in different hydroclimate regimes. While it is a common practice to calibrate snow models with observations at specific point locations, such local calibrations are not feasible for large-domain applications for which

SWE measurements mostly have limited temporal and spatial coverage. Moreover, local calibrations may produce different parameters even for areas that are in close spatial proximity and have physical similarity leading to overfitting of parameter values that may be symptomatic of model structural issues. This challenge also applies to hydrologic prediction and forecasting in snow-dominated regions where a key source of uncertainty is the modeling of SWE.

Figure 5 summarizes the workflow for the development of regional snow parameters for DHSVM's snow submodel. As Hamlet (2018) commented, the BCQC SNOTEL data were used for the model validation. A total of 246 SNOTEL stations with the longest common period of BCQC records from Water Years 2007 through 2013 were selected for this purpose. We also compared the cumulative density function (CDF) of daily SWE observations over the 2007–2013 period to all available observational records and confirmed that the selected 7-year records are generally representative of the long-term SWE dynamics at each station. We first performed a sensitivity test for DHSVM snow parameters to identify sensitive parameters that dominate model behaviors, and then developed regional snow parameters for these sensitive parameters.

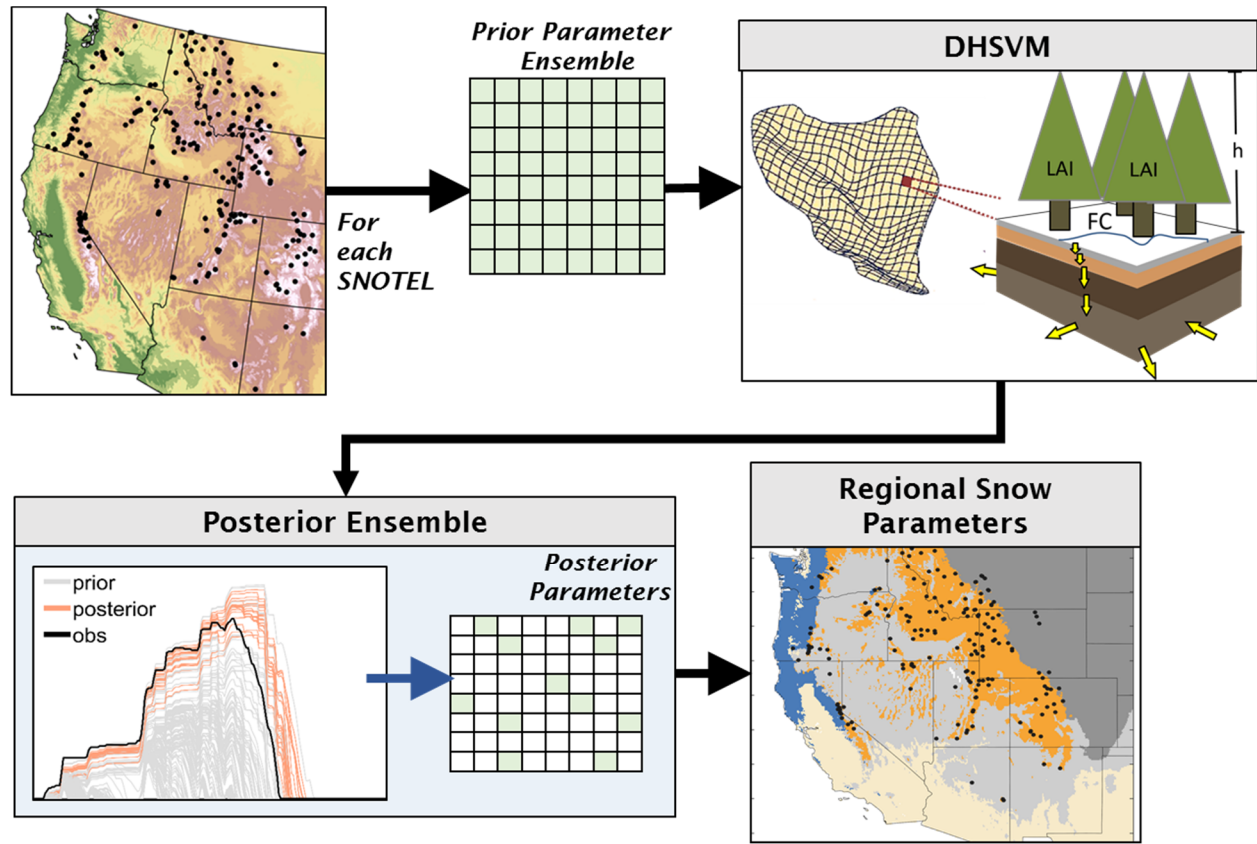


Figure 5. Illustration of flowchart for regional snow parameters development (see Sun et al., 2019).

In the DHSVM snow model, the energy and mass balance is modeled using five parameters: fresh snow albedo (α_{max}), albedo decay coefficient during snow accumulation (λ_A), albedo decay coefficient during melt (λ_M), and precipitation partitioning threshold temperature including the minimum temperature for precipitation to be in the form of rain (T_R , in $^{\circ}\text{C}$) and the maximum temperature for precipitation to be in the form of snow (T_S , in $^{\circ}\text{C}$) (Sun et al., 2019).

In this project, we studied the parameter sensitivity using the Generalized Sensitivity Analysis (GSA) method (Hornberger and Spear, 1981). The GSA method first draws ensemble parameter sets from prior distributions using the Monte Carlo method and then partitions the parameter sets into behavioral/nonbehavioral classes based on a prescribed threshold for model performance metric. A minimum Kolmogorov-Smirnov threshold value of 0.25 was used here to identify sensitive parameters (Sun et al., 2019). Based on the behavior parameter sets, we then obtained posterior parameters that achieved good performance in SWE modeling skill including timing, daily variation, and peak magnitude.

At each of the 246 SNOTEL stations, we constructed the prior ensemble parameter space consisting of 10,000 parameter sets drawn uniformly using the Latin Hypercube Sampling algorithm. Physically implausible parameter combinations were eliminated and an ensemble of 10,000 simulations was performed using an ensemble of prior parameters that were constrained to physically reasonable ranges. Based on the determined posterior parameter space, we developed regionally uniform coherent snow parameters as the ensemble mean over the posterior parameter space for the five regions across the western United States that were classified based on winter (Nov–Mar) mean precipitation, winter air temperature, and winter wind speed. The winter meteorological data cover CONUS using a $1/16^\circ$ spatial grid and include precipitation, minimum and maximum air temperature, and wind speed. These data were acquired from the meteorological data set of Livneh et al. (2013). Figure 5 presents the five regions including alpine (orange), maritime (blue), southern (yellow), northern (dark grey), and prairie (light grey). The developed regional snow parameters were then evaluated at individual SNOTEL stations in each region to confirm their applicability at the site level.

3.3 Objective 3: Validation of NG-IDF Tool in Design Flood Estimates

Before implementing the NG-IDF curves on a real design problem, evaluation and benchmarking the performances of the curves in design flood estimates is essential, because risks related to the United States infrastructure design will involve trillions of dollars in the coming decades (Carnevale and Smith, 2017). In IDF evaluation studies, given the spatially sparse network of stream gauges and the general lack of representative long-record measured flow data, it is common to use validated hydrological models (e.g., progress under Objective 2) to continuously simulate streamflow and provide a performance benchmark of design floods against which to evaluate the IDF technology (Boughton and Droop, 2003; Rogger et al., 2012). Especially in the snow-dominated regions that feature complex topography, streamflow is very poorly observed and most tributary streams have never been gauged due to inherent difficulties of access (Bales et al., 2006; Lundquist et al., 2016). The limited streamflow data further necessitate the use of a well-validated hydrologic model to continuously simulate streamflow and provide a performance benchmark of design flood estimates against which to evaluate the NG-IDF approach. Furthermore, physics-based, continuous simulation models represent the state of the science in hydrologic design, particularly in snow-dominated locations. This validation allows us to identify climatic conditions where NG-IDF-based design may be a viable alternative to continuous simulation models.

Figure 6 presents the flowchart for evaluating the NG-IDF tool using the DHSVM continuous simulation method. First, local meteorological forcing data were used to drive the well-validated DHSVM for generating continuous streamflow simulations and water available for runoff (W).

Second, the NG-IDF curves were developed based on the annual maximum W from DHSVM continuous simulation (e.g., fitting GEV distribution). The PREC-IDF curves were also developed based on annual maximum precipitation from the meteorological forcing data for comparison purposes. Third, flood frequency statistics (e.g., 25-year flood) were derived from the DHSVM-simulated annual maximum streamflow and considered as the benchmark for NG-IDF validation. Fourth, the DHSVM-derived NG-IDF curves and PREC-IDF curves were both used to drive the NRCS TR-55 event-based rainfall-runoff model to compute the associated design floods. Last, the design flood estimates from the TR-55 model with NG-IDF and PREC-IDF input were compared to those from DHSVM continuous simulations. A Monte Carlo (MC) simulation approach suggested by NOAA Atlas 14 was used here to consider sample data uncertainty in both IDF and flood frequency analysis. The Z-statistic (Madsen et al., 2009) is used for the statistically significant test (i.e., to determine whether the difference between the two methods is statistically significant). In this project, we conducted the NG-IDF validation at both hillslope and basin scales.

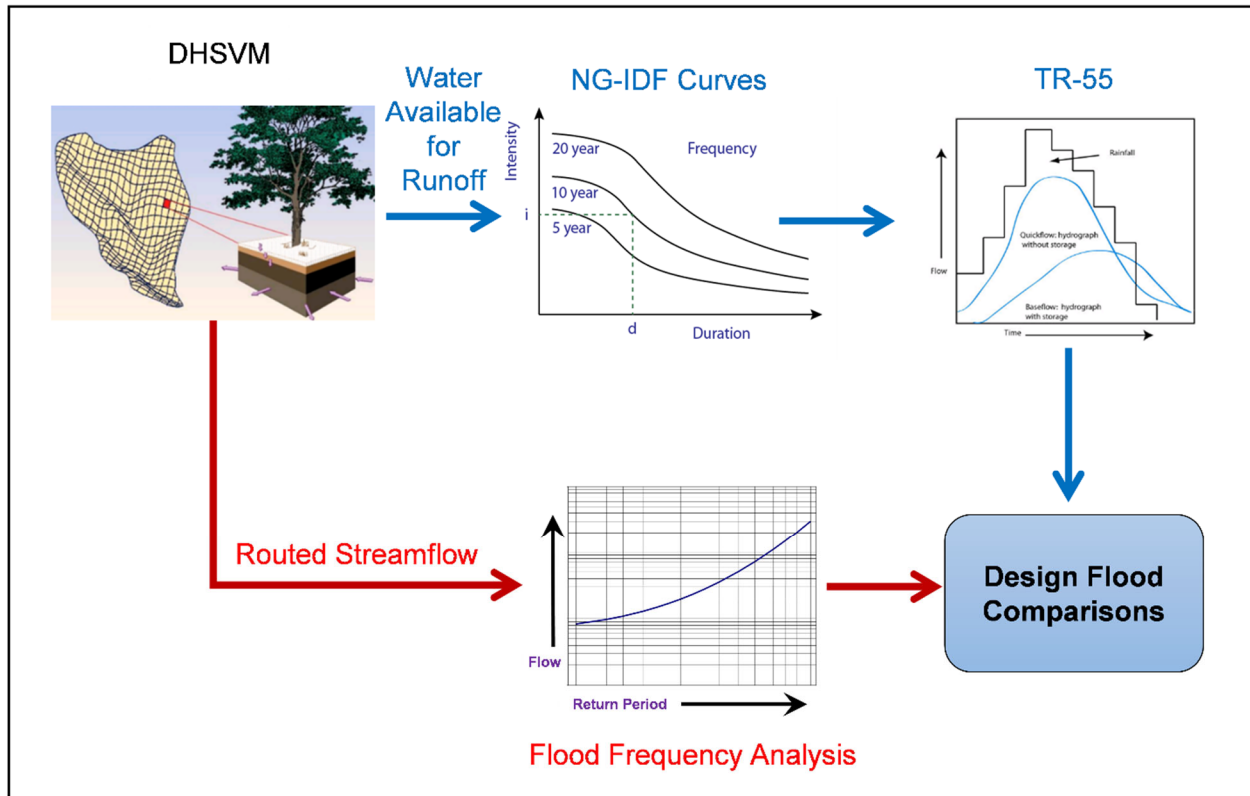


Figure 6. Flowchart for NG-IDF tool validation using the DHSVM continuous simulation method. DHSVM was applied for both hillslope and basin scales for NG-IDF validation.

1) Validation of NG-IDF Tool at Hillslope Scale

A small experimental hillslope appropriate for mountain topography is shown in Figure 7. The planar hillslope had a constant slope of 0.3 and a soil depth of 1.5 m, which were consistent with the hillslope design reported by Wigmosta and Lettenmaier (1999). The hillslope was represented by ten 4 m × 4 m DHSVM grid cells and had bare soil cover approximating SNOTEL open area conditions. The specified soil parameters, shown in Figure 7, were selected

to produce a subsurface time of concentration for the hillslope of 24 hours. We also confirmed that the DHSVM representation resulted in a 24-hour time of concentration using a steady-state test with constant rainfall input. The hillslope was tested at each of the selected 246 SNOTEL stations at which the DHSVM was well-validated under Objective 2 (Yan et al., 2020). For each site, the DHSVM was forced by 15-minute meteorological data including precipitation, humidity, wind speed, air temperature, and shortwave and longwave radiation. The 15-minute forcing data were estimated (if necessary) and disaggregated from the daily BCQC SNOTEL data using the Mountain Microclimate Simulation Model (Hungerford et al., 1989). The use of the experimental hillslope with the same soil characteristic and time of concentration allowed us to examine and compare the NG-IDF performance across the entire western United States under different hydroclimatic conditions, thus providing a robust benchmark report of the NG-IDF tool in design flood estimates.

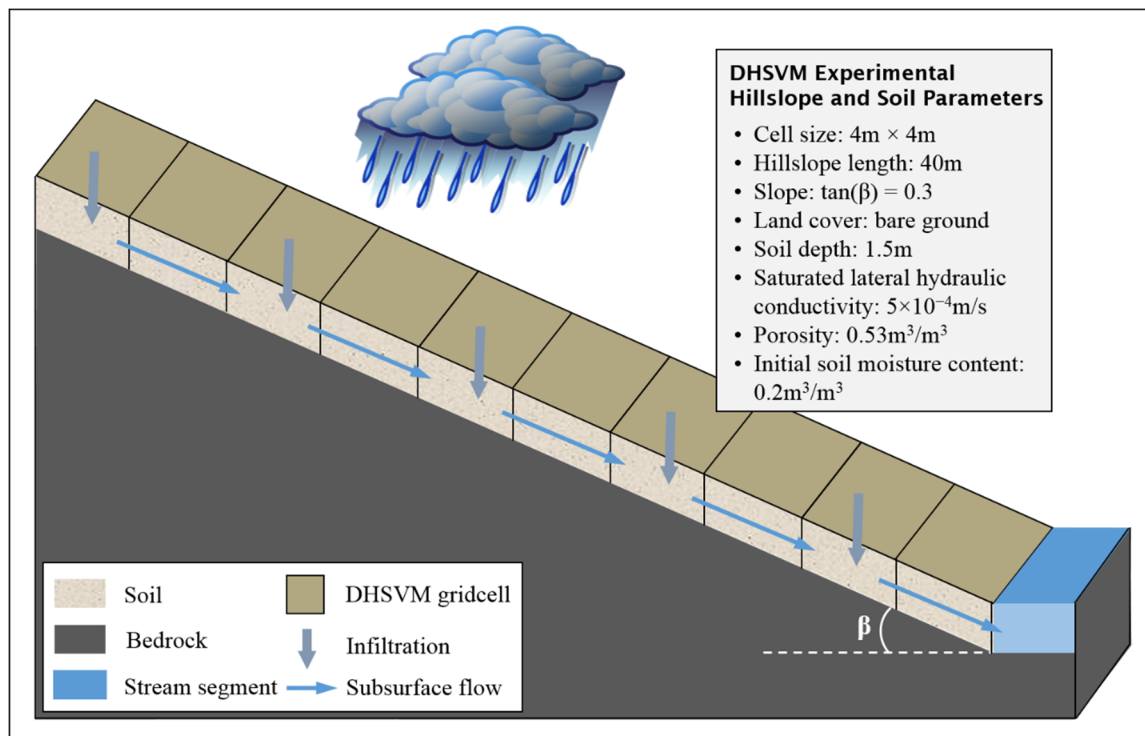


Figure 7. The experimental hillslope and the soil characteristics used in the DHSVM modeling (see Yan et al. 2020).

2) Validation of NG-IDF Tool for a DoD Site at Basin Scale

In a real design problem, the watershed of interest has multiple types of land cover, and streamflow is generated by a variety of surface and subsurface flow processes. Thus, in addition to the validation at the hillslope scale, we also applied and validated the NG-IDF approach at the basin scale, representing practical engineering design conditions with multiple land covers. The selected basin, the Upper West Walker Basin (UWWB), is located on the east side of the Sierra Nevada Mountains and contains the Marine Corps Mountain Warfare Training Center (MCMWTC) (Figure 8). The UWWB has a drainage area of 633 square kilometers and is part of the 8-digit Hydrologic Unit Code (HUC) 16050302. We first calibrated the DHSVM for the

whole UWWB using the daily streamflow data from two U.S. Geological Survey (USGS) gauges (ID 10296500 and 10296000) and the daily BCQC data from the Sonora SNOTEL station (located within the basin). The $1/16^\circ$ grid cell meteorological forcing data, including precipitation, minimum and maximum air temperature, and wind speed, were acquired from the meteorological data set of Livneh et al. (2013). We again used the Mountain Microclimate Simulation Model to disaggregate the daily forcing data into 15-minute time steps for DHSVM simulation. After the calibration and validation of DHSVM, a small test basin flowing to the MCMWTC was selected for the NG-IDF technology assessment (Figure 8). The test basin had an area of 2.36 square kilometers, which aligns with the TR-55 hydrologic model scale (i.e., small watershed design). The small test basin had a uniform bedrock soil but included multiple types of vegetation cover. The dominant land covers in the test basin are grassland (32%) and forest (31%).

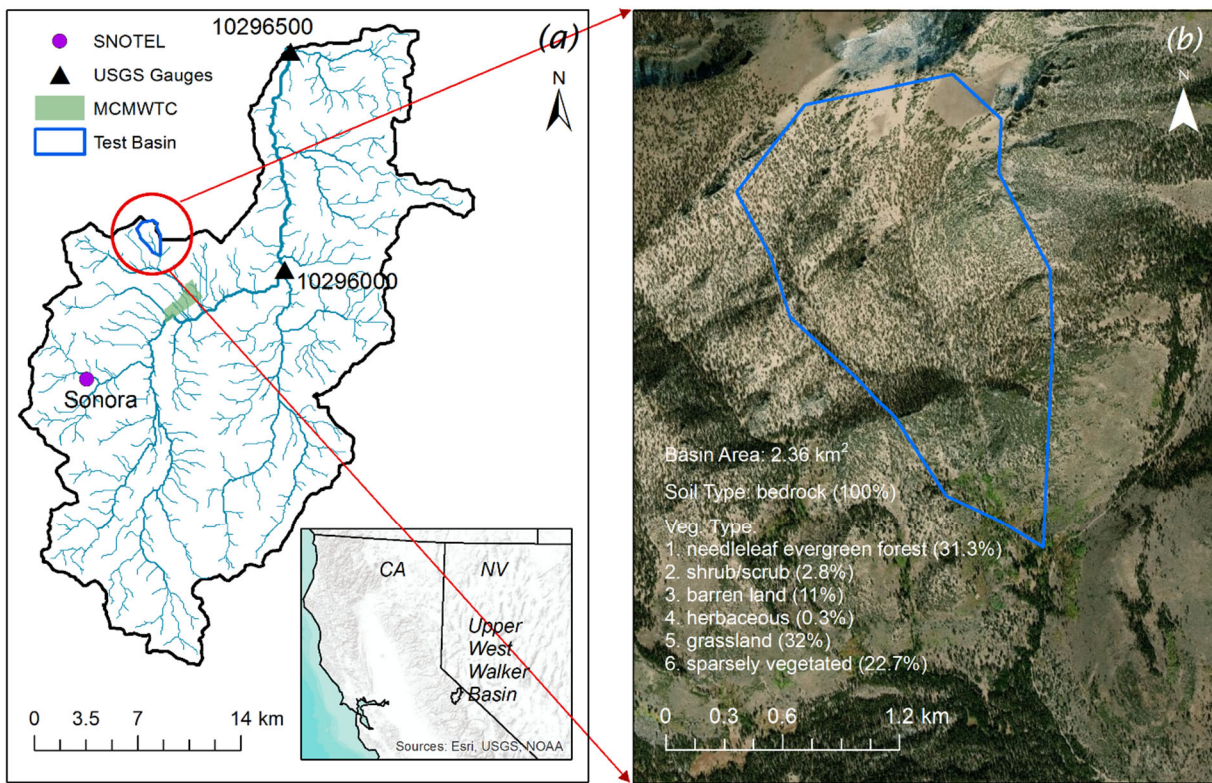


Figure 8. The location of the small test basin that flows into the DoD Marine Corps Mountain Warfare Training Center (MCMWTC).

In the test basin validation, we also explored the effects of hyetograph shape on design flood estimates. Standard rainfall hyetographs such as the uniform hyetograph and NRCS hyetographs (McCuen, 1998) were considered here. Further, for the first time, we proposed the hyetograph of water available for runoff (W) to improve NG-IDF performance. We followed the methodology used by NOAA Atlas 14 (Bonnin et al., 2011) that was originally proposed by Huff (1990). We computed W accumulation for specific time periods to be consistent with the durations used for the NG-IDF curves. A moving window approach was used to estimate W accumulation for each selected duration based on the 15-minute DHSVM basin mean W time series. The three largest

W accumulations were then obtained for each month in the entire simulation record and for each of the selected durations. Following the NOAA Atlas 14 approach, the 2-year return period W magnitude was used as a minimum threshold to ensure only large W events were being captured for developing hyetographs. After acquiring all large W events, each of the accumulations was converted into a ratio of the cumulative 15-minute W to the total W for that duration (i.e., percent of total W), and a ratio of the cumulative time to the total time for that duration (i.e., percent of duration). We could then estimate the cumulative probability of occurrences similar to the probabilistic precipitation hyetographs developed in NOAA Atlas 14. For example, the 10% hyetograph curve indicates that 10% of the corresponding W events had temporal distributions that fell above of the curve.

3.4 Objective 4: NG-IDF Products under Historical Climate

After conducting the DHSVM validation at the five regions in the western United States, we classified the remainder of the CONUS into the five regions based on the same metrics: the winter mean precipitation, air temperature, and wind speed data. Because no in situ SWE network existed in the middle and eastern United States, the whole CONUS was modeled using the five regional snow parameter sets developed based on the SNOTEL observations. For the CONUS application, we ran the DHSVM at more than 200,000 $1/16^\circ$ grid cells using the meteorological data of Livneh et al. (2013) for the period of 1950–2013. The DHSVM was run in parallel at the PNNL high-performance computing (HPC) infrastructure Constance Cluster with 520 nodes. After the simulation, the W time series were estimated using mass balance. The stationarity assumption was tested using the nonparametric Mann-Kendall test, and we detrended the data when significant temporal trends were identified, while maintaining the time series average. Both PREC-IDF and NG-IDF curves were then developed by fitting the detrended annual maximum data into the GEV distribution. We archived these PREC-IDF and NG-IDF curves under historical climate for potential input to a single event rainfall-runoff model to produce flood estimates.

By comparing the PREC-IDF curves to the NG-IDF curves, we can identify the locations subject to the potential for underdesign and overdesign when using PREC-IDF curves. These results are useful for planners to guide their infrastructure design and decide whether snow processes are an important factor in their design process. In addition to producing IDF products, we further explored the spatial variability of the dominant drivers of extreme events across CONUS. The dominant extreme events were classified as snowmelt only, rainfall, and ROS. A ROS event was defined as rainfall of at least 10 mm per day falling on snowpack of at least 10 mm SWE, and the sum of rainfall and snowmelt contained at least 20% snowmelt (Musselman et al., 2018). Rainfall was defined as precipitation falling on snow-free ground and the snowmelt only was defined as a decrease in SWE without precipitation (Yan et al., 2018). The same frequency analysis procedure (e.g., stationarity test, fitting GEV distribution) applied to the W and P were used for estimating the extreme events of rainfall, ROS, and melt at each selected return period or ARI. For each grid cell across CONUS, the mechanism that produced the largest extreme event was identified as the dominant driver.

3.5 Objective 5: Modeling Extreme Precipitation in Both Historical and Future Climate

Although GCMs are the primary tools for projecting future climate under different GHG emissions and land-use/land-cover scenarios, their spatial resolutions are constrained by

computational resources. Hence, even for the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor et al., 2012) that provided climate change scenarios for the most recent Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6), most GCMs were applied at a resolution between 1° to 2°, which is not sufficient for resolving mesoscale features key to the generation of extreme precipitation. Two downscaling approaches were used here to develop regional climate change scenarios for assessing climate change impacts: statistical downscaling and dynamical downscaling. Because of the low computational cost, statistical downscaling approaches can be applied to a large ensemble of GCM projections to estimate uncertainty. However, they suffer from the inherent assumption of stationarity as statistical relationships developed from historical data are applied to future climates. In contrast, dynamical downscaling based on regional climate models (RCMs) is physically based but computationally demanding. Past efforts have mostly used a single RCM driven by a single GCM scenario, thereby limiting the evaluation of uncertainty. In this project, we took advantage of both approaches to project future climate change.

1) WRF Dynamical Downscaling

We used an RCM based on the WRF model (Chen et al., 2018) to dynamically downscale scenarios from five GCMs selected from CMIP5 for the present (1981–2010) and future (2041–2070). Following the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2013), we included five GCMs in this project:

- the second-generation Canadian Earth System Model (*CanESM2*)
- the Community Earth System Model version 1 that includes the Community Atmospheric Model version 5 (*CESM1_CAM5*)
- the Geophysical Fluid Dynamics Laboratory Earth System Model Version 2M (*GFDL_ESM2M*)
- the Hadley Global Environment Model 2-Earth System (*HadGEM2_ES*),
- the Max Planck Institute for Meteorology Earth System Model MR (*MPI_ESM_MR*).

