Evaluating next-generation intensity-duration-frequency curves for design flood estimates in the snow-dominated western United States

Running title: Next-Generation Intensity-Duration-Frequency Curves

Hongxiang Yan[1]; Ning Sun[1]; Mark Wigmosta[1,2]; L. Ruby Leung[3]; Zhangshuan Hou[1]; Andre Coleman[1]; and Richard Skaggs[1]

[1] Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, Washington, United States

[2] Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington, United States

[3] Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, United States

Correspondence

Hongxiang Yan, Pacific Northwest National Laboratory, P.O. Box 999, 902 Battelle Blvd, Richland, WA 99354, United States.

Email: hongxiang.yan@pnnl.gov

Funding information

Strategic Environmental Research and Development Program, Contract No. RC-2546

Abstract

Civil infrastructure such as culverts and bridges are commonly designed using precipitation-based intensity-duration-frequency (PREC-IDF) curves, which assume that the occurrence of precipitation is in the form of rainfall and immediately available for the rainfall-runoff process. In snow-dominated regions, where most winter precipitation occurs as snow that melts during spring to early summer, the use of standard PREC-IDF curves may lead to substantial underestimation of design floods and high failure risk of infrastructure. In this context, we developed next-generation IDF (NG-IDF) curves that characterize the actual water reaching the land surface (i.e., rainfall plus snowmelt) to enhance standard infrastructure design in snow-dominated regions. This study evaluates the performance of NG-IDF curves coupled with U.S. Department of Agriculture Technical Release 55 hydrologic model in estimating design floods for 246 snowy locations in different hydroclimate regimes of the western United States. Design flood estimates from a well-validated continuous simulation using a physics-based hydrologic model, the Distributed Hydrology Soil Vegetation Model (DHSVM), were used as the performance benchmark. Compared with the benchmark estimates, the standard PREC-IDF curves led to substantial errors in design flood estimates while the NG-IDF curves significantly reduced these errors. For example, the averaged error in the 50-year design flood estimates over the 246 locations was reduced from 31% with the use of PREC-IDF curves to 12% with the use of NG-IDF curves. Despite the different model structures, the single-event NG-IDF approach versus the continuous simulation DHSVM did not exhibit statistically significant differences in 91% of the 246 locations for the 50-year design flood estimates. This indicates a satisfactory performance of NG-IDF curves to estimate design flow under the conditions tested in the snow-dominated western United States. This article

also presents technical suggestions and the limitations of infrastructure design using NG-IDF curves for regulatory agencies and practicing engineers.

KEYWORDS

NG-IDF curves, DHSVM, design flood, TR-55, infrastructure design, SNOTEL, snow-dominated regions, western United States

1 INTRODUCTION

Current infrastructure design standards to withstand extreme precipitation and flooding are largely based on local precipitation-based intensity-duration-frequency (PREC-IDF) curves (Chow, Maidment, & Mays, 1988; Perica et al., 2013). The PREC-IDF approach assumes that precipitation is in the form of rainfall that immediately begins the rainfall-runoff process. In snow-dominated regions where a large percentage of annual precipitation accumulates as snow in winter and melts in spring to early summer, the use of standard PREC-IDF curves is questionable as extreme snowmelt or rain-on-snow (ROS) rates may exceed extreme precipitation (Yan, Sun, Wigmosta, Leung, et al., 2019; Yan et al., 2018). For example, in the western United States, a large amount of winter precipitation is temporarily stored as snowpack in mountainous regions. In most river basins of the western United States, especially in the states of California, Oregon, and Washington, snowpack is the largest component of water storage and more than 50% of the total annual runoff originates as melting montane snowpack (Li, Wrzesien, Durand, Adam, & Lettenmaier, 2017). As a result, a large number of severe flood events in the western United States are associated with snowmelt and ROS events (Hou et al., 2019; Yan et al., 2018; Yan, Sun, Wigmosta, Skaggs, Leung, et al., 2019).

To overcome this deficiency and provide a consistent IDF design approach for both rain- and snow-dominated regions, Yan et al. (2018) proposed next-generation IDF (NG-IDF) curves, which characterize the actual water reaching the land surface or the water available for runoff (W) that results from the combined effects of rainfall, snowmelt, and/or ROS. W is obtained through mass balance as $W = P - \Delta SWE$, where P is precipitation and ΔSWE is the change in snow water equivalent (SWE). By analyzing data from nearly 400 U.S. Department of Agriculture (USDA) National Resource Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) stations over the western United States, they showed that the standard PREC-IDF curves at 45% of the stations were subject to underdesign. At these SNOTEL locations, Yan, Sun, Wigmosta, Skaggs, Hou, et al. (2019) further examined the peak design flood estimates using PREC-IDF and NG-IDF curves with the USDA NRCS Technical Release 55 (TR-55) hydrologic model (Cronshey et al., 1986). Their results demonstrated the need to replace PREC-IDF curves with NG-IDF curves for future infrastructure design, as the PREC-IDF curves can underestimate peak design flood by as much as 324% in the snow-dominated western United States.