The five GCMs selected represented a good diversity of models because their earth system components generally followed different developmental paths and approaches. For the mid-century time frame, uncertainty in climate projections due to emission scenarios is less important than uncertainty related to the use of different GCMs (Hawkins and Sutton, 2009), so only the Representative Concentration Pathway mitigation scenario (RCP8.5) was used in our modeling and analysis. The WRF simulations covered the western United States at about 6 km grid resolution to adequately resolve key regional forcing such as orography.

To characterize uncertainty in projecting precipitation extremes, the regional WRF simulations were also analyzed to identify the physical-dynamical mechanisms responsible for changes in extreme precipitation in different climate regimes. In California and the Pacific Northwest, extreme precipitation is often associated with atmospheric rivers (ARs) (Chen et al., 2019b) that transport warm, moist air of tropical origin in narrow bands that produce heavy precipitation upon landfall over the coastal mountains. These are predominantly cold-season processes as frontogenesis plays an important role in the development of ARs. Because ARs can significantly modulate surface hydrological processes (e.g., snow) through the extreme precipitation they produce, we quantified the relationship between ARs and the magnitude of extreme precipitation,

and their impacts on land-surface hydrologic process to illustrate the physical mechanism behind the changes in the NG-IDF curves under future climate conditions.

2) MACA Statistical Downscaling

In addition to the WRF dynamically downscaled products over the western United States, we also used the Multivariate Adaptive Constructed Analogs (MACA) statistically downscaled products from the Climatology Lab at University of Idaho (Abatzoglou and Brown, 2012) for CONUS analyses. MACA is a multi-step procedure that uses bias correction procedures and a constructed analogs approach for developing the fine-scale spatial pattern using a library of observed patterns. MACA is similar in construction to other Constructed Analogues but uses an observational data set to remove historical bias and fit appropriate spatial patterns in the output. Based on the 6 km meteorological data of Livneh (2013), the Climatology Lab downscaled and archived 20 CMIP5 GCMs under RCP4.5 and RCP8.5 for the CONUS and the Canadian portion of the Columbia River Basin for the period of 2006–2099. They also provided the GCM downscaled data for the historical period 1950–2005. The MACAv2-LIVNEH data set includes downscaled data for maximum temperature, minimum temperature, maximum and minimum relative humidity, precipitation accumulation, downward surface shortwave radiation, wind-velocity, and specific humidity. In this project, we selected 10 GCMs that were shown to represent divergence for future changes in temperature and precipitation for the RCP8.5 emission scenario for the CONUS (Naz et al., 2016). The 10 GCMs included the following:

- version 1.1 of the Beijing Climate Center Climate System Model (*bcc-csm1-1*),
- the Beijing Normal University Earth System Model (*BNU-ESM*),
- the fourth version of the Community Climate System Model (*CCSM4*),
- the Centre National de Recherches Meteorologiques Coupled Global Climate Model Version Five (*CNRM-CM5*),
- the Commonwealth Scientific and Industrial Research Organisation Mk3.6.0 model (*CSIRO-Mk3-6-0*),
- the Russian Institute for Numerical Mathematics Climate Model Version 4 (*inmcm4*),
- the Institut Pierre Simon Laplace Model CM5A-LR (*IPSL-CM5A-LR*),
- the Institut Pierre Simon Laplace Model CM5A-MR (*IPSL-CM5A-MR*),
- the Meteorological Research Institute Coupled Global Climate Model Version Three (*MRI-CGCM3*)
- the Norwegian Climate Center's Earth System Model (*NorESM1-M*).

3.6 Objective 6: Ensemble NG-IDF Products under Future Climate

Changes in the frequency, intensity, and duration of snowmelt and/or ROS events respond to local changes in precipitation, air temperature, incoming short- and longwave radiation, wind, and vapor pressure, as well as antecedent snowpack conditions including pack surface albedo, temperature, and SWE. Both dynamically and statically downscaled historical and future climate ensembles derived under Objective 5 were used to drive the validated DHSVM derived under Objectives 2 and 3 to generate a time series of P and W. We used the extreme value theory (i.e., GEV distribution) and L-moments statistics to develop both ensemble PREC-IDF and NG-IDF curves. Specially, we developed (1) 10 ensemble NG-IDF products for future climate across CONUS using the MACA statistically downscaled data and (2) 5 ensemble NG-IDF products for future climate at the DoD site (MCMWTC) using the WRF dynamically downscaled data. The

ensemble NG-IDF curves partially quantified the GCM structure uncertainty for each selected storm duration and return period. This probabilistic NG-IDF tool can support engineering design risk analysis that is adjustable based on design cost and changing operating conditions. With both historical and future climate simulations, the projected climate change impacts on NG-IDF curves, PREC-IDF curves, and dominant drivers of extreme events were also analyzed and discussed.

4. Results and Discussion

The results and discussion for this project are described separately for each section in detail below. Section 4.1 presents the SNOTEL-based NG-IDF results (Objective 1). Section 4.2 presents the DHSVM validation results and the developed regional parameters (Objective 2). Section 4.3 presents the NG-IDF validation results at both hillslope and basin scales containing a DoD site (Objective 3). Section 4.4 presents the NG-IDF products across CONUS under historical climate conditions (Objective 4). Section 4.5 presents the impact of ARs on extreme precipitation and snow events (Objective 5). To better discuss changes of NG-IDF curves from historical to future climate conditions, the results for Objectives 5 and 6 are presented together in Section 4.6.

4.1 NG-IDF Proof-of-Concept Using SNOTEL Observation

We developed both PREC-IDF curves and NG-IDF curves for nearly 400 SNOTEL stations across the western United States that each had at least 30 years of high-quality records. By comparing the NG-IDF curve value with the PREC-IDF curve value, we evaluated the potential design risk for a given duration and return period for three cases (Figure 9):

- underdesign: if the PREC-IDF value is less than the NG-IDF value;
- overdesign: if the PREC-IDF value is greater than the NG-IDF value; and
- proper design: if the differences between the two IDF curve values are less than 10%.

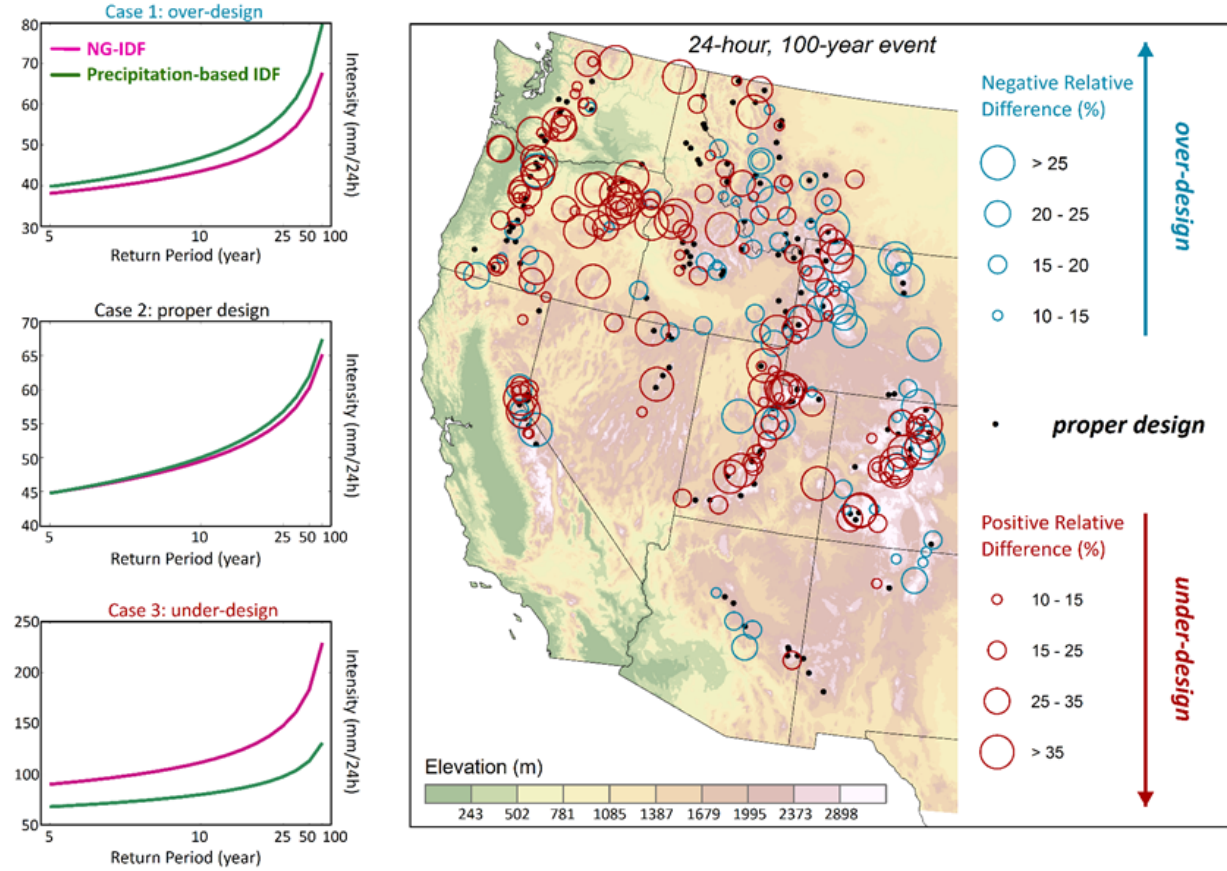


Figure 9. Left panel: examples of three design cases in terms of IDF values. Right panel: relative errors of the 24-hour, 100-year event between the NG-IDF and PREC-IDF curves (see Yan et al., 2018).

We used the 10% threshold for the proper design case to account for data sample uncertainty in IDF frequency analysis. We acknowledge that this value is somewhat arbitrary and could be higher for short records such as the SNOTEL data. Figure 9 shows the spatial relative differences between the NG-IDF and PREC-IDF curves for the 24-hour, 100-year event across all selected SNOTEL stations. In general, results obtained from about 43% of the SNOTEL stations suggested the potential for underdesign across the mountainous regions of the western United States. For these underdesigned stations, the extreme W event exceeded the extreme P event from 10% to as much as 125%. The largest error (125%) occurred at the Idarado SNOTEL station in Colorado. The potential for overdesign occurred at about 20% of the SNOTEL stations, which corresponded to the negative relative errors between the extreme events (W and P) of -75 to -10% . We also found about 37% of the stations to be in the proper design case, where the relative error between the P and W values was within 10%. In Figure 9, it is observed that most of the stations found to be underdesigned were in the Colorado Front Range, the Wasatch Mountains, and the Pacific Northwest. In general, all these regions were characterized by longer snow accumulation seasons and deeper snowpacks (Yan et al., 2018). Note that we only show the 24-hour, 100-year design storm here and these statistical values can vary with different durations and return periods. In summary, the comparisons shown in Figure 9 using the SNOTEL measurements emphasized the potential infrastructure failure risk or unnecessary cost

associated with standard practice when snowmelt and ROS events drive peak W in snow-dominated regions.

We define risk of failure (R) over a planning period (n) following the definition in most hydrology textbooks (Chow et al., 1988) as

$$R = 1 - (1 - p)^n \quad (2)$$

where $p = 1/T$ is the annual exceedance probability and T is the return period of an event. The risk of failure R indicates the likelihood of experiencing at least one event exceeding the design event over a given project life of n years. A risk value closer to 0 indicates lower risk and a higher value suggests a higher risk of failure. If the project design life $n = T$, then $R \rightarrow 0.63$ when $n \rightarrow +\infty$ (Yen, 1970).

As an example, consider SNOTEL (45.21, -117.19) in Cascades, Oregon. The 24-hour, 100-year PREC-IDF and NG-IDF curve values are 125.4 mm and 155.5 mm, respectively (Figure 10a). Therefore, the underestimation percentage for the PREC-IDF is about $(155.5 - 125.4) / 125.4 * 100 = 24\%$. In addition, the 125.4 mm event is not an actual 100-year event at this location, where snowmelt and rain-on-snow dominate or contribute to extreme events. Based on the NG-IDF curve (the actual water available for runoff), the 125.4 mm event is actually a 30-year event (Figure 10b). Therefore, the actual design risk of failure over 100 years is $R = 1 - \left(1 - \frac{1}{30}\right)^{100} = 0.97$ and the standard PREC-IDF approach underestimates the design risk of failure by $(0.97 - 0.63) / 0.63 * 100 = 54\%$ (Figure 10). In this case, a 24% underestimation of the 100-year event results in an underestimated design risk of 54%. Of course, this value varies with the design life of the project.

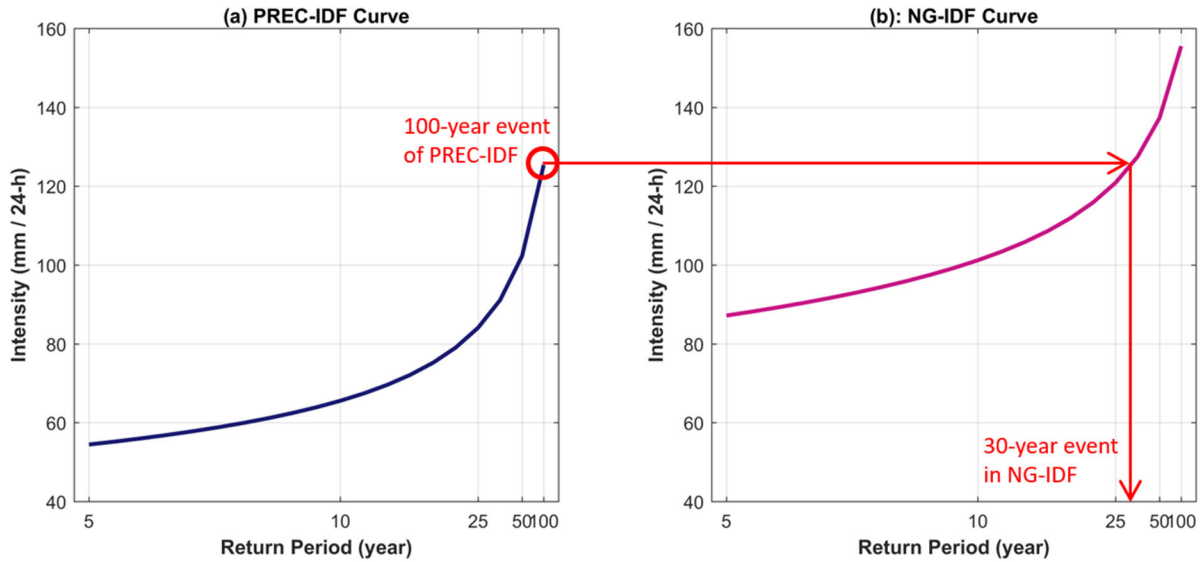


Figure 10. PREC-IDF and NG-IDF curves at the selected SNOTEL (45.21, -117.19) in Cascades, Oregon. The 100-year event of PREC-IDF corresponds to a 30-year event in the NG-IDF curve.

Given a 100-year design life and using the 24-hour, 100-year event in the standard PREC-IDF curves, Figure 11a presents the frequency of their return periods in the corresponding NG-IDF curves among all SNOTELs across the western United States. Values less than 100 indicate

underdesign and values over 100 indicate overdesign. Note only SNOTEL stations with absolute percentages of difference >10% between PREC-IDF and NG-IDF curves are shown in Figure 11. Approximately 14% of the 100-year PREC-IDF curves have a return period less than 25 years when compared to the NG-IDF value and about 33% have a return period of 50 years or less.

Figure 11b further presents the frequency of underestimation/overestimation of the design risk of failure in the 100-year design life period. Positive values indicate the use of PREC-IDF curves results in an underestimation of the risk failure and negative values indicate the opposite. The use of PREC-IDF curves would result in an underestimated risk of failure in excess of 50% at about 24% of the locations and an underestimation in excess of 25% at about 40% of the locations.

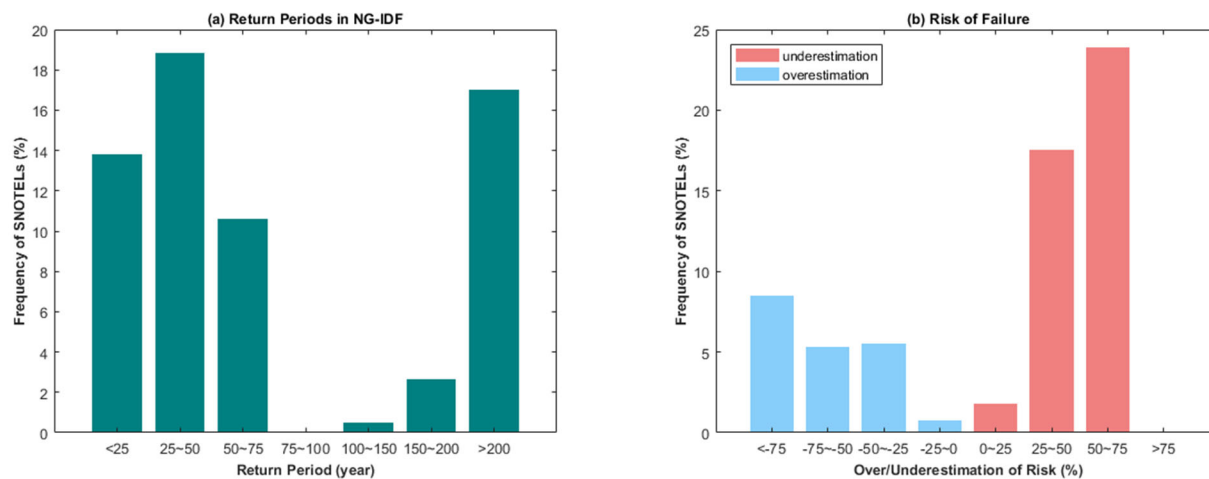


Figure 11. Histogram of the (a) actual associated return periods in NG-IDF curves for a 24-hour 100-year event in PREC-IDF curves, and (b) the associated over/underestimation of risk of failure. Positive values on the x-axis indicate underestimation of risk and negative values indicate overestimation of risk.

The risks associated with underdesign are typically obvious and can be catastrophic, such as those associated with water conveyance (spillways, bridges, culverts, etc.). Therefore, an additional factor of safety may be required by regulatory agencies due to uncertainty in the hydrologic design process and inefficient operation of the facility, such as a reduction in pipe capacity resulting from an accumulation of debris during a storm event or improper maintenance. The factor of safety should not be used to accommodate known deficiencies in the hydrologic design process, such as using PREC_IDF values in place of larger NG-IDF values. In this case the factor of safety should be applied to design calculations based on the NG-IDF curves.

Most modern drainage system design requires the use of water detention facilities to maintain flood peaks after development similar to those prior to development. The required volume of detention is related the difference between the post- and pre-development hydrographs. If the PREC-IDF exceeds the NG-IDF curve and produces a larger difference in pre- and post-development hydrographs (typical with urbanization), the facility will be overdesigned, leading to excess cost and a detention system that does not perform as intended. For example, if a detention pond is designed to store excess stormwater runoff due to upstream urbanization but also required to maintain the pre-development environmental flow requirement downstream. If PREC-IDF value exceeds the NG-IDF value and a detention pond is designed following the

PREC-IDF value, then much more flow reduction would be performed, leading to unnecessary cost and the degradation of the downstream aquatic ecosystem.

Because of the highly nonlinear rainfall-runoff process, the errors between PREC-IDF and NG-IDF curves shown in Figure 9 and the risk associated with the IDF curves can further reduce or increase in the associated design flood estimates after the rainfall-runoff process (Yan et al., 2019b). To better identify locations with the potential for underdesign or overdesign based on flood estimates, it is necessary to quantify the design flood errors using a rainfall-runoff model. Here we evaluated design reliability by coupling the PREC-IDF and NG-IDF curves with the TR-55 event-based rainfall-runoff model. Similar to the IDF curve comparison, we also evaluated the potential design risk for a given return period and duration for three cases by comparing the peak design flood estimates from PREC-IDF (q_P) and NG-IDF (q_W) curves:

- underdesign: if the q_P is less than q_W ;
- overdesign: if the q_P is greater than q_W ; and
- proper design: if the differences between q_P and q_W are less than 10%.

We used the same 10% threshold value as that used in the IDF comparison to consider data uncertainty in IDF frequency analysis and thus provide a fair comparison. Figure 12 presents the relative errors between the peak design floods derived from the NG-IDF curve (q_W) and the PREC-IDF curve (q_P) for the four return periods using the critical design duration approach (i.e., at each site, the duration that produces the largest peak flow for a given return period): 10-year, 25-year, 50-year and 100-year periods. These results are also presented in the form of histograms in Figure 13. We found about 72%, 64%, 63%, and 57% of the SNOTEL stations suggested the potential for underdesign when using the standard PREC-IDF curves for the 10-, 25-, 50-, and 100-year events, respectively. For stations in the underdesign case, the q_W exceeded q_P by as much as 214% for the 10-year event, 242% for the 25-year event, 278% for the 50-year event, and 324% for the 100-year event. We found the largest error —324%—at the Lily Pond station in Colorado. We also found about 7%, 10%, 12%, and 16% of the SNOTEL stations suggested the potential for overdesign for the 10-, 25-, 50-, and 100-year events, respectively. For these stations in the overdesign case, the q_P exceeded q_W by as much as 24% for the 10-year event, 47% for the 25-year event, 64% for the 50-year event, and 75% for the 100-year event. On average, the q_W were 53%, 48%, 44%, and 46% greater than the q_P across all the underdesigned SNOTEL stations for the 10-, 25-, 50-, and 100-year events, respectively. The q_P were on average 17%, 23%, 28%, and 33% greater than the q_W across all the overdesigned SNOTEL stations for the 10-, 25-, 50-, and 100-year events, respectively.

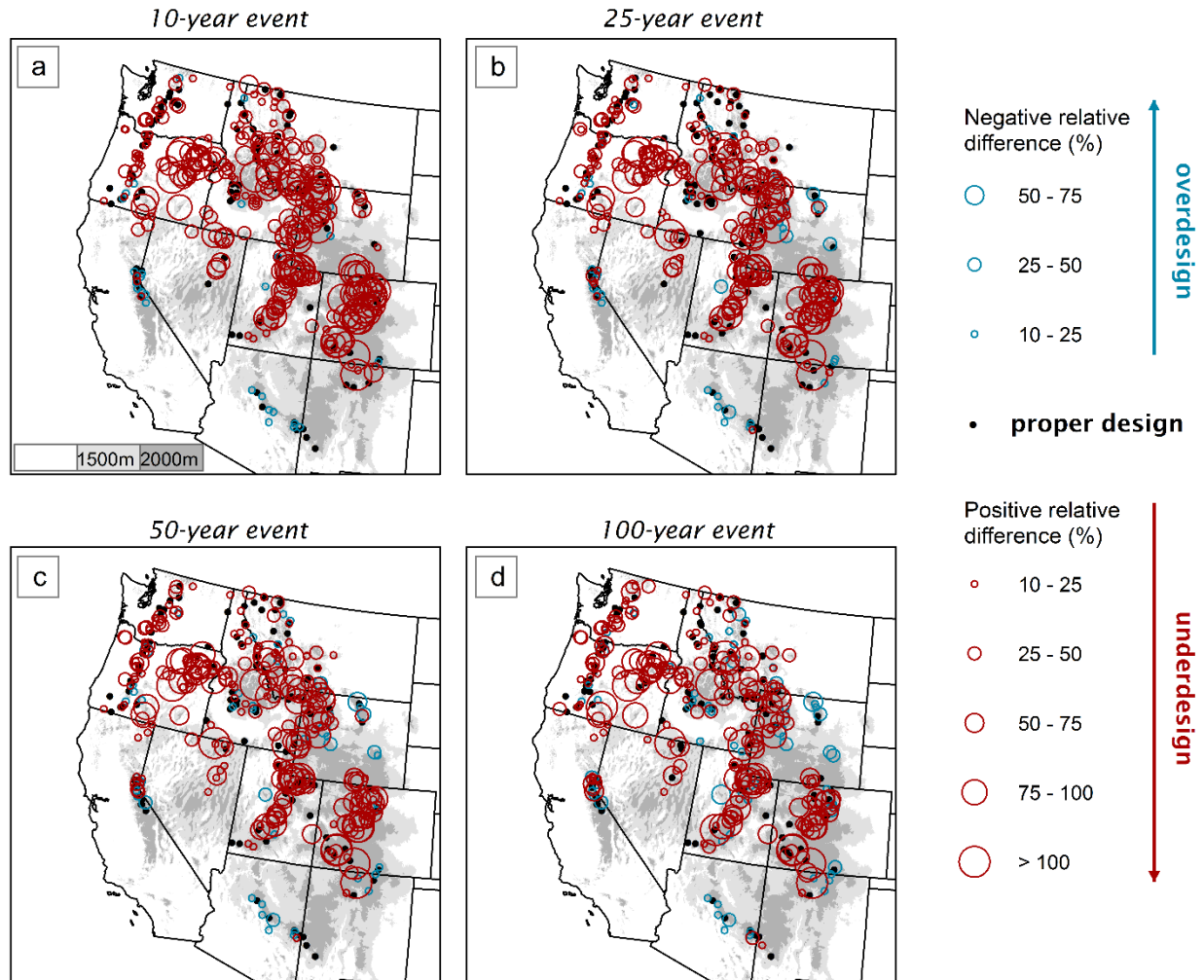


Figure 12. Relative errors between peak design flood estimates from the NG-IDF curves and PREC-IDF curves for the four selected return periods (see Yan et al., 2019b). The examples of the three design cases are in terms of design flood estimates.

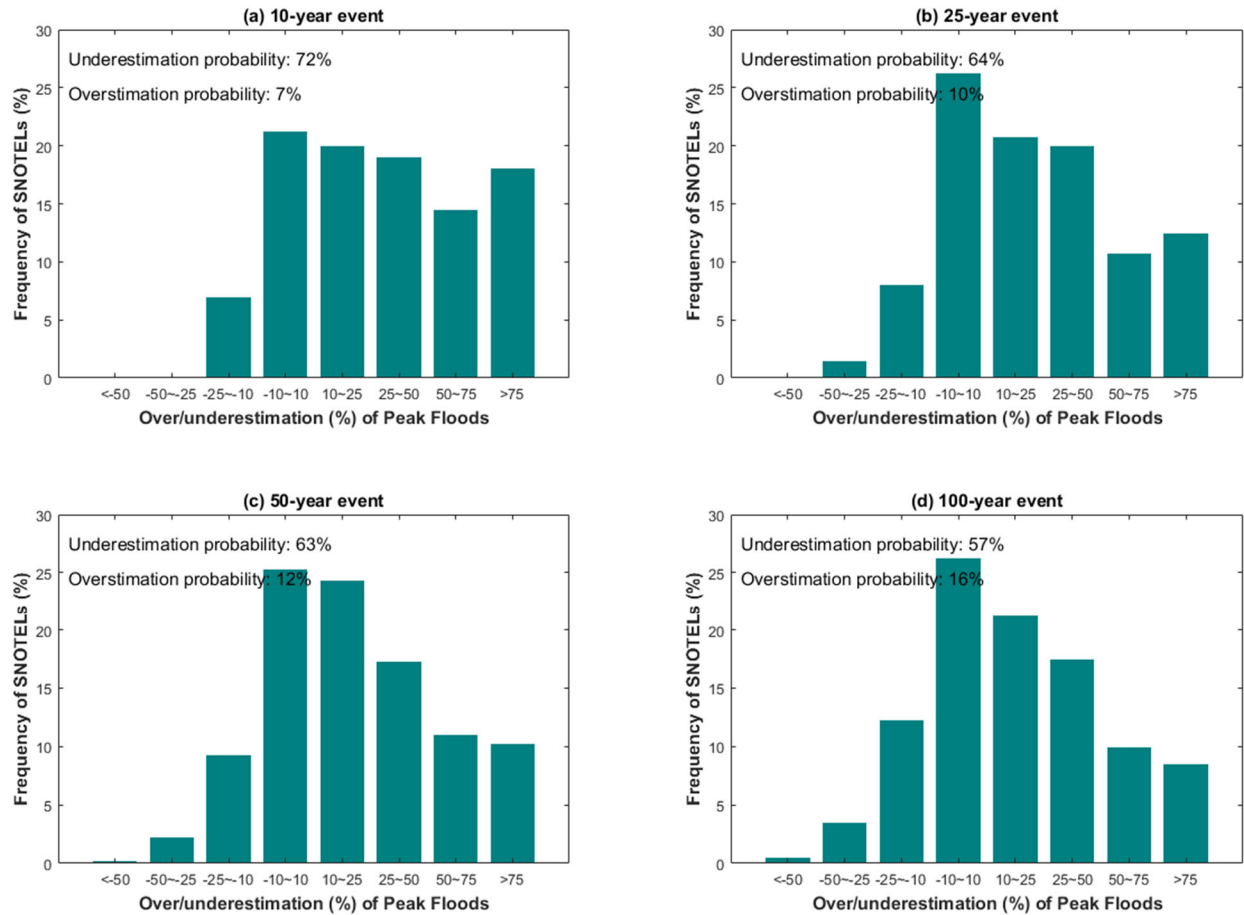


Figure 13. Histogram of the over/underestimation (%) of peak floods for four return periods over all SNOTELs in the western United States. Negative values represent overdesign and positive values represent underdesign. Underdesign or overdesign probabilities are also presented. Flood peaks within -10 to 10 % are considered as properly designed.

Based on Figure 12, it is also observed that the errors showed smaller differences in the maritime regime (Cascade Mountains and Sierra Nevada Mountains) than in the intermountain (Blue Mountains, Eastern Cascades, and Northern Basin and Range) and continental regimes (Colorado Front Range, Wasatch Mountains, and Rocky Mountains). We further explored the climate factors for controlling this spatial heterogeneity and found, as expected, that this spatial pattern was associated with the climate variability across the western United States, which mainly included air temperature, solar radiation, and atmospheric humidity (Yan et al., 2019c).

We also used the SNOTEL observations from 1979–2017 to examine recent, climate-related changes in snowpack properties and the frequency of ROS events. Figure 14 presents the results of the RKT for the 11 ecoregions across the western United States. The RKT results suggested significant trends toward declining April 1 and annual maximum SWE and earlier peak SWE data (i.e., earlier snowmelt). These significant trends coincided with a widespread increasing trend in winter mean temperature. Specifically, 7 and 8 of the 11 ecoregions showed statistically significant declining trends in April 1 and annual maximum SWE, respectively; 7 of the 11 regions suggested significant trends in earlier onset of melt (i.e., inferred from the earlier annual

peak SWE date). All 11 ecoregions showed significant snowpack loss and/or earlier snowmelt, except for the North Cascades region. The insignificant trends in this region were likely related to the relatively large increases in winter precipitation that may offset the effects of warming on snowmelt. Our results are consistent with Mote et al.'s (2018); they also reported widespread declines in April 1 SWE using model simulations in the western United States from 1955 to 2016. Our results are also consistent with the observed and projected trends reported by Clow (2010) and Musselman et al. (2017). Clow (2010) used SNOTEL SWE and USGS streamflow data from the 1978–2007 period in Colorado and found strong, pervasive trends in snowmelt and streamflow timing, which have shifted toward earlier in the year by a median of 2–3 weeks over the 29-year study period. Musselman et al. (2017) used WRF projections and found a contraction of the snowmelt season to a time of lower available energy, leading to earlier and slower snowmelt across the western United States in a warmer world.

Based on Figure 14, it is also observed that 7 of the 11 regions showed significant increasing trends in ROS frequency, and the trend was most significant in the northwestern United States or the Pacific Northwest. The increasing trends in ROS frequency ranged from 0.34 to 2.0 events per decade, and the average change in the northwestern United States was 1.43 events per decade. Over the northwestern United States, this result suggests an increase of 5.6 ROS events from 1979 to 2017 or a 32% increase relative to the regional mean annual ROS events for that period. These observed changes in snowpack have significant implications for water resources management, potentially creating a mismatch between water availability and need under future climate conditions (Sturm et al., 2017; Yan et al., 2019c). In Section 4.6, the changes in snowpack and W under future climate conditions are examined with the use of climate model projections.

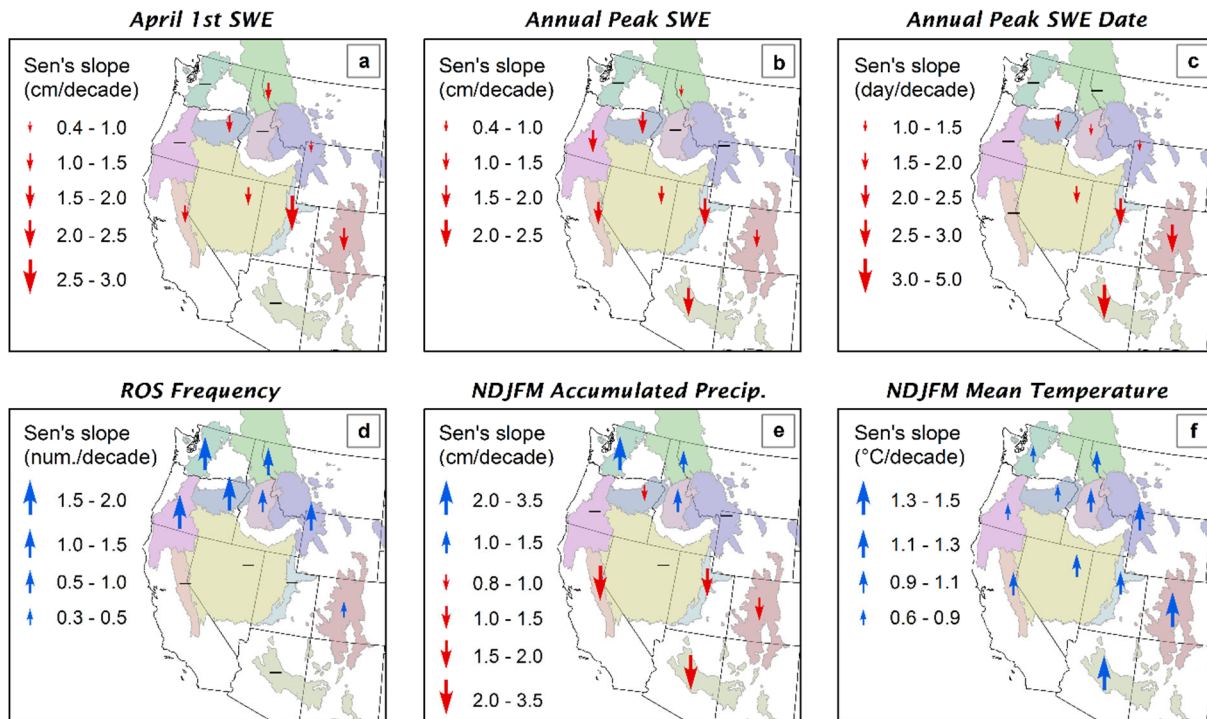


Figure 14. Regional trends at the 5% significance level for the snowpack, precipitation, temperature, and ROS events (see Yan et al., 2019c).

4.2 DHSVM Snow Modeling and Validation

We first validated the enhanced DHSVM snow model in modeling open, forest, and canopy gap conditions (Figure 15). At both open and forest sites, the station SWE observations and the DHSVM grid-average simulation (i.e., assume a uniform snowpack over the grid) were compared. The SWE comparisons showed a very high correlation of coefficient greater than 0.98 and a bias about 0.1 mm at both sites. For the canopy gap site, the SWE was compared between the simulated gap snowpack only and the average SWE measured during the ablation period at three locations at the north edge, center, and south edge of the circular gap floor in the experimental site (Sun et al., 2018). Our results showed good agreement with a correlation of coefficient of 0.94 and bias of 10.4 mm. In summary, the results shown in Figure 15 demonstrated that the enhanced DHSVM was able to reproduce the ablation pattern at all three sites that had distinct vegetation cover, and is capable of representing the key processes that control the snow-cover dynamics in complex forested environments.

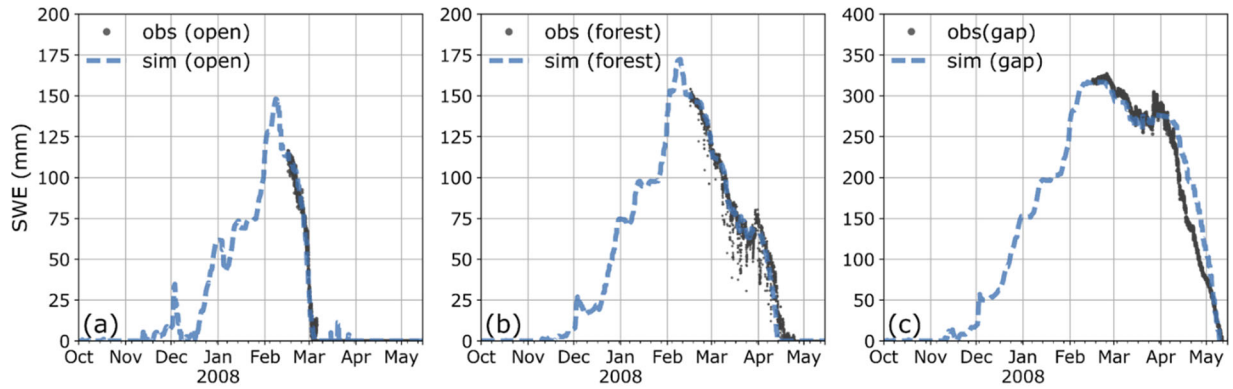


Figure 15. Enhanced DHSVM SWE modeling skill in complex forested environments including open, forest, and canopy gap conditions (see Sun et al., 2018).

After the DHSVM enhancement, we first identified the sensitive parameters of the DHSVM snow model and then developed regional snow parameters for the western United States based on measurements at 246 SNOTEL stations. We found at most stations (>90%) the simulations show sensitivity to the fresh snow albedo (α_{max}) and the albedo decay coefficient during melt (λ_M), and about 77% of the stations show sensitivity to albedo decay coefficient during snow accumulation (λ_A). These results suggested that snow albedo parameterizations are critical for modeling SWE dynamics in DHSVM. We also examined the two parameters used to partition precipitation into rain and snow. Compared to the minimum temperature for precipitation to be in the form of rain (T_R), which only showed sensitivity at about 12% of the stations, the maximum temperature for precipitation to be in the form of snow (T_S) was found to be sensitive at about 47% of stations. We also found that most stations sensitive to T_S were located in the maritime regime such as the Pacific Northwest as well as other lower elevations of mountain ranges like the low-elevation areas of the Sierra Nevada. As might be expected, most of the SNOTEL stations at high elevations (>1500 m) characterized by lower winter temperatures (less

than or equal to -4°C) showed no or low sensitivity to the T_S parameter. Therefore, we developed the regional snow parameters for the four sensitive parameters: α_{max} , λ_M , λ_A , and T_S .

With the 10,000 ensemble simulation, we first explored the single best-fit parameter set that produced the highest Nash-Sutcliffe efficiency (NSE) values at each SNOTEL station. NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance. The value of NSE varies from $-\infty$ to 1 and a value of 1 indicates a perfect fit between observations and simulations. These results confirmed that a simple mass and energy-balance snow model with the four sensitive parameters constrained by a physically reasonable range of values was able to capture the daily variability of SWE as well as annual peak SWE timing and magnitude under the diverse hydroclimatic conditions of the western United States (Sun et al., 2019). Note that only the open area condition was considered in developing regional snow parameters because of the nature of the SNOTEL sites. Figure 16 presents the DHSVM SWE modeling skill with the best-fit parameter set for the 246 SNOTEL stations across the western United States. In Figure 16, it is observed that the NSE-determined best-fit SWE simulations agreed very well with the daily SWE observations; the NSE value was greater than 0.8 at about 91% of the stations and the mean NSE value was 0.9.

Figure 17 further illustrates the SWE simulation results with the value of NSE close to 1 across different hydroclimatic regions across the western United States. We also used three other metrics—PEAK.ERR (mean percentage error of annual peak SWE), PDATE.ERR (mean absolute difference in timing of peak SWE in days), and COV.ERR (mean differences in snow cover duration in days)—to evaluate the model skill relative to the magnitude and timing of peak SWE and snow cover duration. We found that about 88% of the stations have a PEAK.ERR within $\pm 15\%$, about 88% of the stations have a PDATE.ERR within 14 days, and about 74% of the stations have a COV.ERR within 14 days (Figure 16). Only eight exceptions (NSE < 0.6) were found at SNOTEL stations located in central Montana and northern Wyoming, states both characterized by strong winds. At these eight SNOTEL stations, the model underestimated the peak SWE, mainly because of the wind-induced precipitation undercatch, which cannot be captured in the SNOTEL raw data (Sun et al., 2019).

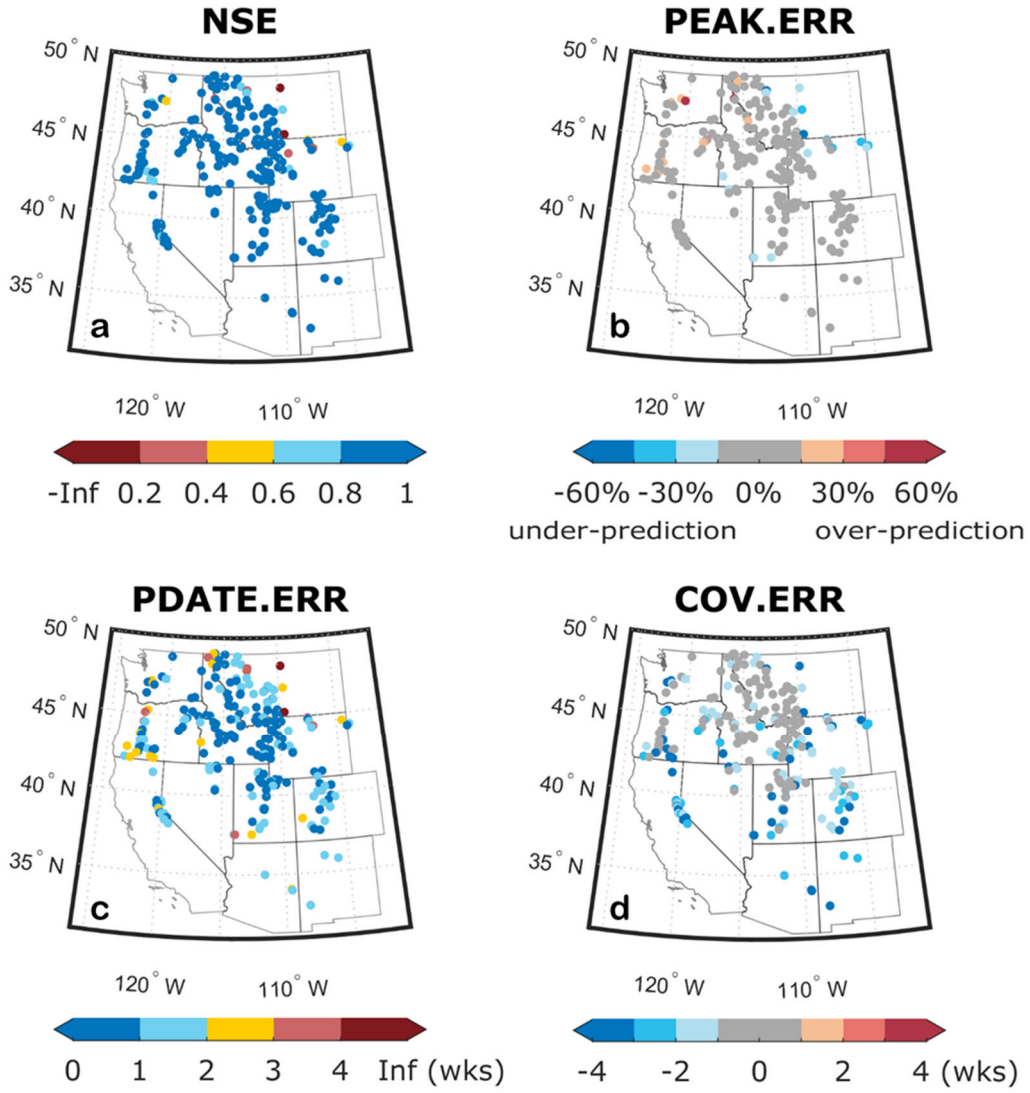


Figure 16. DHSVM SWE modeling skill with the best-fit parameter set for the 246 SNOTELs across the western United States. Four metrics are used to assess the model skill: the Nash-Sutcliffe efficiency (NSE), mean percentage error of annual peak SWE (PEAK.ERR), mean absolute difference in timing of peak SWE in days (PDATE.ERR), and mean differences in snow cover duration in days (COV.ERR) (see Sun et al., 2019).

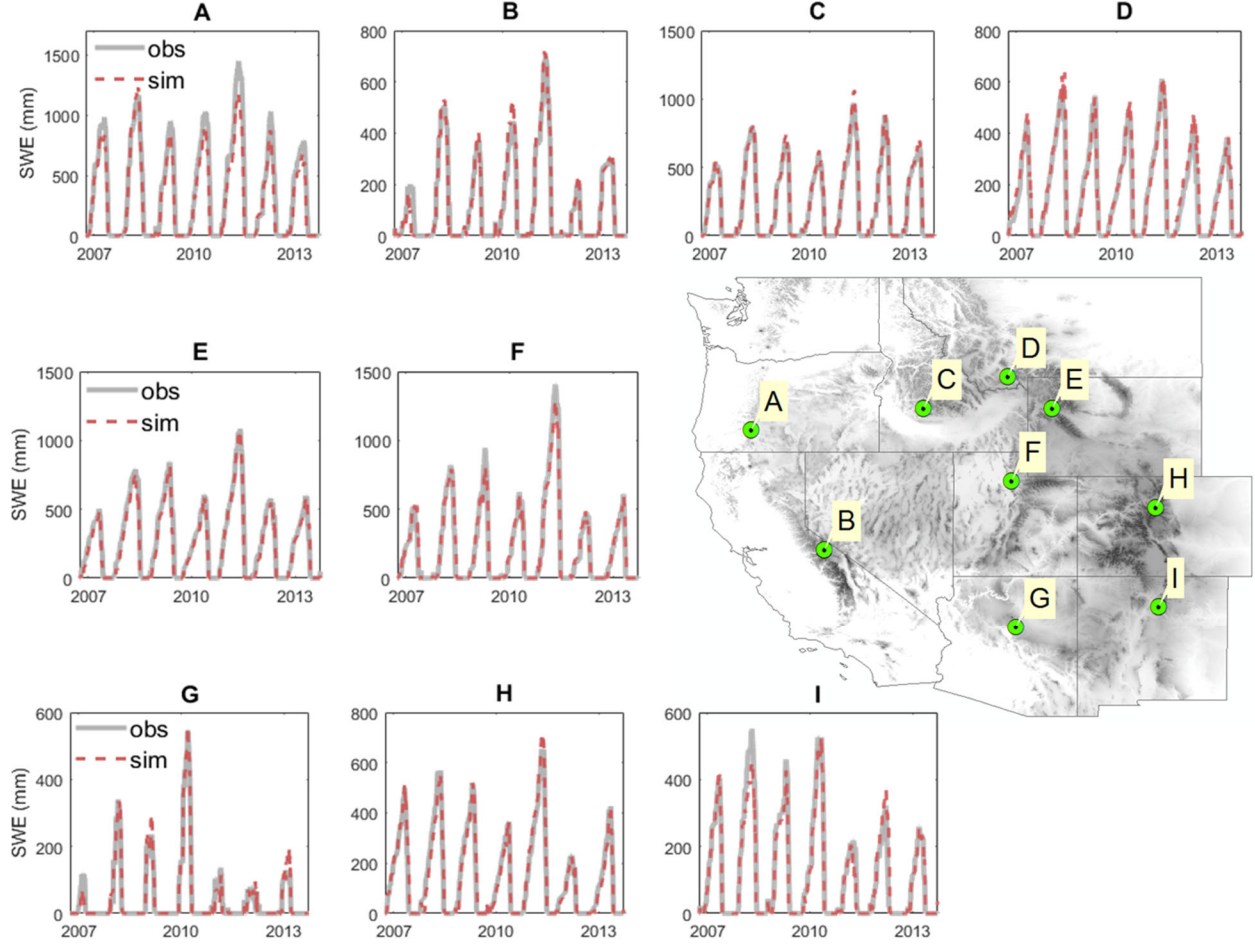


Figure 17. Illustration of DHSVM-simulated and observed SWE for selected SNOTEL stations across different hydroclimatic regions for which the NSE value is close to 1.

With the 10,000 ensemble parameter sets, we then developed the posterior distributions using the resampling criteria as $NSE \geq 0.8$, $PEAK.ERR \leq 20\%$, $PDATE.ERR \leq 14$ days, and $COV.ERR \leq 21$ days. We next developed the regional parameter set using the ensemble mean over the posterior parameter space over all stations within a region. Note that the western United States was divided into five parameter regions based on winter mean precipitation, winter air temperature, and winter wind speed (over Nov–Mar). In our developed regional parameters, the α_{max} , λ_M , and λ_A showed limited spatial variability across all regions, while the T_s showed large spatial variability. We found this large spatial variability was largely related to humidity: a lower relative humidity can enhance evaporative cooling through latent heat exchange, thereby enabling a snow crystal to maintain its frozen state in above-freezing temperatures (Yan et al., 2019c). After developing the regional parameters, we applied them to daily SWE simulation at each individual station and summarized the performance metrics by region.

Model performance in simulating SWE using regional snow parameters was evaluated by the same three metrics that compare the simulated and observed SWE: (1) NSE, (2) the mean percentage error of annual peak SWE (PEAK.ERR), and (3) the mean difference in the timing of annual peak SWE (PDATE.ERR). NSE can range from $-\infty$ to 1. In our application, NSE equal to

1 corresponds to a perfect match of the modeled to the observed daily SWE. The closer to 1, the more accurate the model is. $NSE = 0$ indicates that the model predictions are as accurate as the mean of the observed data. Based on reported literature values (Moriasi et al., 2007; Newman et al., 2015), $NSE > 0.65$ indicates satisfactory to very good hydrologic simulations. $PEAK.ERR$ and $PDATE.ERR$ can range from $-\infty$ to ∞ . $PEAK.ERR$ equal to 0 suggests no difference between modeled and observed annual peak SWE; $PDATE.ERR$ equal to 0 suggests no difference between modeled and observed timing of peak SWE. In our evaluation, $PEAK.ERR$ within $\pm 25\%$ and $PDATE.ERR$ within 2 weeks were considered reasonably good simulations of peak SWE.

Figure 18 presents the results of the validation of DHSVM in SWE modeling skill with regional snow parameters. In Figure 18, it is observed that the regional parameters were able to capture the snow processes at a majority of the stations. The simulations of daily SWE using the regional parameters have an NSE greater than 0.65 at about 80% of the stations, a $PEAK.ERR$ within $\pm 25\%$ at about 62% of the stations, and a $PDATE.ERR$ within 2 weeks at about 66% of the stations, respectively. Relative to the single best simulations, regional parameter simulations produced comparable NSE , $PEAK.ERR$, and $PDATE.ERR$ values (Sun et al., 2019). In general, these validation results suggest that areas that had a similar hydroclimate shared a similar snow regime and regionally coherent snow parameterizations were able to capture daily variations in SWE observations.

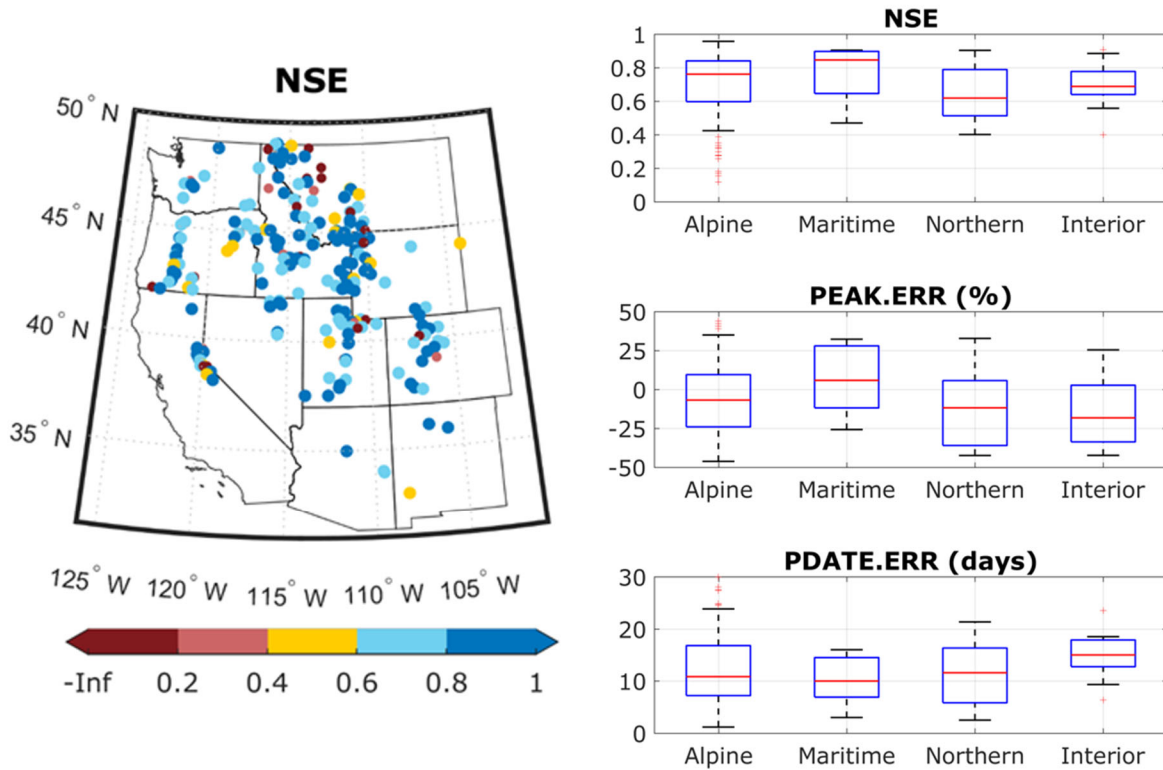


Figure 18. Validation of DHSVM in SWE modeling skill with regional snow parameters. The boxes in the boxplot show 25th–75th quantiles; the median value is indicated by the red solid line; the whisker length

is specified as the interquartile range (i.e., the difference between 75th and 25th percentiles), and data points beyond the whiskers are considered outliers.

4.3 Validation of NG-IDF Tool in Design Flood Estimates

Before implementing the NG-IDF curves on a real design problem, evaluation and benchmarking of the performances of the curves in design flood estimates are essential, because risks related to the United States infrastructure design will involve trillions of dollars in the coming decades (Carnevale and Smith, 2017). In the hillslope-scale validation, the DHSVM design flood estimates (q_{dhsvm}) were used as performance benchmarks to evaluate the peak design flood estimates from both PREC-IDF (q_P) and NG-IDF (q_W) curves. Figure 19 presents the scatterplots of q_P or q_W versus q_{dhsvm} for the selected return periods or ARIs. Our results suggested that for all three selected return periods, q_P produced much larger errors than q_W in regard to q_{dhsvm} at most of the 246 stations. We found about 73%, 69%, and 66% of the 246 SNOTEL stations had errors between q_P and q_{dhsvm} larger than $\pm 20\%$ for the 10-, 25-, and 50-year events, respectively. With the use of the NG-IDF curves, the percentage of stations that had errors larger than $\pm 20\%$ was reduced significantly to 18% for the 10-year event, 15% for the 25-year event, and 13% for the 50-year event (Yan et al., 2020). These differences were mainly attributed to the exclusion of the snowmelt process in the PREC-IDF developments.

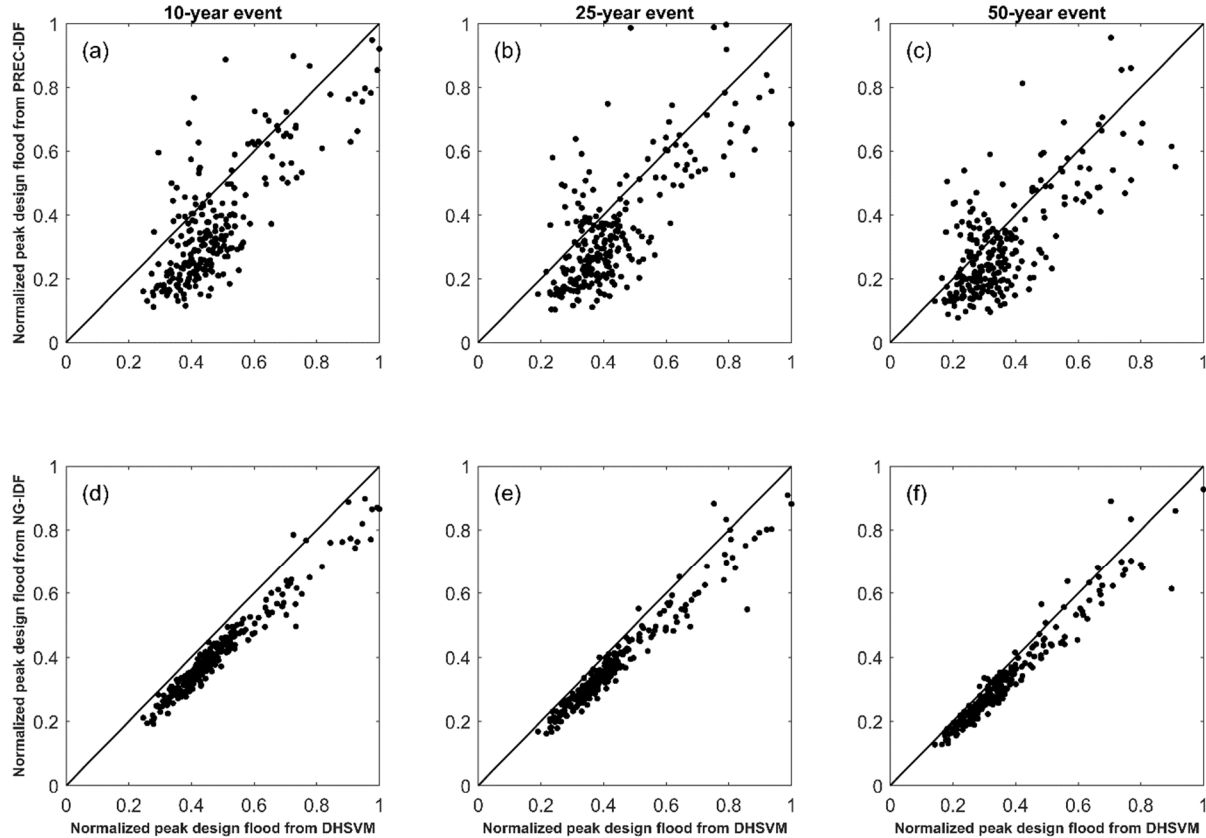


Figure 19. Scatterplots of the normalized peak design floods from PREC-IDF curves (top) and NG-IDF curves (bottom) versus normalized design floods from DHSVM continuous simulation for the 246 SNOTEL stations across the western United States (see Yan et al., 2020).

Figure 20 further presents the relative errors spatially between q_P or q_W and q_{dhsvm} across the western United States. In Figure 20, it is observed that considerable underestimation of design floods (with q_P) occurred in the continental regime (Colorado Front Range, Wasatch Mountains, and Rocky Mountains) and intermountain regime (Blue Mountains and Northern Basin and Range). This is because the magnitude of extreme precipitation events is smaller than the extreme snowmelt events in these regions (Yan et al., 2018). Note that a few stations also showed overestimation when using the PREC-IDF curves, and these stations are primarily located in the southern Sierra Nevada and Arizona/New Mexico. This is because in these regions, the PREC-IDF values were significantly higher than the NG-IDF values (Yan et al., 2019b). In the southern Sierra Nevada, the higher values of PREC-IDF than NG-IDF were caused by large winter precipitation and slower melt in spring related to lower humidity (Yan et al., 2019c); while in Arizona/New Mexico, the overall warm conditions in winter led to shallow snowpack accumulation and very small snowmelt events in mid-winter to early spring (Yan et al., 2018). In Figure 19, it is noted that compared to q_{dhsvm} , q_W suggested a consistent underestimation of design floods across all stations. In our TR-55 modeling, we used the wet initial condition and critical design duration approach to avoid the potential underestimation of design flood; this tendency of underestimation was most likely attributed to the use of a uniform hyetograph in TR-55 modeling. In general, a design hyetograph that had a pronounced peak produced a larger flood peak and these results suggested that it is important to develop W hyetographs to further improve NG-IDF performance.

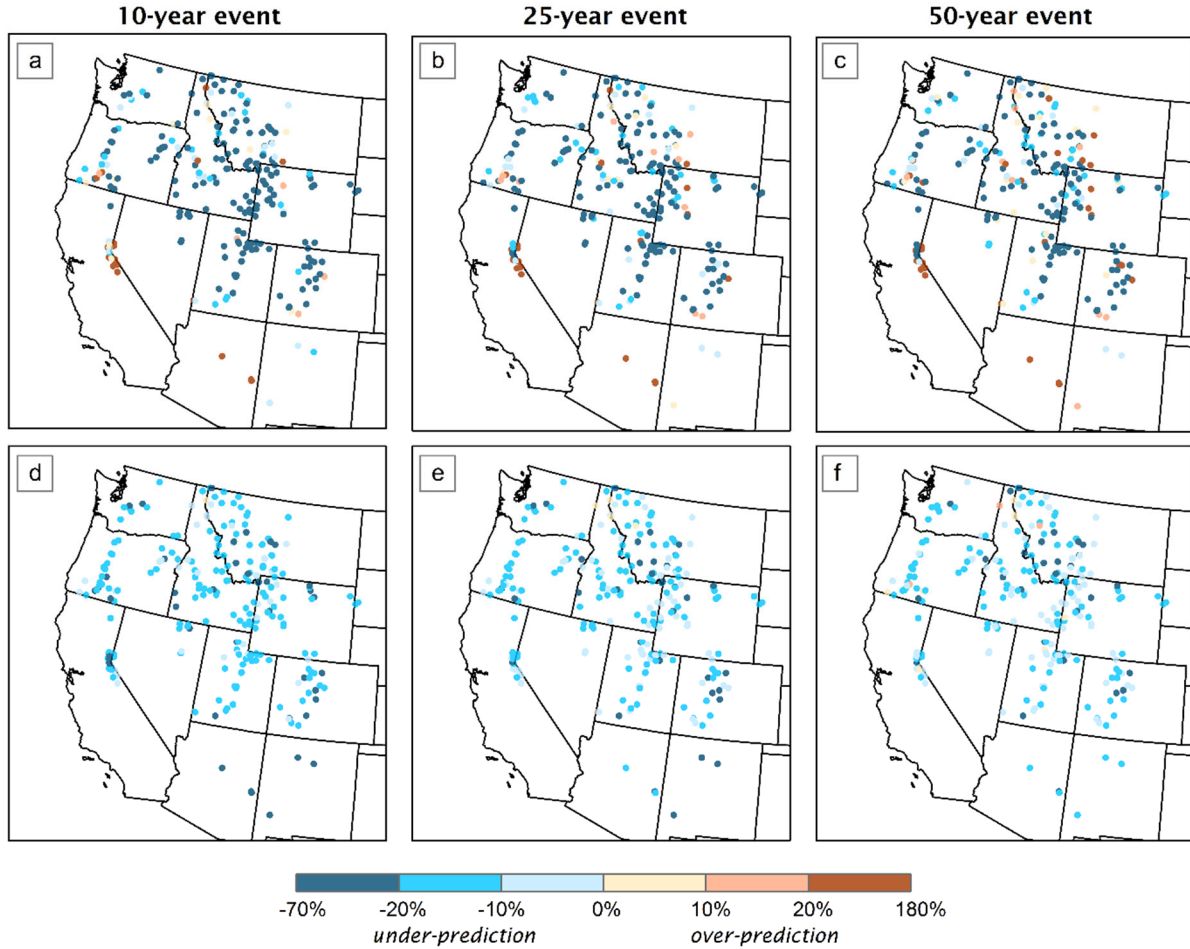


Figure 20. Relative peak design flood errors between the estimates from PREC-IDF curves (a, b, and c) and NG-IDF curves (d, e, and f) versus the DHSVM continuous simulations (see Yan et al., 2020).

Using the hillslope applied to 246 SNOTEL stations, we provided a systematic evaluation of NG-IDF curves over the mountainous regions of the western United States that had no vegetation cover. While the experimental hillslope itself had uniform bare soil cover conditions, in a real design problem, the watershed of interest has multiple types of land cover, and streamflow is generated by a variety of surface and subsurface flow processes. In addition to the validation at hillslope scale, we provided the results for NG-IDF validation for a real basin-scale location that contained a DoD site. Figure 21 presents the DHSVM performance in both calibration and validation periods for the UWWB. In the calibration period, the DHSVM achieved NSE values of 0.69, 0.70, and 0.78, and Kling-Gupta efficiency values of 0.80, 0.83, and 0.77 in streamflow and SWE simulation compared to the observations from the two USGS gauges and one SNOTEL station, respectively. During the validation period, we also achieved similar performance—NSE values of 0.60, 0.64, and 0.77, and KGE values of 0.75, 0.80, and 0.72 in both streamflow and SWE simulation, respectively. These results confirmed that the validation DHSVM in reproduced the observed streamflow and SWE in the whole UWWB. After the model validation, we extracted the simulated streamflow and W for the small test basin that flowed into the DoD MCMWTC site.

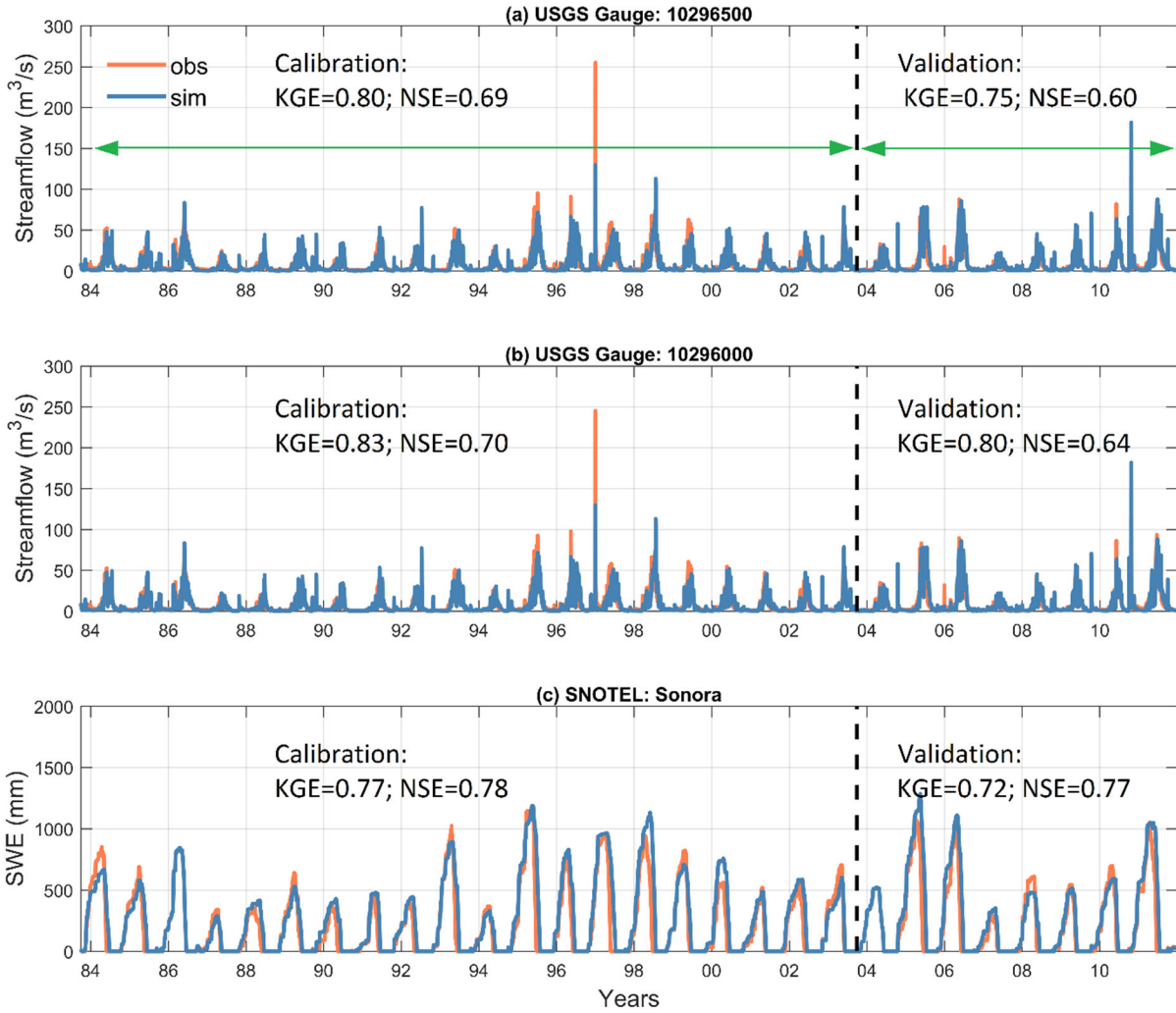


Figure 21. DHSVM calibration and validation results in streamflow and SWE for the Upper West Walker Basin.

Using the small basin mean 15-minute W time series, we first developed the W hyetograph for the test basin. Figure 22 presents the probabilistic W hyetograph for the small test basin. The data in Figure 22 represent the average of many large W events throughout the simulation period, illustrating the cumulative probability of occurrence in 20% increments. For example, the 10% of events in which W is concentrated closest to the beginning of the time period will have distributions that fall above and to the left of the 10% curve (more front-loaded). At the other end of the spectrum, only 10% of events are likely to have a temporal distribution falling to the right and below the 90% curve (more end-loaded). In Figure 22, an explicit diurnal cycle is observed compared to other standard rainfall hyetographs such as the NRCS hyetograph (McCuen, 1998). This is because snowmelt is generally greatest during daylight hours. The W hyetograph in Figure 22 also explains why the uniform hyetograph led to underestimation of design floods in the NG-IDF validation for the hillslope.

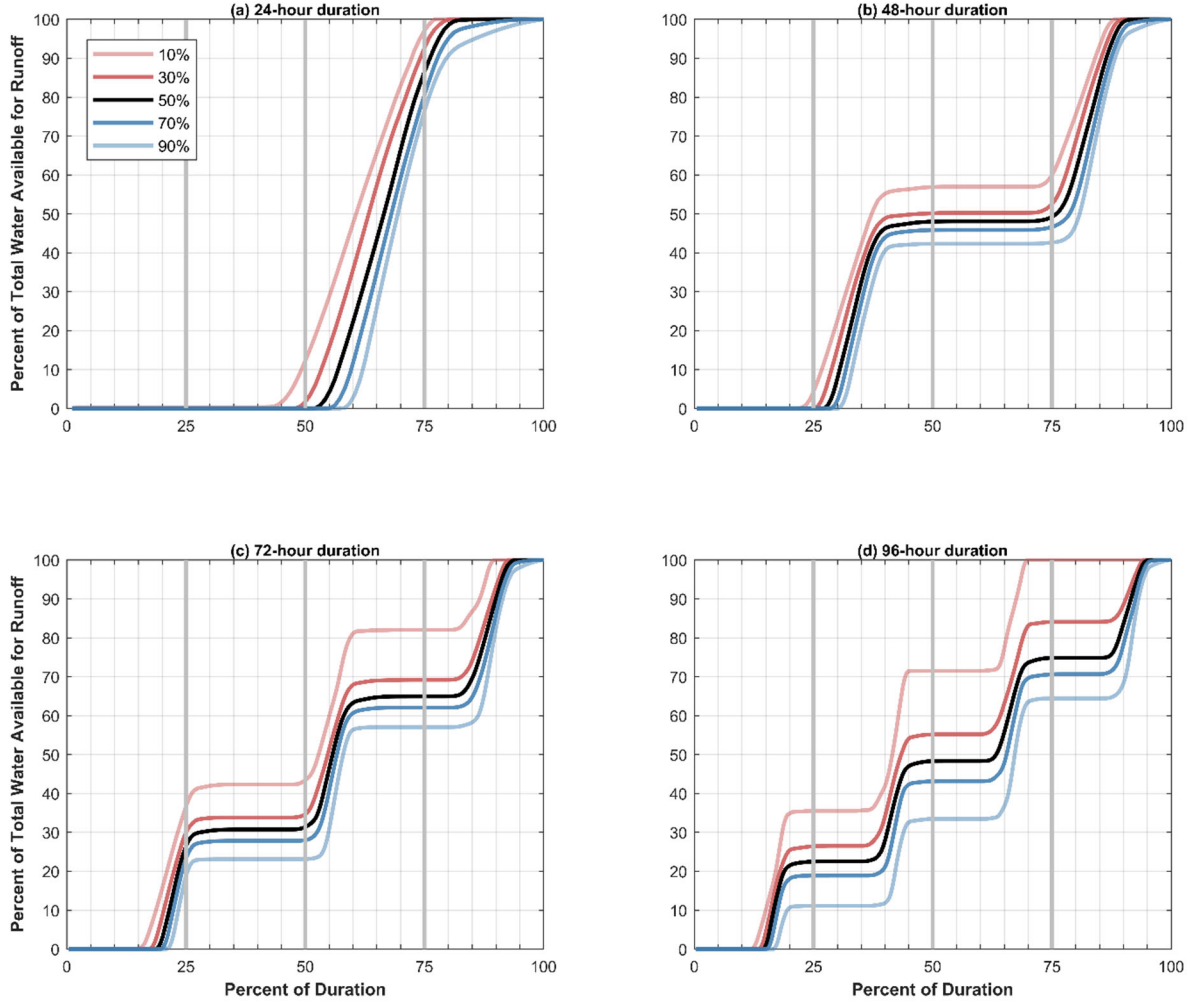


Figure 22. Probabilistic W hyetograph for the small test basin. The graph indicates the cumulative probability of occurrence in 20% increments.

Figure 23 presents the comparisons of design floods estimated from DHSVM continuous simulation and NG-IDF curves with the use of standard rainfall hyetographs and also the developed W hyetograph. For the W hyetograph, we used the median (50%) case and also the optimized case (varying the W hyetograph from 10% to 90%) and found the best estimates in design flood. As discussed in Section 3.3, an MC procedure was used to capture the sample data uncertainty in both flood and NG-IDF frequency analysis. A p-value for Z-statistics was also reported to test the statistically significant difference between the two estimates. In Figure 23, it is observed that the both uniform and NRCS rainfall hyetograph led to significant underestimation of DHSVM design floods with a p-value close to zero. The differences at the basin scale were much larger than the differences at the hillslope scale using the same uniform hyetograph. This is because the flood at the hillslope was mainly generated from subsurface flow with a long time of concentration (24 hours), while the flood at the basin was mainly from surface flow that had a short time of concentration (42 minutes).

In Figure 23, it can also be observed that with the use of W hyetograph, the differences between the two design flood estimates were significantly reduced. The optimized W hyetograph led to the relative errors between the DHSVM- and NG-IDF-derived design floods of less than 2% on average for the selected four return periods or ARIs. Even with the use of the 50% (median) case, we found statistically insignificant differences between the design floods at three return periods. The only exception was the 5-year event, but the error was about 20% between the two design floods. Given this hydrologically challenging context, Calver et al. (2009) recommended that errors within the 20% threshold are very good in practice. These results suggested the W hyetograph should be used together with NG-IDF curves (instead of the standard rainfall hyetograph) in design flood estimates.

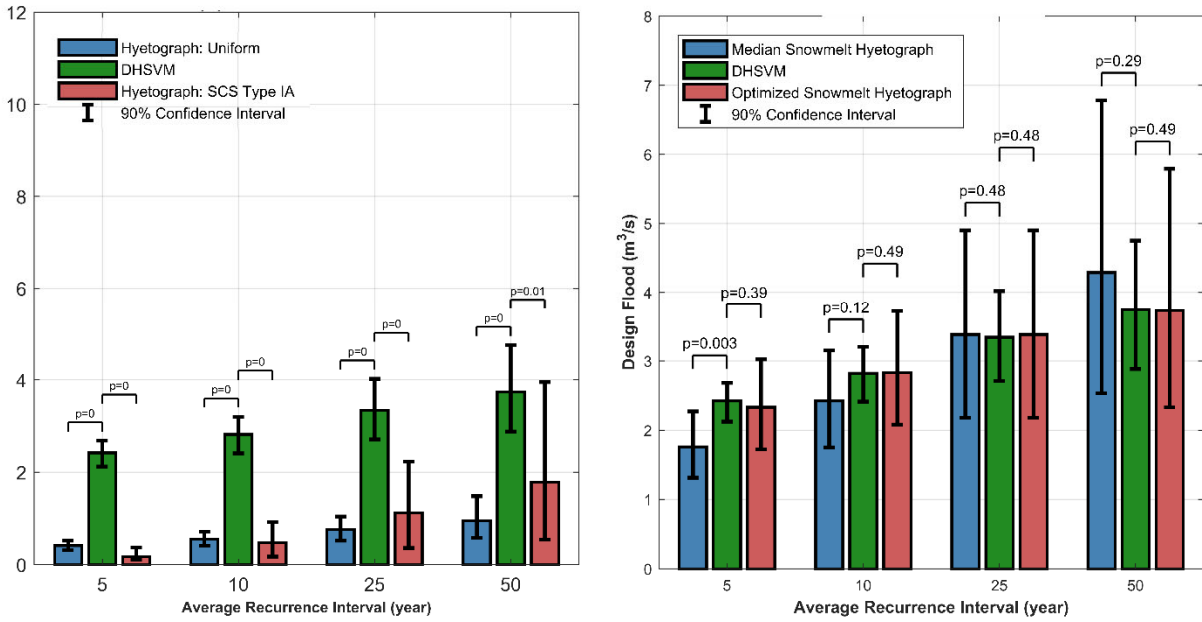


Figure 23. Comparison of design flood estimates derived from DHSVM and NG-IDF technology with the use of different hyetographs. Left: using the standard rainfall hyetograph. Right: using the developed W hyetograph. The bracket p-value represents the Z-statistics in significance test. Note the difference in the y-axis between left and right graphs.

4.4 NG-IDF Products over CONUS under Historical Climate

After validating both the DHSVM and NG-IDF tool (Sections 4.2 and 4.3), we developed the NG-IDF products using DHSVM simulations for CONUS at $1/16^\circ$ grid cell (~ 6 km) resolution. This required development of regional DHSVM snow parameters. Four sensitive snow parameters were identified (Sun et al., 2019) to be critical for capturing daily SWE dynamics. The four parameters consist of the albedo decay coefficient during snow accumulation (λ_A) and snowmelt (λ_M) as applied in the snow albedo decay curve of Laramie and Schaake (1972) for estimating snow surface albedo evolution, fresh snow albedo (α_{max}), and the maximum temperature for precipitation to be completely snowfall (T_S).

The CONUS was divided into five spatial clusters (Figure 24) based on the mean precipitation, air temperature, and wind speed during the general snow accumulation season from November

through March for the 1950–2013 period. We determined one set of “regional” parameters for each of the five clusters (Figure 25). For λ_A and α_{max} (Figure 25a, c), more than one cluster share the same parameter value, so there are only two distinct parameter values across the CONUS. Based on these regional parameters, we ran the DHSVM snow model for CONUS over 200,000 grid cells for the 1950–2013 period and developed NG-IDF and PREC-IDF curves at every grid location.

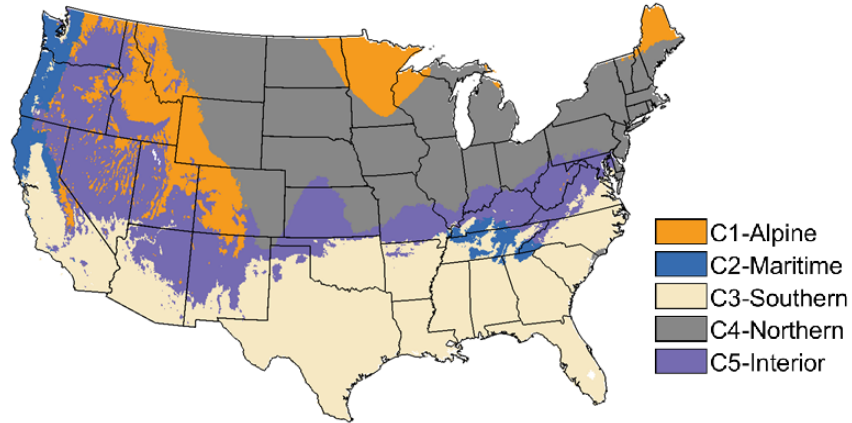


Figure 24. The CONUS is divided into 5 spatial clusters based on the mean precipitation, air temperature, and wind speed during the general snow accumulation season from November through March over 1950–2013.

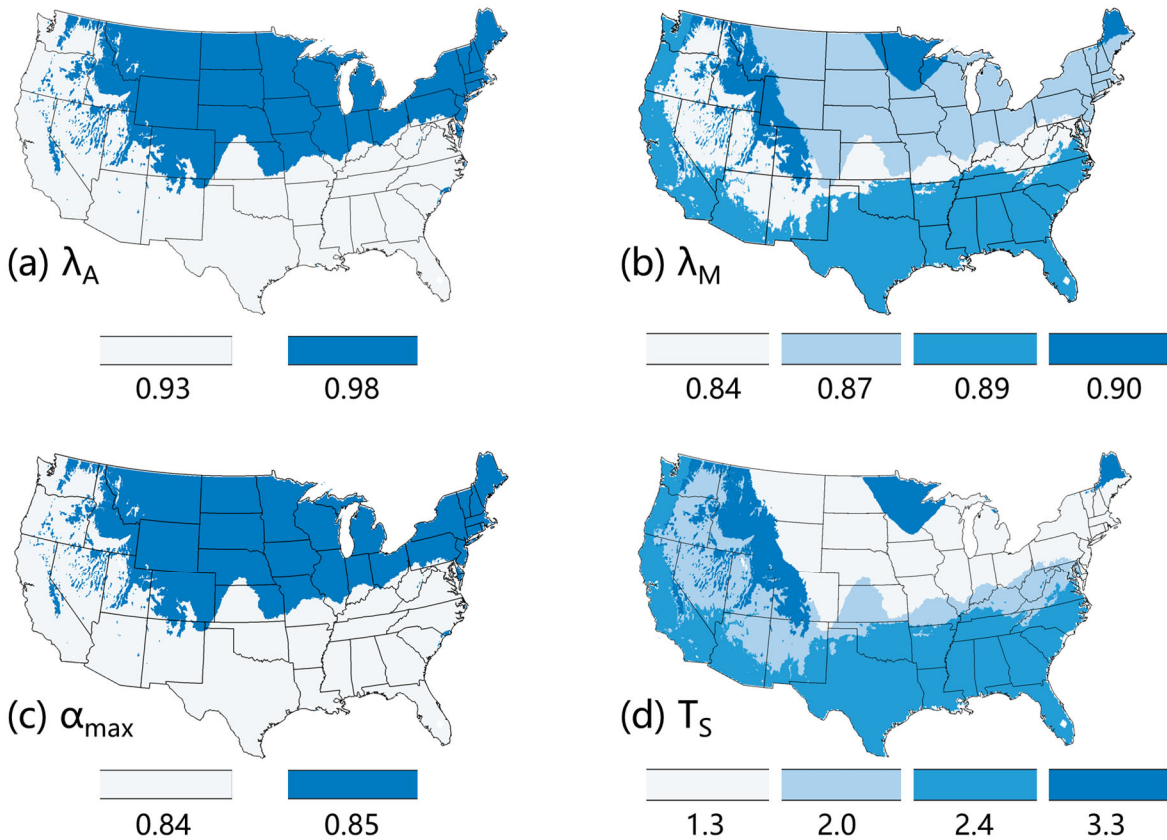


Figure 25. *Four regional DHSVM snow parameters over the CONUS. The CONUS is divided into five parameter regions based on winter mean precipitation, winter air temperature, and winter wind speed for 1950–2013.*

Figure 26 presents NG-IDF curve values under historical climate conditions across the CONUS for the three selected durations (24 hour, 48 hour, and 72 hour) and return periods or ARIs (10 year, 50 year, and 100 year). Figure 27 presents the dominant drivers of extreme events at each grid cell over CONUS. It is observed that snowmelt and ROS were dominant drivers of extreme events of 24-hour to 72-hour durations in the mountainous western United States and the Northeast Appalachians. The dominant mechanism shifted for different return periods and durations. In the Northwest, the 10-year extreme W events were dominated by rain and snowmelt (~70% locations), and snowmelt and ROS in spring dominated the 100-year events (~80% locations). In the Northeast, the 10-year events were dominated by snowmelt and rain (~80% locations), and ROS and rain-dominated 100-year events (~80% locations). Snowmelt dominated the extreme events for 24-hour to 72-hour durations in the snowbelts near the Great Lakes that were strongly affected by lake-effect snowstorms. In general, the NG-IDF W extremes were higher in magnitude in the rain-dominated Southeast and in the snowmelt- or ROS-dominated mountainous regions in the Northeast and western United States.

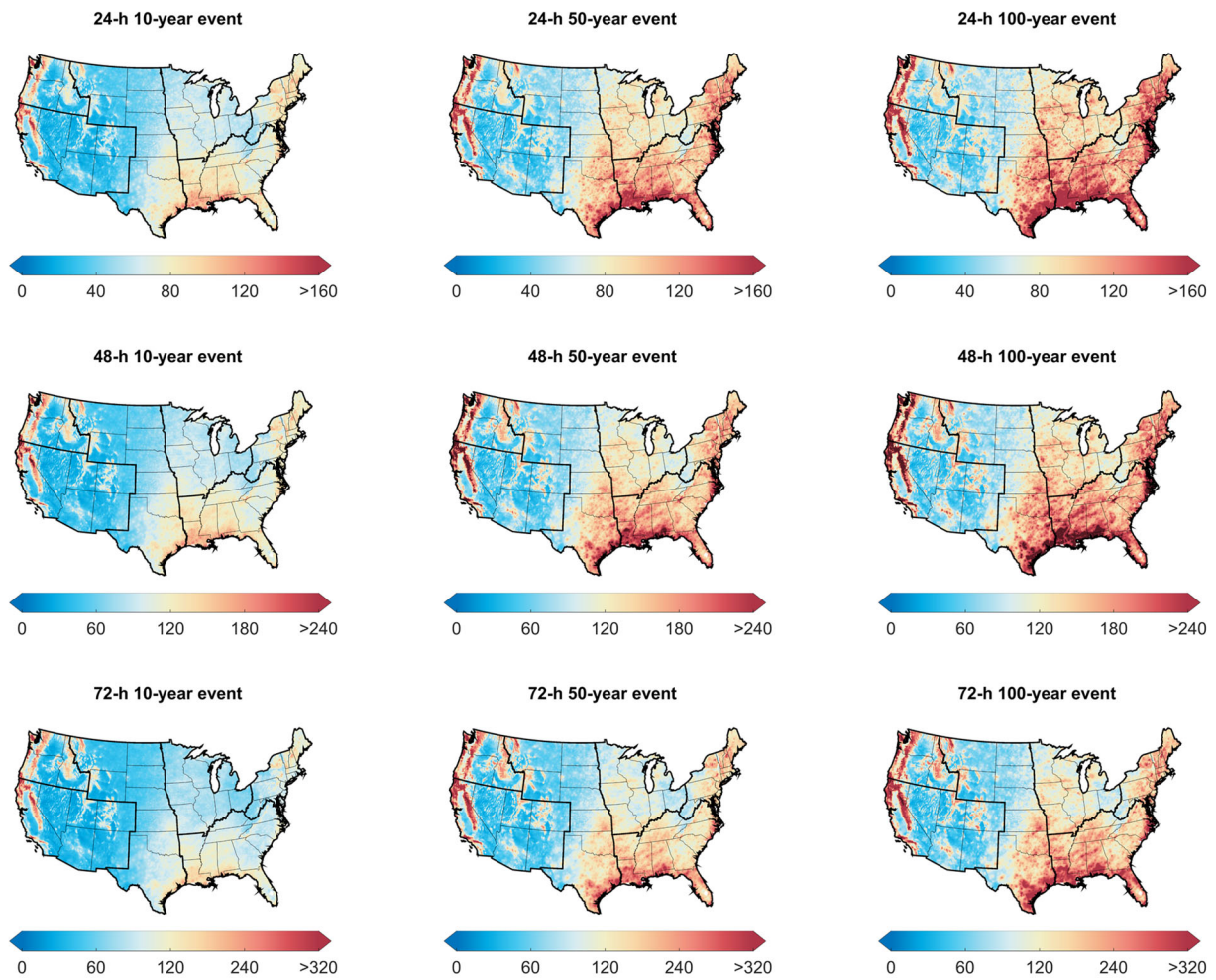


Figure 26. NG-IDF curve values for the selected duration and return periods over CONUS under historical climate condition from 1950–2013. Unit in the figure: mm.

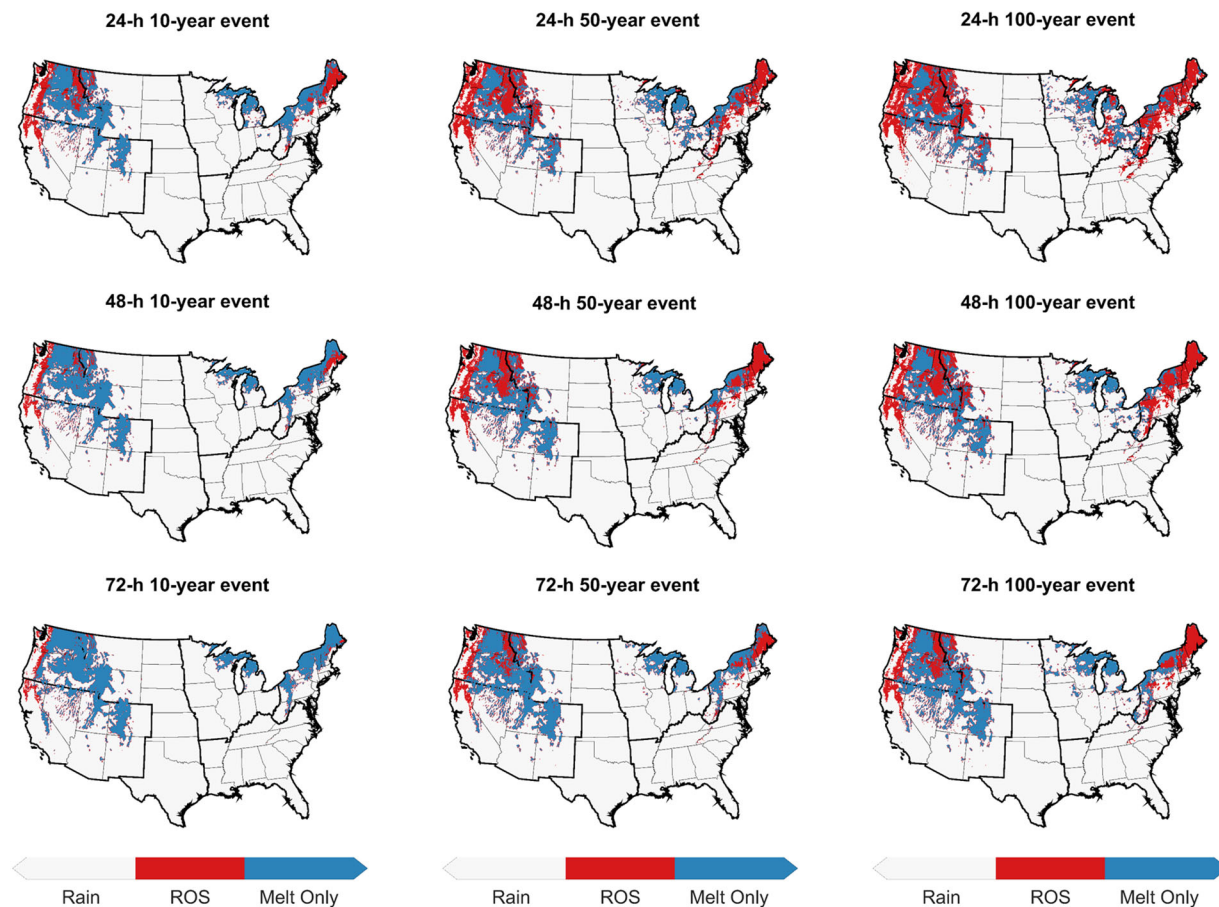


Figure 27. Dominant drivers of extreme events over CONUS under historical climate conditions from 1950–2013. The dominant driver was classified as rainfall, snowmelt only, and ROS.

We also compared the ratio of the NG-IDF curve value to the standard PREC-IDF curve value at each grid cell for the selected durations and return periods, and these results are presented in Figure 28. Specifically, the ratio was estimated as W/P at each $1/16^\circ$ grid cell across the CONUS. We evaluated the potential design risk for a given duration and return period for three cases:

- underdesign: if the W/P greater than 1.2;
- overdesign: if the W/P less than 0.8;
- proper design: if the differences between the two IDF curve values are less than 20%.

We used the 20% threshold for the proper design case to account for both data sample uncertainty in IDF frequency analysis and hydrologic modeling uncertainty. We acknowledge that this value is somewhat arbitrary and needs to be quantified explicitly in future studies. In Figure 28, we found that about 13%, 13%, and 15% of the total CONUS land area indicated the potential for underdesign for the 24-hour, 48-hour, and 72-hour events, respectively. Less than 1% of the CONUS area indicated the potential for overdesign. Consistent with the dominant extreme drivers shown in Figure 28, most of the locations subject to underdesign were in the

snowmelt- or ROS-dominated mountainous regions in the Northeast and western United States. Water resource managers and planners in these regions should pay special attention to the safety of infrastructure and potential snowmelt or ROS flooding events. We also archived NG-IDF products for historical products, which can be directly used as input to a single event hydrologic model to produce flood estimates.

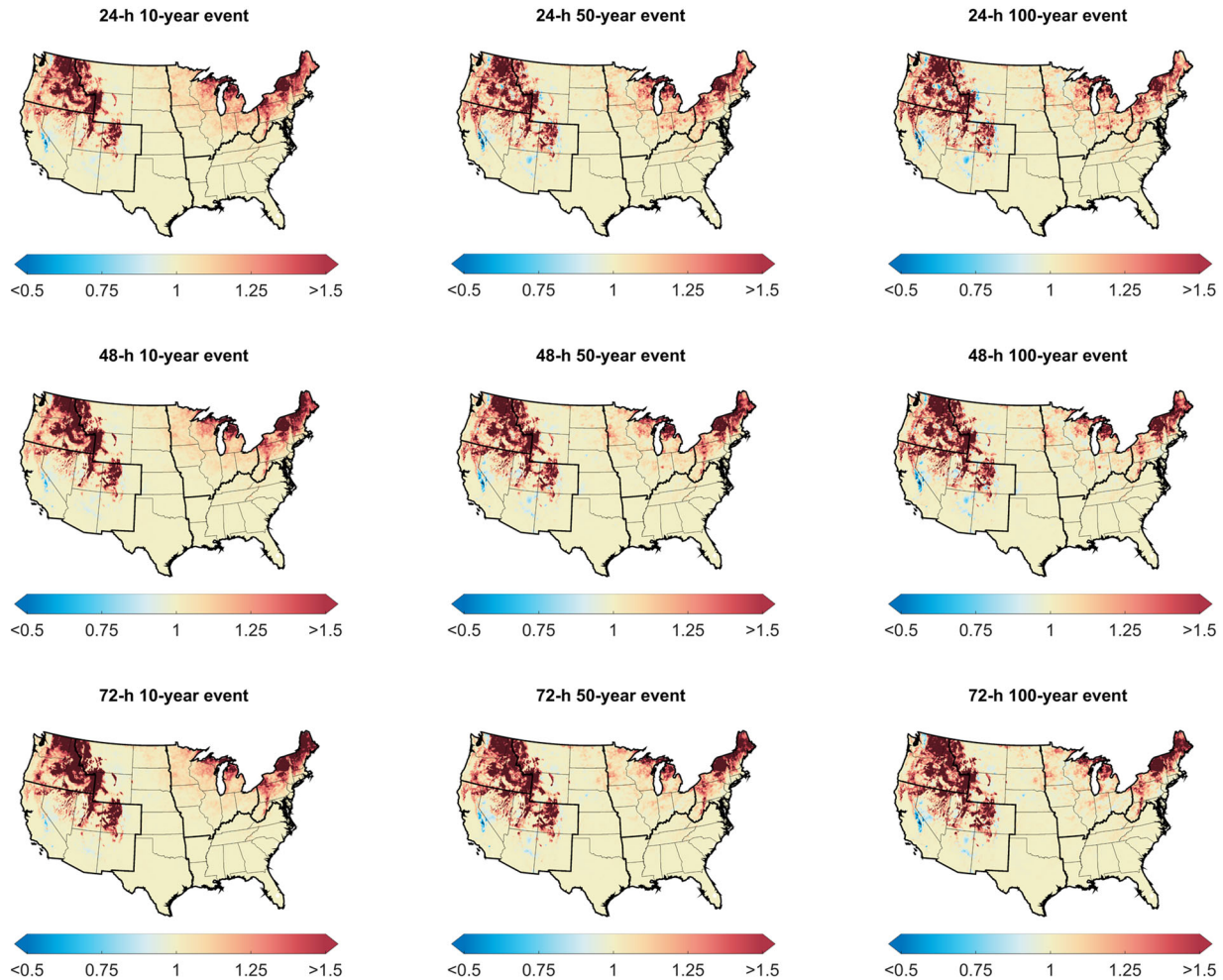


Figure 28. Ratio of NG-IDF curve values to PREC-IDF curve values for the selected durations and return periods over CONUS under historical climate conditions from 1950 to 2013.

4.5 Impacts of Atmospheric Rivers on Extreme Events

Before discussing the projected climate change impacts on NG-IDF curves, it is important to understand the physical-dynamical mechanisms responsible for these changes. Contrary to the statistically downscaled GCM data, which only provided for limited climate variables, the regional dynamical WRF simulations provided the opportunity to illustrate the physical mechanism responsible for the extreme events at relatively fine spatial resolutions. Here we focused on investigating the impacts of ARs on extreme events because extreme precipitation is often associated with ARs in the western United States (Chen et al., 2019a, 2018).

In this project, we used six AR-tracking algorithms to provide an improved understanding of the effects of AR events on extreme precipitation. The six algorithms included the Rutz, Gershunov, and Guan algorithms that identified the largest number of ARs, and the Goldenson and two PNNL methods (PNNL1 and PNNL2) that identified the lowest number of ARs (Chen et al., 2018). The details for each algorithm can be seen in Table S1 of Chen et al. (2018). Figure 29 presents the correlation between AR intensity-duration-area (AR-IDA) and the extreme precipitation amount at the monthly scale. The performance of different AR-tracking algorithms varied widely, but in general, the Goldenson and two PNNL methods performed better because fewer ARs identified were generally associated with more intense or prolonged ARs and they were more likely to be related to heavy precipitation. We also used other metrics such as Gilbert Skill Scores (GSSs) to compare the different AR-tracking algorithms (Figure 30). The GSS is a metric that evaluates how well the predicted events (hit or miss) correspond to the observations. Based on both GSSs and correlations, we concluded that the Goldenson and PNNL models performed better than the others.

In Figure 29, higher correlations (~ 0.5) were observed between ARs and extreme precipitation events across the western United States especially in the coastal regions, suggesting that ARs can provide useful predictability for extreme precipitation occurrence and magnitude in western United States watersheds. We also highlighted that AR-tracking algorithms with higher integrated vapor transport thresholds should be considered to further improve extreme precipitation skill.

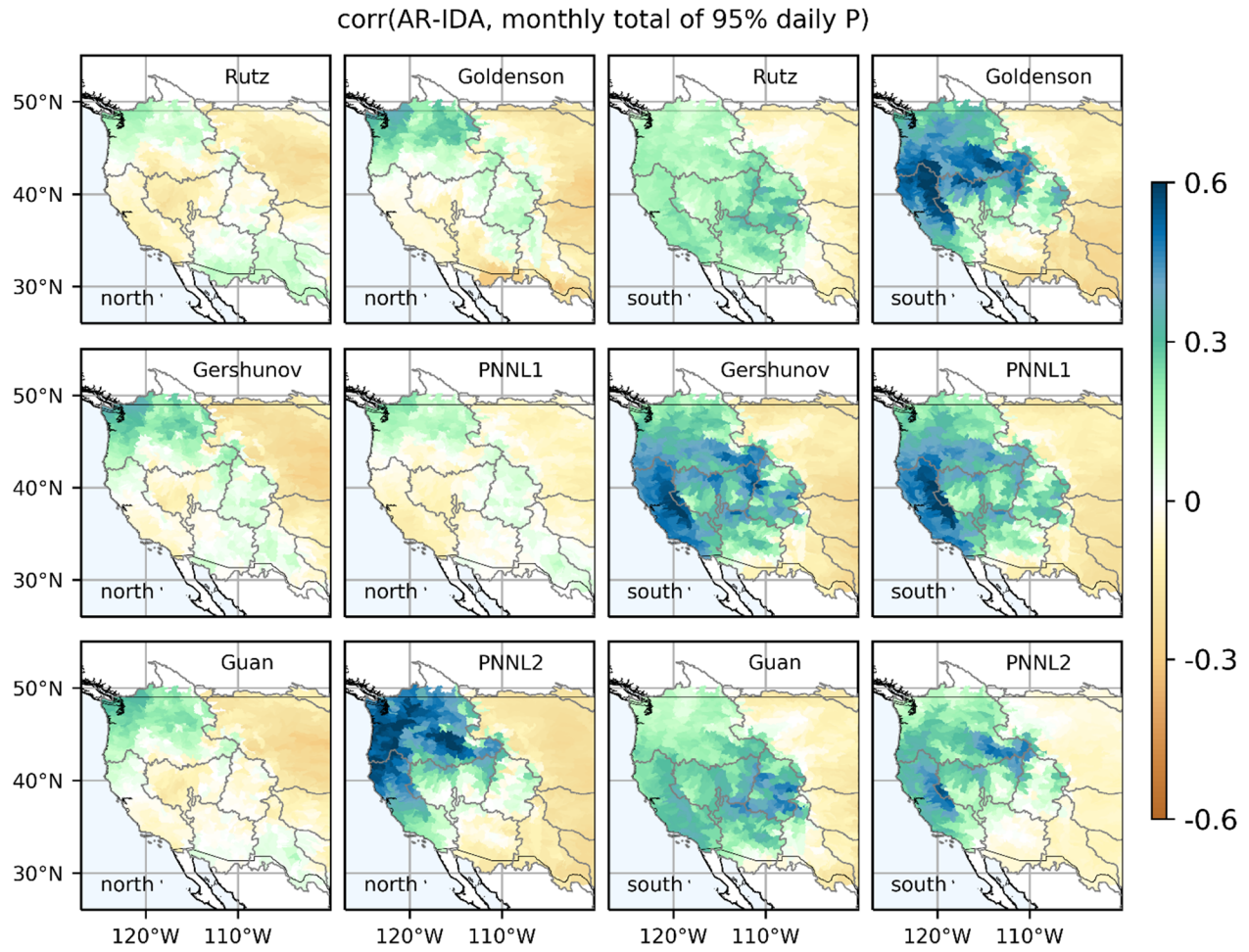


Figure 29. Correlation coefficients between AR intensity-duration-area (IDA) and the monthly total of extreme daily precipitation amount. The first and second columns are for ARs in the north (making landfall between 40°N and 60°N) derived from six AR-tracking algorithms. The third and fourth columns are for ARs in the south (making landfall between 25°N and 40°N) (see Chen et al., 2018).

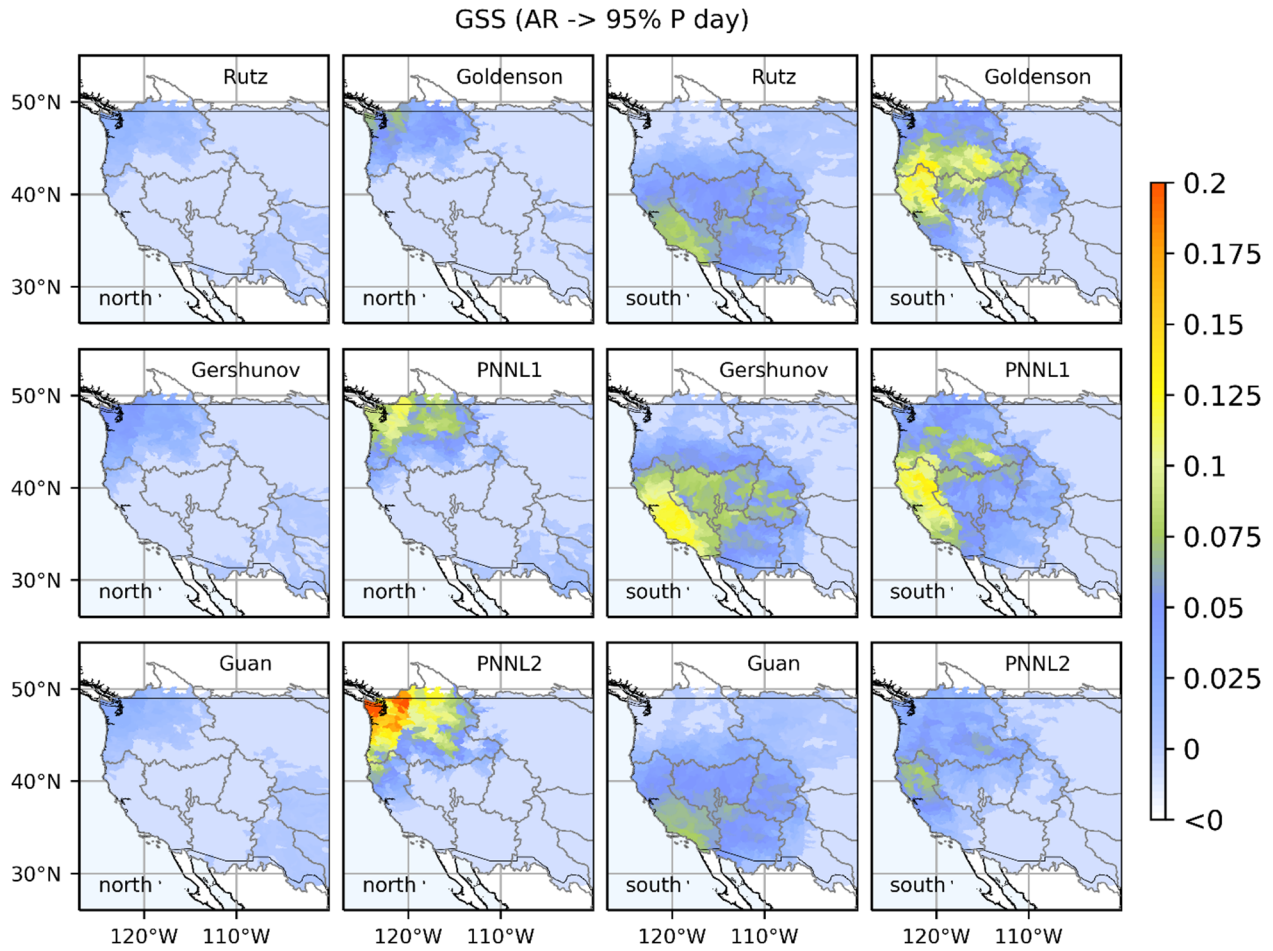


Figure 30. Gilbert Skill Scores (GSSs) of predicting the number of extreme precipitation days using ARs. A GSS of zero equals no skill and a GSS of 1 indicates a perfect score (see Chen et al., 2018).

In addition to extreme precipitation, we further investigated the impacts of ARs on watershed water budget using the WRF simulations in the western United States. Figure 31 presents the overall differences in major water budget components, including P, surface runoff, evapotranspiration (ET), and soil moisture between AR and non-AR events. As expected, precipitation was significantly higher during AR events. Surface runoff resembled the difference in precipitation, where coastal watersheds such as the western Cascades and the Sierra Nevada experienced the largest increase in runoff during ARs. In winter, when not limited by soil moisture, the ET difference mainly reflected the difference in total radiation (Chen et al., 2019b). In summer, the ET difference was controlled by radiation and/or soil moisture, depending on whether ET is energy- or moisture-limited in specific areas. The soil moisture increased across all areas during AR events for both the summer and winter seasons, but more spatial heterogeneity was observed for SWE. In summer, all precipitation events caused the snow to melt, but ARs led to moderately more snow ablation in the Pacific Northwest, which could be attributed to warmer temperature or increased longwave radiation. In winter, snow generally accumulated during winter non-AR precipitation events, while ARs led to widespread snow ablation (i.e., the mechanisms for ROS events).

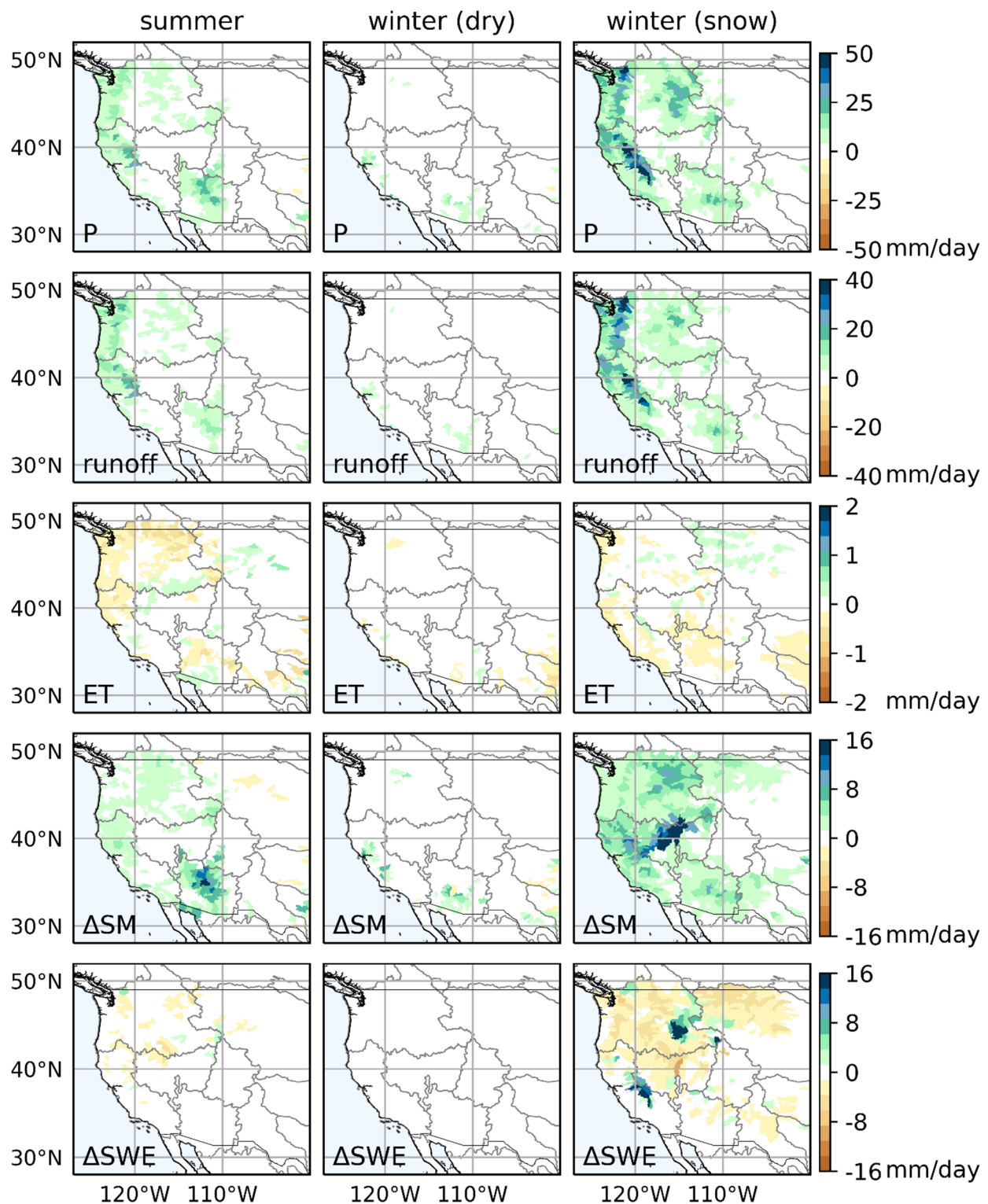


Figure 31. The difference in surface water budget associated with AR and non-AR precipitation events (AR minus non-AR). Top to bottom: precipitation, runoff, evapotranspiration (ET), soil moisture change (ΔSM), and SWE change (ΔSWE) (mm/day) (see Chen et al., 2019b).

Similar to the “water cycle” chart presented in hydrology textbooks, we developed a “water budget chart” here (Figure 32) to illustrate the AR impacts on surface hydrology averaged over the HUC 8 watersheds in the western United States. In Figure 32, it is observed that compared to non-AR events, ARs produced much higher precipitation but slightly suppressed ET. This substantially increased the water input to the surface hydrological system. During ARs, strong radiation and warm temperature increased the likelihood and the magnitude of ROS events, which caused enhanced snow ablation. On average, ΔSWE was -0.1 mm/day (melt) during AR events, while it was 1.46 mm/day (accumulation) during non-AR events. Our regression analyses attributed this differences in snow accumulation and ablation to the change of temperature/radiation during ARs. Specifically, warmer temperature caused snow ablation of 3.49 mm/day and increased net radiation triggered snow ablation of 0.97 mm/day during ARs. In summary, we identified two physical mechanisms through which ARs produced higher flooding events than non-AR events: (1) ARs produced higher P and modification of initial soil moisture, and (2) ARs enhanced ROS events that provided extra snowmelt water into soil.

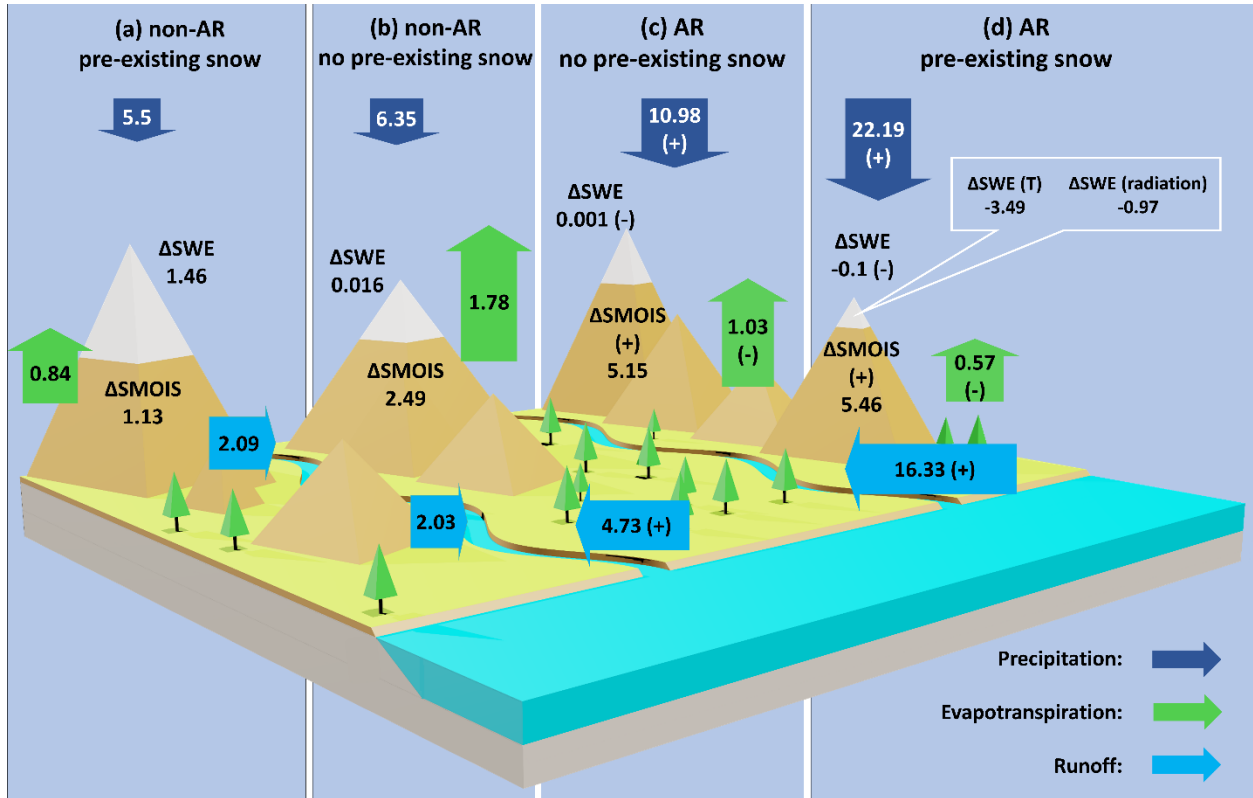


Figure 32. Illustration of the surface hydrological response to non-AR and AR precipitation events with and without pre-existing snowpack. Unit: mm/day. (See Chen et al., 2019b)

4.6 Climate Change Impacts on NG-IDF Products

To explore the projected changes in PREC-IDF and NG-IDF curves, we used both ensemble statistically and WRF dynamically downscaled GCM data to run DHSVM for both historical and future climate conditions.

1) Projected Changes over CONUS

Figure 33 summarizes the variability/uncertainty of the regional mean across the CONUS of changes in PREC-IDF and NG-IDF values across the selected 10 GCMs with boxplots. We show 24-hour and 72-hour, 100-year events as examples to represent the GCM uncertainty. The 10 GCMs were selected to represent divergence in future changes in temperature and precipitation for the RCP8.5 emission scenario for the CONUS. The 10 GCMs were bcc-csm1-1, BNU-ESM, CCSM4, CNRM-CM5, CSIRO-Mk3-6-0, inmcm4, IPSL-CM5A-LR, IPSL-CM5A-MR, MRI-CGCM3, and NorESM1-M. In Figure 33, it is observed that both W and P extreme events increased significantly under future climate conditions, and the Southwest (SW), Southeast (SE), and Great Plains (GP) showed the largest uncertainties in future projections.



Figure 33. Regional variability of future projections of extreme P and W events over CONUS using the data from the 10 selected statistically downscaled GCMs under RCP8.5.

Figure 34 presents the ensemble mean ratio of the NG-IDF curve values to the PREC-IDF curve values (in terms of W/P) for historical (1950–2005) and future (2044–2099) climate conditions

across CONUS using the data for the 10 selected statistically downscaled GCMs under RCP8.5. Similar to Figure 33, we used 24-hour and 72-hour, 100-year events as examples to show the change in the W/P ratio. We also evaluated the potential design risk for a given duration and return period for three cases:

- underdesign: if the W/P greater than 1.2;
- overdesign: if the W/P less than 0.8;
- proper design: if the differences between the two IDF curve values are less than 20%.

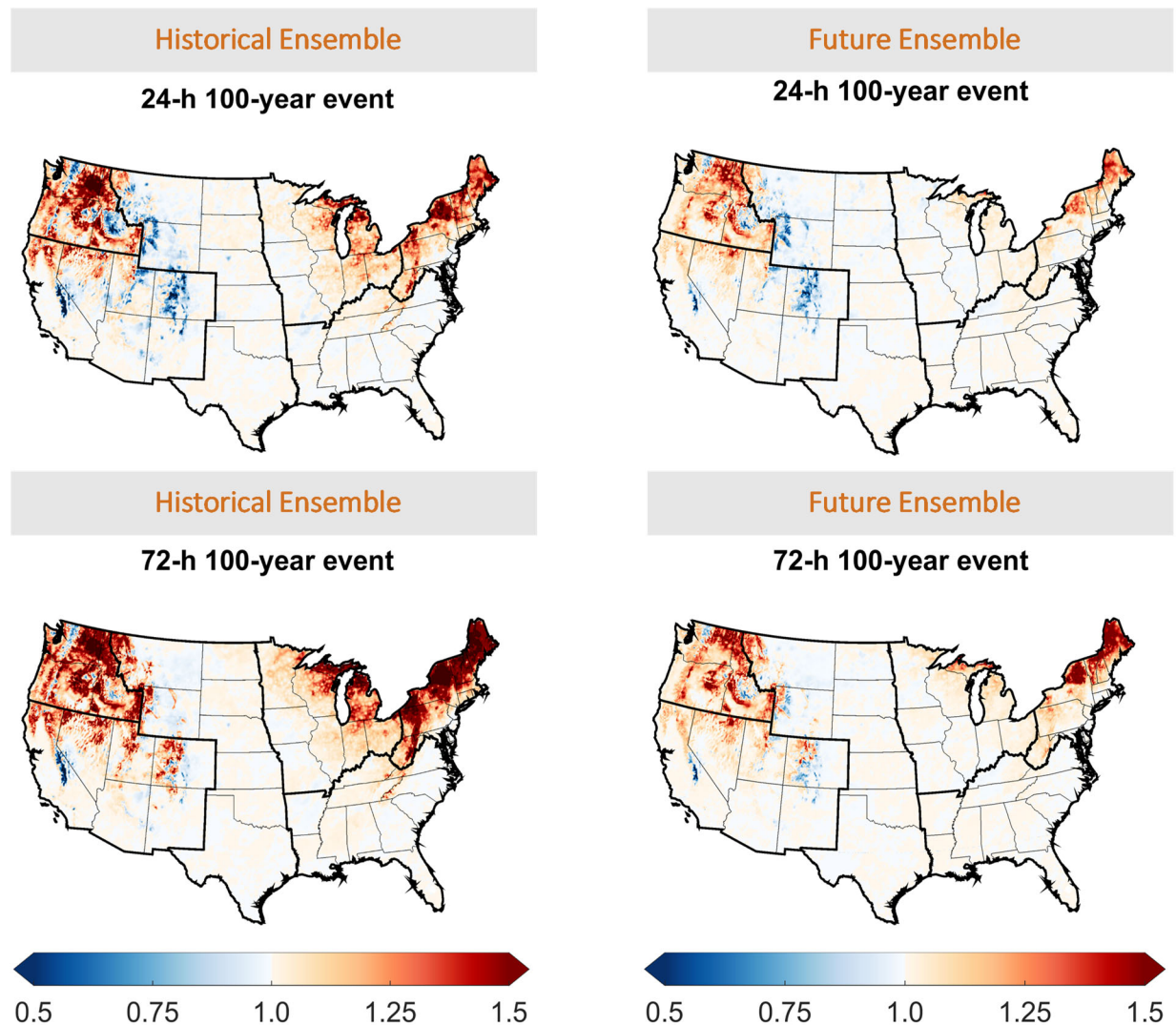


Figure 34. Ensemble mean ratio of the NG-IDF curve values to the PREC-IDF curve values (W/P) for historical (1950–2005) and future (2044–2099) climate conditions over CONUS using the data for the 10 selected statistically downscaled GCMs under RCP8.5.

For the 24-hour, 100-year event, the locations subject to underdesign (red color in Figure 34) decreased from 14% to 9% across the CONUS from the historical period to the future period. For the 72-hour, 100-year event, the locations subject to underdesign decreased from 23% to 9% of the CONUS from the historical period to the future period. The historical percentages are based

on the ensemble mean of the 10 downscaled GCMs over the historical period (1950–2005), and they differ slightly from those in Section 4.4 that are based on the meteorological data set of Livneh et al. (2013), which were developed based on historical observations from more than 20,000 NOAA stations. However, it is important to note that it does not mean the W or P value will decrease during the future period. As shown in Figure 33, under a generally warm future climate, the extreme P and W design events both increased in magnitude. The results in Figure 34 only show that there is a lower risk in infrastructure underdesign when comparing W extremes to P extremes. This is mainly due to the decreasing snowfall and snowpack (i.e., a shift in dominant extreme driver as shown in Figure 35) or earlier and slower snowmelt under a warming climate (Yan et al., 2019c).

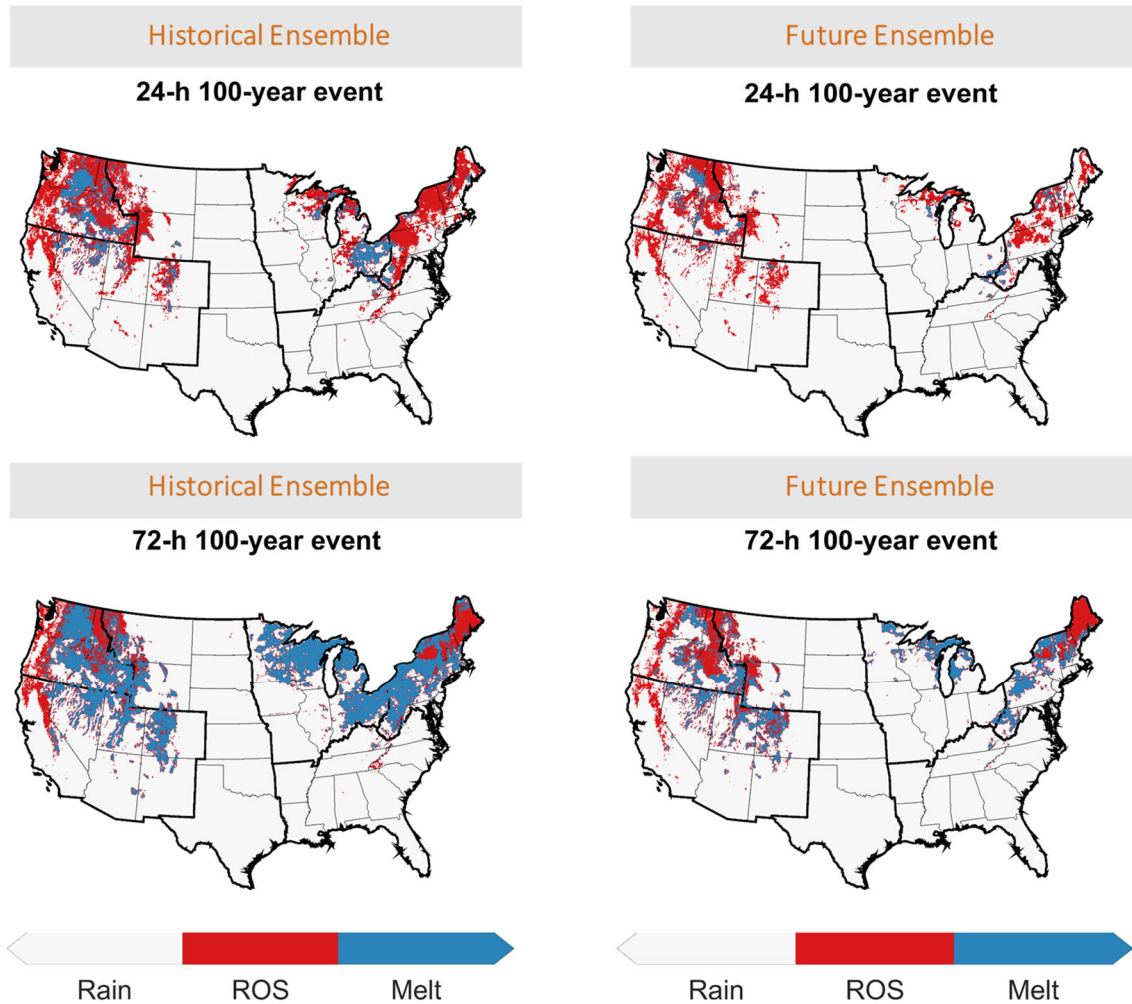


Figure 35. Dominant drivers of extreme events from historical (1950–2005) to future (2044–2099) periods over CONUS using the data for the 10 selected statistically downscaled GCMs under RCP8.5. Dominant drivers were estimated using the ensemble mean for each grid cell.

Figure 35 and Table 2 present the dominant drivers for the 14-, 72-, 100-year events for both historical and future climate conditions across CONUS. These results suggest that under future

climate conditions, extreme W events at more locations shift from melt-dominated or ROS-dominated to rain-dominated and from 24-hour to 72-hour durations.

Table 2. Dominant drivers of extreme events by percent of events over CONUS for historical and future climate conditions.

Climate	24-Hour, 100-Year			72-Hour, 100-Year		
	Rain	Melt	ROS	Rain	Melt	ROS
Historical	81%	6%	13%	75%	19%	6%
Future	89%	2%	9%	85%	9%	6%

2) Projected Changes at DoD Site

In addition to assessing the climate change impacts at the CONUS scale, we also used the WRF dynamically downscaled data to assess the climate change impacts on the DoD MCMWTC site. Figure 36 and Table 3 present basin-scale annual maximum W depths for historical (1981–2010) and future (2041–2070) climate conditions across the small test DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5. The WRF model was archived at the hourly scale so it provided more insights into the subdaily W values. We compared the basin mean annual maximum W depth for the 1-, 3-, 6-, 12-, 24-, 48-, and 72-hour durations for the historical and the future climate conditions averaged among the five GCMs along with the uncertainty range for the future climate conditions. For all the durations, all but the GFDL_ESM2M GCM projected an increase in annual maximum W depth. The mechanism behind the slightly decreased projection of W in the GFDL_ESM2M model deserves further investigation.

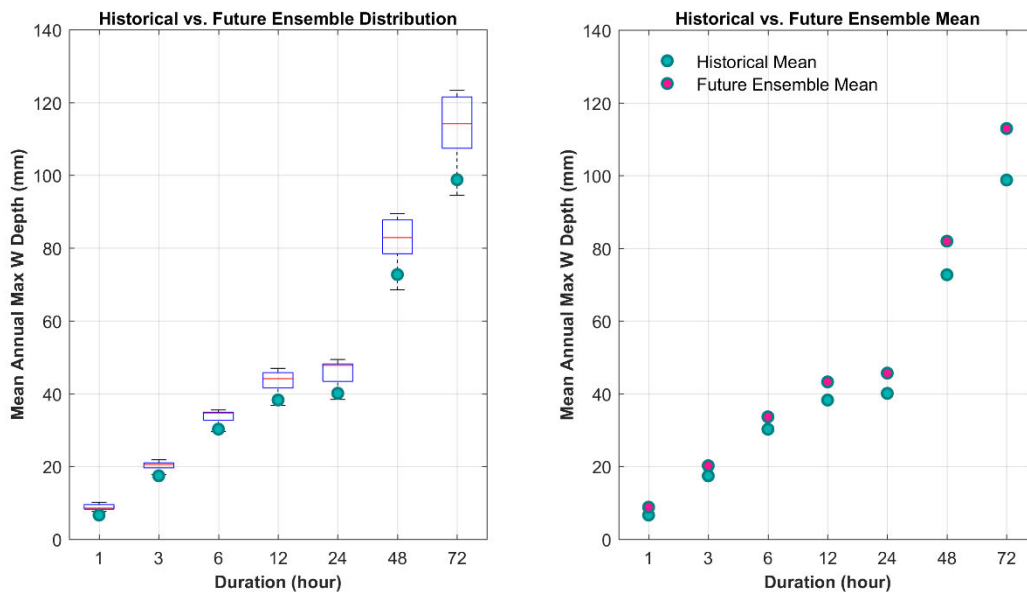


Figure 36. Basin-scale annual maximum W depth for historical (1981–2010) and future (2041–2070) climate conditions over the small test DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5.

Table 3. The basin mean annual maximum *W* depth (mm) for the 1-, 3-, 6-, 12-, 24-, 48-, and 72-hour durations.

Climate	Duration						
	1 Hour	3 Hours	6 Hours	12 Hours	24 Hours	48 Hours	72 Hours
Historic	6.6	17.4	30.3	38.8	40.1	72.7	98.8
Future	8.8	20.1	33.6	43.3	45.7	82.0	112.9
Uncertainty	7.6-11.1	17.8-21.9	29.6-35.5	36.8-47.0	38.4-49.4	68.6-89.5	94.5-123.3

After extracting the annual maximum *W* data set, we developed the NG-IDF curves for both historical and future climate conditions using the GEV distribution. Figure 37 presents the historical and future ensemble mean NG-IDF curves for durations from 1 hour to 72 hours and return periods from 5 years to 100 years. It is observed that the *W* extremes increased significantly for the future period for the MCMWTC DoD site. The *W* extremes also showed a higher rate of increase at the shorter durations (e.g., 1 hour) than the longer duration (e.g., 24 hours), suggesting that not only the magnitude of *W* is changing but the temporal distribution of *W* also is changing. The temporal distribution of *W* tended to become sharper under a warming climate. For example, the 1-hour, 100-year event increased from 9.7 mm under historical climate conditions to 18.2 mm (ensemble mean) under future climate conditions or an 87% increase under the RCP8.5 scenario, while the 24-hour, 100-year event increased from 82.2 mm to 91.5 mm, or an 11.4% increase. These results are consistent with the findings of Wasko and Sharma (2015), who, based on used observation data, found a sharper hyetograph was associated with higher temperatures.

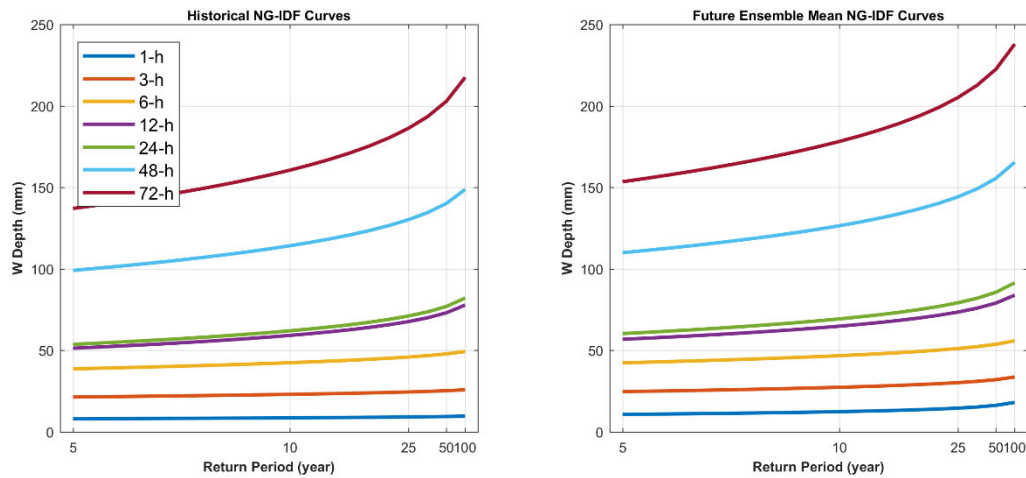


Figure 37. NG-IDF curves for historical (1981–2010) and future (2041–2070) climate conditions over the DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5.

Besides presenting the ensemble mean NG-IDF curves, Figure 38 and Table 4 also show the variability/uncertainty of future changes in the NG-IDF curve values, using two return periods—10 years and 50 years—as examples. These results are similar to those depicted in Figure 36;

four GCMs showed an increase in the extreme W events and only the GFDL_ESM2M GCMs showed a moderate increase or decrease in extreme W events.

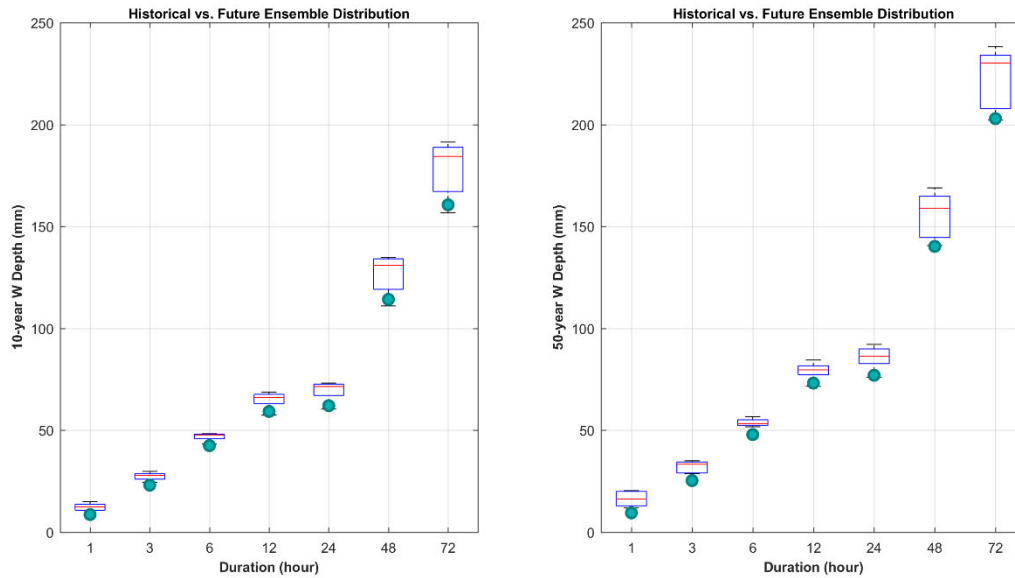


Figure 38. Two examples of extreme W events for historical (1981–2010; blue circle) and future (2041–2070, box and whisker) climate conditions over the small test DoD site basin using the data for the five selected WRF dynamically downscaled GCMs under RCP8.5.

Table 4. Variability/uncertainty of future changes in the NG-IDF curve values (mm) for 10- and 50-year return periods and durations of 1, 3, 6, 12, 24, 48, and 72 hours.

Return Period	Duration						
	1 Hour	3 Hours	6 Hours	12 Hours	24 Hours	48 Hours	72 Hours
10-Year	10.7-15	24.5-29.9	43.3-48.4	57.4-68.7	60.5-73.2	111.1-134.7	156.8-191.4
100-Year	11.9-20.4	28.8-35.1	51.8-56.6	71.6-84.5	76.1-92.1	140.7-168.9	202.3-238.3

In summary, these results suggest a significant increase in extreme W events under future (2041–2070) climate conditions at the DoD MCMWTC site. They also have significant implications for future hydrologic design because not only is the W magnitude projected to increase but the W hyetograph also is projected to become sharper (i.e., a twofold effect), leading to a much higher failure risk of infrastructure in the future. For example, Hettiarachchi et al. (2018) found that the changes in the projected precipitation hyetograph alone could increase the risk of flood magnitude up to 35% in an urban basin in Minnesota.

5. Conclusions and Implications for Future Research/Implementation

In this project, our research team of climate scientists, hydrologic scientists, water resource engineers, and data scientists collaborated to understand fundamental atmospheric and hydrologic processes and have developed scientifically defensible methods—the NG-IDF curves—for infrastructure design in the face of persistent and increasing flood threats. Given the

lack of coordinated and consistent surface water design manuals for snow-dominated regions of the United States, by developing the NG-IDF curves we aim to provide a consistent IDF-based design approach for both rain- and snow-dominated regions. For the first time, NG-IDF curves have been developed at more than 200,000 locations at 1/16° grid cell (~6 km) resolution across the CONUS, considering nonstationary and snow processes under both historical and future climate conditions (10 GCM ensemble products). The NG-IDF tool has the potential to benefit both practicing engineers, planners, and DoD and jurisdictions responsible for developing unified facility criteria and approving subsequent infrastructure designs. Consistent with the study objectives, the main conclusions of this project are listed in Table 5:

Table 5. Research objectives and summary of results.

Objective	Description	Results
1	NG-IDF Proof-of-Concept using SNOTEL Observation	We enhanced the technical basis of a 50-year old PREC-IDF approach predicated on all precipitation being in the form of rain that is immediately available for runoff. We developed a new approach based on the concept of water available for runoff (W)—the NG-IDF curve—which accounts for both rain and snow processes. We analyzed historical SNOTEL station observations in the western United States and found that the traditional PREC-IDF approach underestimated W intensity at 43% of nearly 400 SNOTEL stations across the western United States. When coupling the NG-IDF curves with the TR-55 rainfall-runoff model, we found that about 70% of the stations in the western United States showed the potential for underdesign, for which the PREC-IDF curves underestimated peak design floods by as much as 324%.
2	DHSVM Snow Modeling and Validation	We developed new understanding of historical and future trends of rain, snow, and ROS that affect NG-IDF curves. For example, using 40-year historical observations, we found that extreme W events in the southwestern United States declined by 15%, while the frequency of ROS events in the northwestern United States increased by 32%. We had made advancements in accounting for snow processes in local forest and regional-scale DHSVM hydrologic modeling, effectively enabling continental-scale development of NG-IDF curves. Specific advancements include new representation of forest-snow interactions by explicitly accounting for forest canopy and gaps that affect snow retention, and the development of new regional parameters that account for snow processes in snow models. We validated the enhanced the DHSVM canopy snow model and regional snow parameters at SNOTEL stations across the western United States. The well-validated DHSVM provides the opportunity to develop NG-IDF curves for complex forested environments and ungauged basins across the CONUS.

Objective	Description	Results
3	Validation of NG-IDF Tool in Design Flood Estimates	We validated the NG-IDF tool in design floods using the DHSVM continuous simulation at both hillslope and basin scales. This validation confirmed the accuracy and robustness of the NG-IDF tool in practical applications. In addition, for the first time, we developed a hyetograph of W to improve NG-IDF tool performance. We found that the use of standard rainfall hyetographs led to significant underestimates of design floods and the use of a hyetograph of W played an important role in NG-IDF design. We also highlighted the potential change of the W and rainfall hyetographs under a warming climate, which could further increase future flood risk.
4	NG-IDF Products under Historical Climate	We found that snowmelt and ROS were dominant drivers of extreme events of 24-hour to 72-hour durations in the mountainous western United States and the Northeast Appalachians. The dominant mechanism shifted for different return periods and durations. In the Northwest, the 10-year extreme W events were dominated by rain and snowmelt (~70% locations); and snowmelt and ROS in spring dominated the 100-year events (~80% locations). In the Northeast, the 10-year events were dominated by snowmelt and rain (~80% locations); and ROS and rain-dominated 100-year events (~80% locations). Snowmelt dominated the extreme events for 24-hour to 72-hour durations in the snowbelts near the Great Lakes that were strongly affected by lake-effect snowstorms. In general, the NG-IDF W extremes were higher in magnitude in the rain-dominated Southeast, and in the snowmelt- or ROS-dominated mountainous regions in the Northeast and western United States.
5	Modeling Extreme Precipitation in Future Climate	We provided new understanding of the impacts of ARs on extreme precipitation and watershed budgets including runoff and snow. In particular, we identified two physical mechanisms through which ARs produced higher flooding events than non-AR events. These findings can shed light on the physical understanding of changes in NG-IDF curves under future climate conditions in the western United States.
6	Ensemble NG-IDF Products under Future Climate	We compared the ensemble mean ratio of the NG-IDF curve values to PREC-IDF curve values (in terms of W/P) for historical (1950–2005) and future (2044–2099) climate conditions across the CONUS using statistically downscaled data from 10 GCMs under RCP8.5. For the 24-hour, 100-year event, the locations subject to underdesign decreased from 14% to 9% of the CONUS from the historical period to future period. For the 72-hour, 100-year event, the locations subject to underdesign decreased from

Objective	Description	Results
		23% to 9% of the CONUS from the historical period to future period. However, it is important to note that it does not mean the W or P value will decrease during the future period; under a generally warm future climate, the extreme P and W design events both increased in magnitude. These results only indicate that there is a lower risk of infrastructure underdesign when comparing W to P extremes. This is mainly due to the decreasing snowfall and snowpack or earlier and slower snowmelt under a warming climate. These results suggest that under future climate conditions, extreme W events at more locations shift from melt-dominated or ROS-dominated to rain-dominated and from 24-hour to 72-hour durations.

These accomplishments support the potential of NG-IDF curves to be used as an alternative to the standard PREC-IDF curves for hydrologic design in both rain- and snow-dominated regions (i.e., providing a consistent IDF design method). Nevertheless, further investigations are necessary to enhance the NG-IDF tool for use in design practice. Here we outline the five directions for future development:

1. Develop regional hyetographs for W. We proved the critical role of hyetograph shape in NG-IDF modeling in the basin-scale validation. Regional hyetographs are necessary for hydrologic design in ungauged basins. DHSVM- (or another suitable model) simulated W time series can be used to construct regional hyetographs across CONUS and test them against different hillslope and basin configurations to evaluate the framework across different hydroclimatic regions including more DoD sites. For NG-IDF curves under future climate conditions, it is also important to quantify the projected changes in W hyetograph.
2. Provide NG-IDF products for multiple land covers, such as the TR-55 land-cover types, to be consistent with the standard hydrologic design workflow. Vegetation plays a significant role in snow process through canopy interception and sublimation, and changes the snowpack energy balance. The NG-IDF curves produced in this paper were based on bare ground approximating the SNOTEL open condition. Infrastructure design at the basin scale may contain multiple types of land cover and thus we need to estimate the weighted combination of W for different land covers. DHSVM can be used to produce W data sets for different vegetation covers, perhaps following the TR-55 classifications.
3. Examine the areal reduction factor (ARF) for W. IDF-based design in larger basins commonly uses a local or regional ARF to convert point IDF depths to basin depths by accounting for the spatial variability of storm intensity in a basin. In our hillslope validation studies, it is reasonable to assume the point W rate falls uniformly across the hillslope; however, it is unrealistic to assume this for larger basins. DHSVM-simulated W can be used to estimate the ARFs for different basins (e.g., HUC8 or HUC10) by calculating the ratio between the average W for the basin and the areal maximum W for different storm durations across CONUS. For hydrologic design under future climate conditions, it is also necessary to quantify the projected changes in ARF.

4. Generate ensemble dynamically downscaled data over CONUS. It is desirable to use WRF dynamically downscaled climate data for the historical and future periods from the GCM ensemble to represent the variability across climate projections. Because of the large computational demand, we had five GCMs for future climate conditions across the western United States at 6 km spatial resolution in this project. The WRF model can be used to downscale more GCMs to better quantify the GCM uncertainty in future W events and better understand the changes in ARs and other physical mechanisms responsible for extreme events. To provide CONUS NG-IDF products, it would also be necessary to extend the WRF model domain from the western United States to CONUS.
5. Develop probabilistic NG-IDF curves that have transparent uncertainty quantification and a cyberinfrastructure for data distribution. Figure 39 illustrates a cyberinfrastructure for distributing probabilistic NG-IDF curves. For the historical NG-IDF curves, it is desirable to produce the uncertainty sources, including observed meteorological data uncertainty, hydrologic model parameter uncertainty, and sample data uncertainty in frequency analysis. For the future NG-IDF curves, two additional uncertainty sources are scenario uncertainty and GCM climate model uncertainty. The final probabilistic NG-IDF design tool would provide a probability design function for each selected storm duration and return period (Figure 39). The probability design function can include all sources of uncertainty and will provide information about the most likely condition and the associated 95% confidence interval (the least condition, the worst condition) to support engineering design risk analysis that is adjustable based on design cost and changing operating conditions.

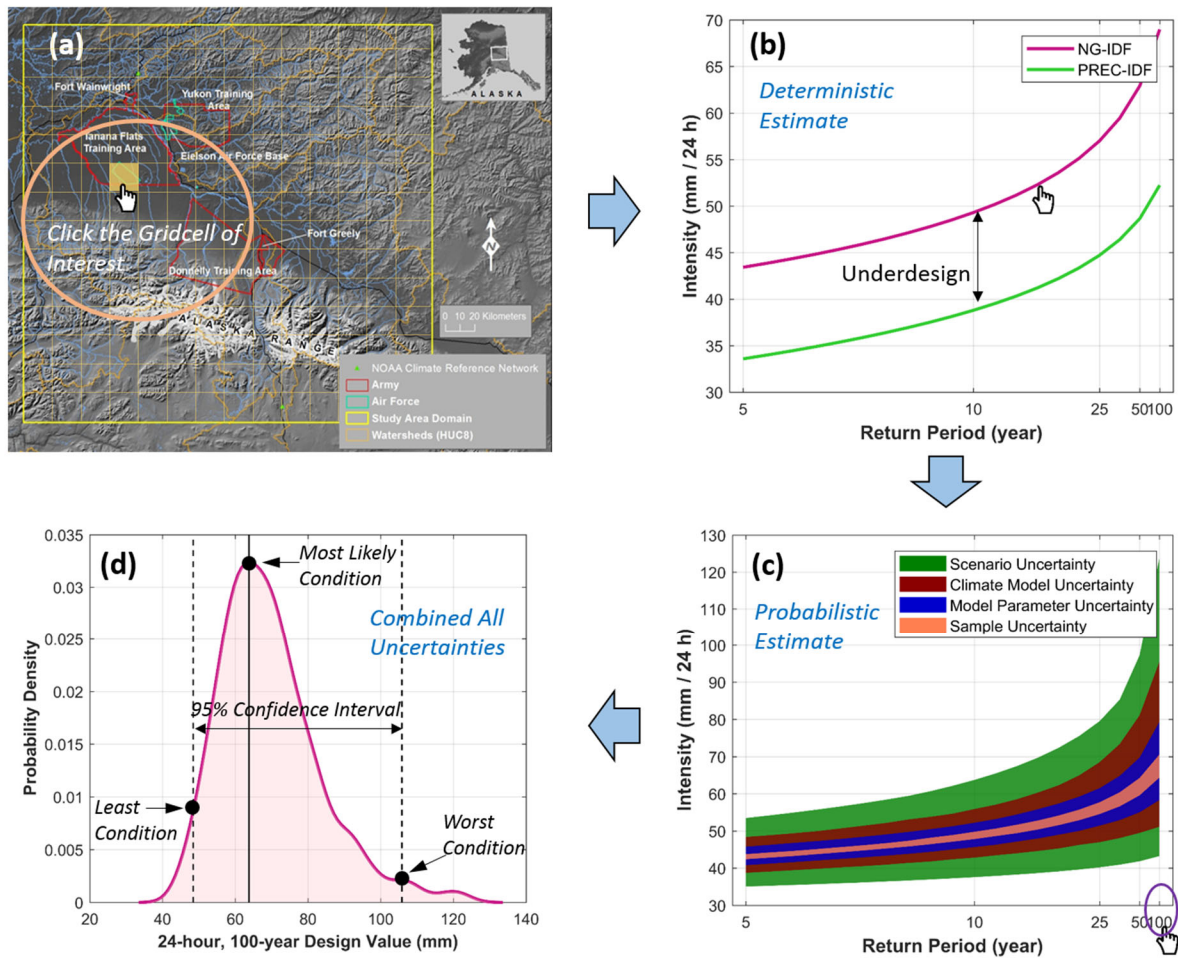


Figure 39. An example of a cyberinfrastructure for distributing probabilistic NG-IDF curves for a DoD site in Alaska with explicit and complete quantification of uncertainty for end users. The uncertainty source includes the scenario divergence, climate model uncertainty, hydrologic model parameter uncertainty, and sample data uncertainty. The dominant uncertainty source is identified and end users are able to select which types of source of uncertainty they would like to include in their design risk analysis.

6. Literature Cited

- Abatzoglou, J.T., Brown, T.J., 2012. A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.* 32, 772–780. <https://doi.org/10.1002/joc.2312>
- Bales, R.C., Molotch, N.P., Painter, T.H., Dettinger, M.D., Rice, R., Dozier, J., 2006. Mountain hydrology of the western United States. *Water Resour. Res.* 42, W08432. <https://doi.org/10.1029/2005WR004387>
- Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438, 303–309. <https://doi.org/10.1038/nature04141>
- Beckers, J., Smerdon, B., Wilson, M., 2009. Review of hydrologic models for forest

- management and climate change applications in British Columbia and Alberta, FORREX Series 25. Forum for Research and Extension in Natural Resources Society, Kamloops, British Columbia, Canada.
- Bonnin, G.M., Martin, D., Lin, B., Parzybok, T., Yekta, M., Riley, D., 2011. Precipitation-Frequency Atlas of the United States, NOAA Atlas 14. vol. 1, version 5.0, U.S. Department of Commerce, National Oceanic and Atmospheric Administration.
- Boughton, W., Droop, O., 2003. Continuous simulation for design flood estimation—a review. *Environ. Model. Softw.* 18, 309–318. [https://doi.org/10.1016/S1364-8152\(03\)00004-5](https://doi.org/10.1016/S1364-8152(03)00004-5)
- Calver, A., Stewart, E., Goodsell, G., 2009. Comparative analysis of statistical and catchment modelling approaches to river flood frequency estimation. *J. Flood Risk Manag.* 2, 24–31. <https://doi.org/10.1111/j.1753-318X.2009.01018.x>
- Carnevale, A.P., Smith, N., 2017. Trillion Dollar Infrastructure Proposals Could Create Millions of Jobs. Center on Education and the Workforce, Georgetown University.
- CEC, 2009. Commission for Environmental Cooperation Ecological Regions of North America, Level III.
- Chen, X., Duan, Z., Leung, L.R., Wigmosta, M., 2019a. A Framework to Delineate Precipitation-Runoff Regimes: Precipitation Versus Snowpack in the Western United States. *Geophys. Res. Lett.* 46, 13044–13053. <https://doi.org/10.1029/2019GL085184>
- Chen, X., Leung, L.R., Gao, Y., Liu, Y., Wigmosta, M., Richmond, M., 2018. Predictability of Extreme Precipitation in Western U.S. Watersheds Based on Atmospheric River Occurrence, Intensity, and Duration. *Geophys. Res. Lett.* 45, 11,693–11,701. <https://doi.org/10.1029/2018GL079831>
- Chen, X., Leung, L.R., Wigmosta, M., Richmond, M., 2019b. Impact of Atmospheric Rivers on Surface Hydrological Processes in Western U.S. Watersheds. *J. Geophys. Res. Atmos.* 124, 8896–8916. <https://doi.org/10.1029/2019JD030468>
- Chow, V.T., 1964. Handbook of Applied Hydrology: A Compendium of Water-Resources Technology. McGraw-Hill, New York.
- Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied Hydrology. McGraw-Hill, New York.
- Clow, D.W., 2010. Changes in the Timing of Snowmelt and Streamflow in Colorado: A Response to Recent Warming. *J. Clim.* 23, 2293–2306. <https://doi.org/10.1175/2009JCLI2951.1>
- Cronshey, R., McCuen, R.H., Miller, N., Rawls, W., Robbins, S., Woodward, D., 1986. Urban Hydrology for Small Watersheds—TR-55. U.S. Department of Agriculture, Washington, D.C.
- ESTCP, 2018. Nonstationary Weather Patterns and Extreme Events: Informing Design and Planning for Long-Lived Infrastructure. Project RC-201591. Retrieved from <https://www.serdp-estcp.org/News-and-Events/Blog/Nonstationary-Weather-Patterns-and-Extreme-Events-Workshop-Report>.
- Fassnacht, S.R., Records, R.M., 2015. Large snowmelt versus rainfall events in the mountains. *J. Geophys. Res. Atmos.* 120, 2375–2381. <https://doi.org/10.1002/2014JD022753>
- Hamlet, A.F., 2018. New Observed Data Sets for the Validation of Hydrology and Land Surface Models in Cold Climates. *Water Resour. Res.* 54, 5190–5197. <https://doi.org/10.1029/2018WR023123>
- Hawkins, E., Sutton, R., 2009. The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bull. Am. Meteorol. Soc.* 90, 1095–1108. <https://doi.org/10.1175/2009BAMS2607.1>

- Helsel, D.R., Frans, L.M., 2006. Regional Kendall Test for Trend. *Environ. Sci. Technol.* 40, 4066–4073. <https://doi.org/10.1021/es051650b>
- Hettiarachchi, S., Wasko, C., Sharma, A., 2018. Increase in flood risk resulting from climate change in a developed urban watershed – the role of storm temporal patterns. *Hydrol. Earth Syst. Sci.* 22, 2041–2056. <https://doi.org/10.5194/hess-22-2041-2018>
- Hornberger, G.M., Spear, R.C., 1981. Approach to the preliminary analysis of environmental systems. *J. Environ. Manage.* 12, 7–18.
- Hosking, J.R.M., Wallis, J.R., 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, U. K.
- Hou, Z., Ren, H., Sun, N., Wigmosta, M.S., Liu, Y., Leung, L.R., Yan, H., Skaggs, R., Coleman, A., 2019. Incorporating Climate Nonstationarity and Snowmelt Processes in Intensity–Duration–Frequency Analyses with Case Studies in Mountainous Areas. *J. Hydrometeorol.* 20, 2331–2346. <https://doi.org/10.1175/JHM-D-19-0055.1>
- Huff, F.A., 1990. *Time Distributions of Heavy Rainstorms in Illinois*. Illinois State Water Survey, Champaign, 173, 17 pp.
- Hungerford, R.D., Nemani, R.R., Running, S.W., Coughlan, J.C., 1989. MTCLIM: A Mountain Microclimate Simulation Model. U.S. Department of Agriculture, Ogden, Utah, Utah.
- Kendall, M.G., 1975. *Rank Correlation Methods*. Griffin, London.
- Laramie, R.L., Schaake, J.C.J., 1972. Simulation of the continuous snowmelt process.
- Livneh, B., Rosenberg, E.A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K.M., Maurer, E.P., Lettenmaier, D.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., Nijssen, B., 2013. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *J. Clim.* 26, 9384–9392. <https://doi.org/10.1175/JCLI-D-12-00508.1>
- Lundquist, J.D., Roche, J.W., Forrester, H., Moore, C., Keenan, E., Perry, G., Cristea, N., Henn, B., Lapo, K., McGurk, B., Cayan, D.R., Dettinger, M.D., 2016. Yosemite Hydroclimate Network: Distributed stream and atmospheric data for the Tuolumne River watershed and surroundings. *Water Resour. Res.* 52, 7478–7489. <https://doi.org/10.1002/2016WR019261>
- Madsen, H., Arnbjerg-Nielsen, K., Mikkelsen, P.S., 2009. Update of regional intensity–duration–frequency curves in Denmark: Tendency towards increased storm intensities. *Atmos. Res.* 92, 343–349. <https://doi.org/https://doi.org/10.1016/j.atmosres.2009.01.013>
- Mann, H.B., 1945. Nonparametric Tests Against Trend. *Econometrica* 13, 245. <https://doi.org/10.2307/1907187>
- McCabe, G.J., Hay, L.E., Clark, M.P., 2007. Rain-on-Snow Events in the Western United States. *Bull. Am. Meteorol. Soc.* 88, 319–328. <https://doi.org/10.1175/BAMS-88-3-319>
- McCuen, R.H., 1998. *Hydrologic analysis and design*, 2nd ed. Prentice-Hall, Upper Saddle River, N.J.
- Mearns, L.O., Sain, S., Leung, L.R., Bukovsky, M.S., McGinnis, S., Biner, S., Caya, D., Arritt, R.W., Gutowski, W., Takle, E., Snyder, M., Jones, R.G., Nunes, A.M.B., Tucker, S., Herzmann, D., McDaniel, L., Sloan, L., 2013. Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Clim. Change* 120, 965–975. <https://doi.org/10.1007/s10584-013-0831-3>
- Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity Is Dead: Whither Water Management? *Science* 319, 573–574. <https://doi.org/10.1126/science.1151915>
- Moriasi, D., Arnold, J., Liew, M. Van, 2007. Model evaluation guidelines for systematic

- quantification of accuracy in watershed simulations. *Am. Soc. Agric. Biol. Eng.* 50, 885–900.
- Mote, P.W., Li, S., Lettenmaier, D.P., Xiao, M., Engel, R., 2018. Dramatic declines in snowpack in the western US. *npj Clim. Atmos. Sci.* 1, 2. <https://doi.org/10.1038/s41612-018-0012-1>
- Musselman, K.N., Clark, M.P., Liu, C., Ikeda, K., Rasmussen, R., 2017. Slower snowmelt in a warmer world. *Nat. Clim. Chang.* 7, 214–219. <https://doi.org/10.1038/nclimate3225>
- Musselman, K.N., Lehner, F., Ikeda, K., Clark, M.P., Prein, A.F., Liu, C., Barlage, M., Rasmussen, R., 2018. Projected increases and shifts in rain-on-snow flood risk over western North America. *Nat. Clim. Chang.* 8, 808–812. <https://doi.org/10.1038/s41558-018-0236-4>
- Naz, B.S., Kao, S.-C., Ashfaq, M., Rastogi, D., Mei, R., Bowling, L.C., 2016. Regional hydrologic response to climate change in the conterminous United States using high-resolution hydroclimate simulations. *Glob. Planet. Change* 143, 100–117. <https://doi.org/10.1016/j.gloplacha.2016.06.003>
- Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett, D., Brekke, L., Arnold, J.R., Hopson, T., Duan, Q., 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci.* 19, 209–223. <https://doi.org/10.5194/hess-19-209-2015>
- Perica, S., Martin, D., Pavlovic, S., Roy, I., Laurent, M. St., Trypaluk, C., Unruh, D., Yekta, M., Bonnin, G., 2013. *Precipitation-Frequency Atlas of the United States*, NOAA Atlas 14. vol. 8, version 2.0, U.S. Dep. of Commer., National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Md.
- Ren, H., Hou, Z.J., Wigmosta, M., Liu, Y., Leung, L.R., 2019. Impacts of Spatial Heterogeneity and Temporal Non-Stationarity on Intensity-Duration-Frequency Estimates—A Case Study in a Mountainous California-Nevada Watershed. *Water* 11, 1296. <https://doi.org/10.3390/w11061296>
- Rogger, M., Kohl, B., Pirkel, H., Viglione, A., Komma, J., Kirnbauer, R., Merz, R., Blöschl, G., 2012. Runoff models and flood frequency statistics for design flood estimation in Austria – Do they tell a consistent story? *J. Hydrol.* 456–457, 30–43. <https://doi.org/10.1016/j.jhydrol.2012.05.068>
- Sen, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall’s Tau. *J. Am. Stat. Assoc.* 63, 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Sturm, M., Goldstein, M.A., Parr, C., 2017. Water and life from snow: A trillion dollar science question. *Water Resour. Res.* 53, 3534–3544. <https://doi.org/10.1002/2017WR020840>
- Sugiyama, M., Shiogama, H., Emori, S., 2010. Precipitation extreme changes exceeding moisture content increases in MIROC and IPCC climate models. *Proc. Natl. Acad. Sci.* 107, 571–575. <https://doi.org/10.1073/pnas.0903186107>
- Sun, N., Wigmosta, M., Zhou, T., Lundquist, J., Dickerson-Lange, S., Cristea, N., 2018. Evaluating the functionality and streamflow impacts of explicitly modelling forest-snow interactions and canopy gaps in a distributed hydrologic model. *Hydrol. Process.* 32, 2128–2140. <https://doi.org/10.1002/hyp.13150>
- Sun, N., Yan, H., Wigmosta, M.S., Leung, L.R., Skaggs, R., Hou, Z., 2019. Regional Snow Parameters Estimation for Large-Domain Hydrological Applications in the Western United States. *J. Geophys. Res. Atmos.* 124, 5296–5313. <https://doi.org/10.1029/2018JD030140>
- Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An Overview of CMIP5 and the Experiment Design. *Bull. Am. Meteorol. Soc.* 93, 485–498. <https://doi.org/10.1175/BAMS-D-11->

00094.1

- Wasko, C., Sharma, A., 2015. Steeper temporal distribution of rain intensity at higher temperatures within Australian storms. *Nat. Geosci.* 8, 527–529.
<https://doi.org/10.1038/ngeo2456>
- Wigmosta, M.S., Lettenmaier, D.P., 1999. A comparison of simplified methods for routing topographically driven subsurface flow. *Water Resour. Res.* 35, 255–264.
<https://doi.org/10.1029/1998WR900017>
- Wigmosta, M.S., Vail, L.W., Lettenmaier, D.P., 1994. A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.* 30, 1665–1679.
<https://doi.org/10.1029/94WR00436>
- Yan, H., Sun, N., Wigmosta, M., Leung, L.R., Hou, Z., Coleman, A., Skaggs, R., 2020. Evaluating next-generation intensity–duration–frequency curves for design flood estimates in the snow-dominated western United States. *Hydrol. Process.* 34, 1255–1268.
<https://doi.org/10.1002/hyp.13673>
- Yan, H., Sun, N., Wigmosta, M., Leung, L.R., Hou, Z., Skaggs, R., Coleman, A., 2019a. Enhancing hydrologic design by next-generation intensity-duration-frequency curves considering snowmelt and climate nonstationarity, in: *World Environmental and Water Resources Congress 2019: Groundwater, Sustainability, Hydro-Climates/Climate Change, and Environmental Engineering*. American Society of Civil Engineers, Pittsburgh, PA, pp. 278–286. <https://doi.org/10.1061/9780784482346.028>
- Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., Leung, L.R., 2019b. Next-Generation Intensity–Duration–Frequency Curves to Reduce Errors in Peak Flood Design. *J. Hydrol. Eng.* 24, 04019020. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001799](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001799)
- Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., Leung, R., 2018. Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design in Snow-Dominated Environments. *Water Resour. Res.* 54, 1093–1108. <https://doi.org/10.1002/2017WR021290>
- Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Leung, L.R., Coleman, A., Hou, Z., 2019c. Observed Spatiotemporal Changes in the Mechanisms of Extreme Water Available for Runoff in the Western United States. *Geophys. Res. Lett.* 46, 767–775.
<https://doi.org/10.1029/2018GL080260>
- Yen, B.C., 1970. Risks in hydrologic design of engineering projects. *J. Hydraul. Div.* 96, 959–965.

Appendix A. Supporting Data

Appendix A.1. BCQC SNOTEL Data Set

The bias corrected and quality controlled (BCQC) snow telemetry (SNOTEL) data consist of bias-corrected and quality-controlled daily snow water equivalent (SWE) and meteorological records including daily minimum, maximum, and mean air temperature and daily total precipitation through 09/30/2018 for 829 active stations located in the western United States and Alaska (Figure A1). For details, the users are referred to Sun et al. (2019) for the bias correction (BC) procedure and Yan et al. (2018) for the quality control (QC) procedure.

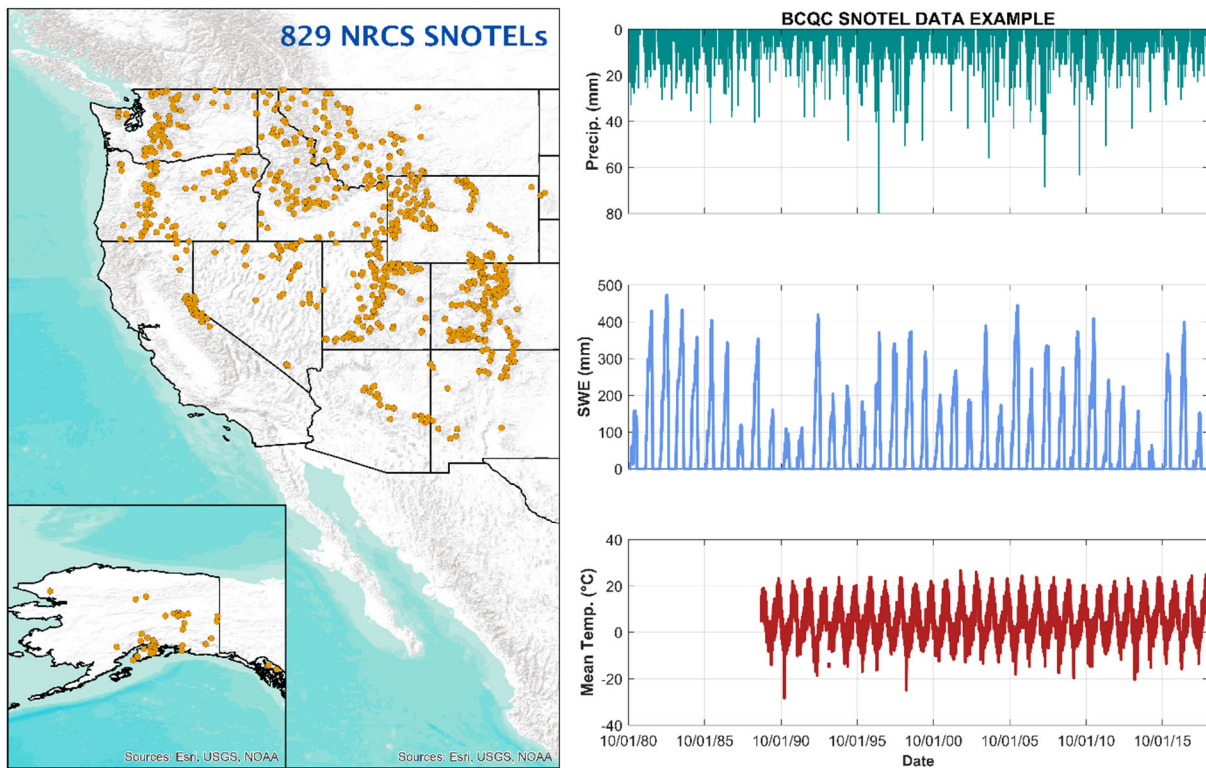


Figure A1. Left: 829 active NRCS SNOTELs across the western United States and Alaska. Right: daily time series of BCQC precipitation, SWE, and air temperature data for one SNOTEL as an example.

The BCQC data are free for downloading at: <https://dhsvm.pnnl.gov>. The file of each SNOTEL station is named as “bcqc_<latitude>_<longitude>.txt” (i.e., bcqc_44.43000_-120.33000.txt). In each text file, there are eight columns separated by a space delimiter:

- 1st col: year
- 2nd col: month
- 3rd col: day

- 4th col: daily precipitation, in inch
- 5th col: maximum air temperature, in °F
- 6th col: minimum air temperature, in °F
- 7th col: mean air temperature, in °F
- 8th col: SWE, in inch, reset to zero on 1 October at the start of each water year.

In each text file, “nan” indicates missing or filtered out data after the BCQC procedures. The raw daily SNOTEL SWE measurements and meteorological records were acquired from USDA National Resource Conservation Center (NRCS). The raw data for 829 SNOTEL stations were compiled and compressed into one zip file and are available for download on our website. We maintain BCQC SNTOTEL data and will make updates periodically to extend the data to the latest available SNOTEL records.

Appendix B. List of Scientific/Technical Publications

Appendix B.1. Articles in Peer-Reviewed Journals

1. Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., Leung, R., 2018. Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design in Snow-Dominated Environments. *Water Resour. Res.* 54, 1093–1108. <https://doi.org/10.1002/2017WR021290>
2. Sun, N., Wigmosta, M., Zhou, T., Lundquist, J., Dickerson-Lange, S., Cristea, N., 2018. Evaluating the functionality and streamflow impacts of explicitly modelling forest-snow interactions and canopy gaps in a distributed hydrologic model. *Hydrol. Process.* 32, 2128–2140. <https://doi.org/10.1002/hyp.13150>
3. Chen, X., Leung, L.R., Gao, Y., Liu, Y., Wigmosta, M., Richmond, M., 2018. Predictability of Extreme Precipitation in Western U.S. Watersheds Based on Atmospheric River Occurrence, Intensity, and Duration. *Geophys. Res. Lett.* 45, 11,693–11,701. <https://doi.org/10.1029/2018GL079831>
4. Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., Leung, L.R., 2019. Next-Generation Intensity–Duration–Frequency Curves to Reduce Errors in Peak Flood Design. *J. Hydrol. Eng.* 24, 04019020. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001799](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001799)
5. Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Leung, L.R., Coleman, A., Hou, Z., 2019. Observed Spatiotemporal Changes in the Mechanisms of Extreme Water Available for Runoff in the Western United States. *Geophys. Res. Lett.* 46, 767–775. <https://doi.org/10.1029/2018GL080260>
6. Sun, N., Yan, H., Wigmosta, M.S., Leung, L.R., Skaggs, R., Hou, Z., 2019. Regional Snow Parameters Estimation for Large-Domain Hydrological Applications in the Western United States. *J. Geophys. Res. Atmos.* 124, 5296–5313. <https://doi.org/10.1029/2018JD030140>
7. Chen, X., Leung, L.R., Wigmosta, M., Richmond, M., 2019. Impact of Atmospheric Rivers on Surface Hydrological Processes in Western U.S. Watersheds. *J. Geophys. Res. Atmos.* 2019JD030468. <https://doi.org/10.1029/2019JD030468>
8. Ren, H., Hou, Z.J., Wigmosta, M., Liu, Y., Leung, L.R., 2019. Impacts of Spatial Heterogeneity and Temporal Non-Stationarity on Intensity-Duration-Frequency Estimates—A Case Study in a Mountainous California-Nevada Watershed. *Water* 11, 1296. <https://doi.org/10.3390/w11061296>
9. Hou, Z., Ren, H., Sun, N., Wigmosta, M., Liu, Y., Yan, H., Skaggs, R., and Coleman, A., 2019. Incorporating Climate Non-stationarity and Snowmelt Processes in Intensity-Duration-Frequency Analyses with Case Studies in Mountainous Areas. *J. Hydrometeorol.*, 20, 2331–2346. <https://doi.org/10.1175/JHM-D-19-0055.1>
10. Chen, X., Z. Duan, L. R. Leung, and M. Wigmosta (2019), A Framework to Delineate Precipitation-Runoff Regimes: Precipitation Versus Snowpack in the Western United States, *Geophys. Res. Lett.* 46, 13044–13053. <https://doi.org/10.1029/2019GL085184>
11. Yan, H., Sun, N., Wigmosta, M., Leung, L.R., Hou, Z., Coleman, A., and Skaggs, R. (2020), Evaluating next-generation intensity-duration-frequency curves for design flood

estimates in the snow-dominated western United States. *Hydrol. Process.* 34, 1255–1268.
<https://doi.org/10.1002/hyp.13673>

Articles under Revision or in Preparation

1. Sun, N., Yan, H., Wigmosta, M., and Zhou, T. Variability of Canopy Effect on Snow Accumulation and Duration across the Western United States Mountain Snow Regimes: Comparison between Open, Forest, and Canopy Gap. In preparation.
2. Yan, H., Sun, N., Wigmosta, M., and Leung, L.R. Assessment of Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design at Snow-Dominated River Basin Scale: The Need to Develop Hyetograph of Water Available for Runoff. In preparation.
3. Sun, N., Yan, H., Wigmosta, M., and Leung, L.R. Characterizing extreme events relevant to hydrologic design using next-generation intensity-duration-frequency curves over the conterminous United States. In preparation.

Appendix B.2. Technical Reports

None

Appendix B.3. Conference Symposium Proceedings

1. Yan, H., Sun, N., Wigmosta, M., Leung, L.R., Hou, Z., Skaggs, R., Coleman, A., 2019. Enhancing hydrologic design by next-generation intensity-duration-frequency curves considering snowmelt and climate nonstationarity, in: *World Environmental and Water Resources Congress 2019: Groundwater, Sustainability, Hydro-Climates/Climate Change, and Environmental Engineering*. American Society of Civil Engineers, Pittsburgh, PA, pp. 278–286. <https://doi.org/10.1061/9780784482346.028>

Appendix B.4. Conference or Symposium Abstracts

1. Sun, N., Wigmosta, M., Zhou, T., Dickerson-Lange, S., Lundquist, J., and Cristea, N., 2016. Enhancement of DHSVM for spatially distributed modeling of forest-gap-snow dynamics. Abstract (C51D-0688) presented at 2016 AGU Fall Meeting, San Francisco, CA, 12–16 Dec.
2. Lundquist, J., Ris Lambers, J., Dickerson-Lange, S., Sun, N., Cristea, N., and Wigmosta, M., 2016. Do forest-snow feedbacks mitigate or exacerbate a mountain's hydrologic response to a warming world? Abstract (H11D-01) invited presentation at 2016 AGU Fall Meeting, San Francisco, CA, 12–16 Dec.
3. Yan, H., Sun, N., Wigmosta, M., and Hou, Z., 2017. Are precipitation-based intensity-duration-frequency curves appropriate for cost effective and resilient infrastructure design in snow-dominated regions? Next-generation curves with inclusion of rain-on-snow events. Presented at PNNL Post-Graduate Research Symposium, Richland, WA, Jul. 25–26.
4. Wigmosta, M., Yan, H., Sun, N., Skaggs, R., Hou, Z., and Leung, R., 2017. Are precipitation-based intensity-duration-frequency curves appropriate for cost-effective and

- resilient infrastructure design in snow-dominated regions? Next-generation curves with inclusion of rain-on-snow events. Presented at SERDP and ESTCP Symposium, Washington, D.C., Nov. 28–30.
5. Yan, H., Sun, N., Wigmosta, M., and Hou, Z., 2017. Are precipitation-based intensity-duration-frequency curves appropriate for cost effective and resilient infrastructure design in snow-dominated regions? Next-generation curves with inclusion of rain-on-snow events. Abstract (H23L-06) invited presentation at 2017 AGU Fall Meeting, New Orleans, LA, 11–15 Dec.
 6. Sun, N., Yan, H., and Wigmosta, M., 2017. Ensemble-Based Parameter Evaluation for Distributed Snow Modeling in the Western United States. Abstract (H43C-1657) presented at 2017 AGU Fall Meeting, New Orleans, LA, 11–15 Dec.
 7. Hou, Z., Ren, H., Sun, N., Leung, R., Liu, Y., Coleman, A., Skaggs, R., Wigmosta, M., 2017. Impacts of Non-Stationarity in Climate on Flood Intensity-Duration-Frequency: Case Studies in Mountainous Areas with Snowmelt. Abstract (GC43E-1103) presented at 2017 AGU Fall Meeting, New Orleans, LA, 11–15 Dec.
 8. Wigmosta, M., Yan, H., Sun, N., Skaggs, R., Hou, Z., and Leung, R., 2018. Assessing extreme events with next-generation intensity-duration-frequency curves over the Conterminous United States. Presented at SERDP and ESTCP Symposium, Washington, D.C., Nov. 27–29.
 9. Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., Leung, R., and Coleman, A., 2018. Next-Generation IDF Curves for Hydrologic Design. Abstract (H23G-1962) presented at 2018 AGU Fall Meeting, Washington, D.C., 10–14 Dec.
 10. Sun, N., Yan, H., Wigmosta, M., Leung, R., Skaggs, R., Coleman, A., and Hou, Z., 2018. Characterizing Extreme Events Relevant to Hydrologic Design using Next-Generation Intensity-Duration-Frequency Curves over the Conterminous United States. Abstract (H22A-02) presented at 2018 AGU Fall Meeting, Washington, D.C., 10–14 Dec.
 11. Wigmosta, M., Yan, H., Sun, N., Coleman, A., Leung, R., Skaggs, R., and Hou, Z., 2018. Spatiotemporal Changes in the Mechanisms of Extreme Runoff Generation in the Western United States. Abstract (H23G-1974) presented at 2018 AGU Fall Meeting, Washington, D.C., 10–14 Dec.
 12. Chen, X., Leung, R., Gao, Y., Liu, Y., Wigmosta, M., and Richmond, M., 2018. Impact of Atmospheric Rivers on the Seasonal Surface Water Balance and Water Resources of Western U.S. Watersheds. Abstract (H11V-1763) presented at 2018 AGU Fall Meeting, Washington, D.C., 10–14 Dec.
 13. Cristea, N., Reynolds, D., Hinkelman, L., Sun, N., Wigmosta, M., and Lundquist, J., 2018. Importance of Model Structure and Irradiance Input to Modeling Snow Surface Temperature and Snowmelt in Complex Terrain. Abstract (C42B-06) presented at 2018 AGU Fall Meeting, Washington, D.C., 10–14 Dec.
 14. Wigmosta, M., Yan, H., Sun, N., Leung, R., Hou, Z., Skaggs, R., and Coleman A., 2019. Enhancing Hydrologic Design by Next-Generation Intensity-Duration-Frequency Curves Considering Snowmelt and Climate Nonstationarity. Abstract (555773) presented at 2019 ASCE EWRI World Environmental & Water Resources Congress, Pittsburgh, PA, 19–23 May.
 15. Wigmosta, M., 2019. Next-generation intensity-duration-frequency curves for improved hydrologic design. Webinar (#92) presented at SERDP and ESTCP Webinar Series, 20 Jun.

16. Wigmosta, M., 2019. Snowmelt-Based IDF. Invited presentation presented at SERDP and ESTCP Symposium, Washington, D.C., Dec. 3–5.
17. Yan, H., Sun, N., Wigmosta, M., and Leung, R., 2019. Evaluating Next-Generation Intensity-Duration-Frequency Curves for Design Flood Estimates at Snow-Dominated River Basin Scale: The Need to Develop Snowmelt Hyetograph. Abstract (H53H-1844) presented at 2019 AGU Fall Meeting, San Francisco, CA, 9–13 Dec.
18. Sun, N., Yan, H., Wigmosta, M., Zhou, T., Lundquist, J., and Dickerson-Lange, S., 2019. Variability of canopy effect on snow accumulation and duration across the western United States mountain snow regimes: Comparison between open, forest, and canopy gap. Abstract (H11P-1753) presented at 2019 AGU Fall Meeting, San Francisco, CA, 9–13 Dec.
19. Wigmosta, M., Yan, H., Sun, N., Leung, R., Hou, Z., and Coleman, A., 2019. Next-generation intensity-duration-frequency curves for hydrologic design: Concepts, advances, and future prospects. Invited presentation (abstract H52A-02) presented at 2019 AGU Fall Meeting, San Francisco, CA, 9–13 Dec.

Appendix B.5. Test Book or Book Chapters

None