Despite the advantages of NG-IDF curves over PREC-IDF curves, application of NG-IDF curves is limited by snow data availability. The above NG-IDF studies were based on SNOTEL observations, which are limited in both spatial coverage and temporal extent. It is unlikely that the location of infrastructure is coincident with the SNOTEL stations. In this regard, Hamlet (2018) commented that a well-validated hydrologic model is key for extending these limited data in space and time to make NG-IDF curves useful in practical applications. In addition, before implementing the NG-IDF curves to a real design problem, evaluation and benchmarking the performances of the curves in design flood estimates are essential, because risks related to the United States infrastructure design involve trillions of dollars in the coming decades (Carnevale & Smith, 2017). Unfortunately, streamflow in snow-dominated regions with complex topography is often very poorly observed due to inherent difficulties of access (Lundquist et al., 2016). The generally spatially sparse streamflow gauge network in snow-dominated environments presents additional challenges in the evaluation of the NG-IDF approach (Curran, Barth, Veilleux, & Ourso, 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 to evaluate the NG-IDF approach. On the other hand, this hydrological modeling approach also brings uncertainties in both snow and streamflow simulations. Two sources of uncertainties associated with W estimates are hydrological model structure (i.e., different process parameterization) and hydrological model parameter estimation (Clark et al., 2015; Sun et al., 2019). In order to provide robust NG-IDF curves for ungauged sites, quantifying contribution of each uncertainty source and reducing uncertainties in W estimates are necessary and deserve further studies. For instance, Sun et al. (2019) recently developed regional snow parameters across the mountain ranges of the western United States, which helped reduce uncertainties in regional to continental snow and hydrological modeling.

Several studies have used validated hydrologic models to evaluate the PREC-IDF approach in design flood estimates (Boughton & Droop, 2003; Boughton, Srikanthan, & Weinmann, 2002; Calver, Stewart, & Goodsell, 2009; Camici, Tarpanelli, Brocca, Melone, & Moramarco, 2011; Grimaldi, Petroselli, & Serinaldi, 2012; Nnadi, Kline, Wray, & Wanielista, 1999; Pathiraja, Westra, & Sharma, 2012; Rogger et al., 2012). For example, Boughton et al. (2002) found the PREC-IDF method overestimated design floods for watersheds with large baseflow. Grimaldi et al. (2012) suggested that the PREC-IDF method tended to underestimate the design hydrograph's peak, volume, and duration in southern Italy. Calver et al. (2009) performed comparisons over 107 watersheds in the United Kingdom and concluded the PREC-IDF method tended to underestimate design floods of return periods between 2 and 50 years. All these studies attributed the identified overestimation or underestimation in design flood estimates to the inherent limitations within the IDF-based design method.

In comparison to the hydrologic model continuous simulation approach, the IDF method over-simplifies the physical hydrologic process and suffers from several key assumptions including: 1) the subjective choices of antecedent moisture condition prior to the storm event and selection of the design storm hyetograph; and 2) the assumption of equal return periods between the design storm and the resulting flood event (Yan, Sun, Wigmosta, Skaggs, Hou, et al., 2019). Although studies discussed above evaluated the errors associated with the PREC-IDF method, they ignored data sample uncertainties in both the PREC-IDF and flood frequency analyses, which may lead to a conclusion of no statistically significant differences in the comparisons (Ganguli & Coulibaly, 2017; Madsen, Arnbjerg-Nielsen, & Mikkelsen, 2009).

Although the use of continuous simulation from a physics-based hydrologic model can relax the assumptions in the IDF-based method to provide more robust design flood estimates, it can be cost-prohibitive especially for small infrastructure design (Yan et al., 2018). Further, local surface water design manuals may require or recommend the use of IDF-based design methods (SCDM, 2016; SMMEW, 2004). In this context, the NG-IDF technology is more likely to be implemented in the near-term by regulators and agencies as it does not require significant changes to the standard IDF-based design workflow. To evaluate the reliability of the NG-IDF method in design flood estimates, we propose a new evaluation framework based on a well-validated physics-based hydrologic model with explicit uncertainty quantification. In particular, the Distributed Hydrology Soil Vegetation Model (DHSVM) (Wigmosta, Vail, & Lettenmaier, 1994) is used to benchmark the performances of NG-IDF curves coupled with the TR-55 single event-based model in design flood estimates. Here we address the following questions: 1) how well does the new NG-IDF method estimate design floods in different hydroclimate regimes across the snow-dominated western United States; and 2) what are the implications and limitations with use of NG-IDF curves for infrastructure design?

This paper is organized as follows: section 2 describes the evaluation framework in detail, which includes the study area and data sources, DHSVM and TR-55 event-based models, DHSVM snow parameter calibration, NG-IDF curves development, design flood estimate, and uncertainty quantification. Section 3 presents the evaluation results of NG-IDF curves in design flood estimates in different hydroclimate regimes of the snow-dominated western United States. Finally, section 4 concludes the paper and provides technical suggestions for use of NG-IDF curves in infrastructure design.

2 METHODOLOGY

In this section, we describe the framework for evaluating the NG-IDF methods in design flood estimates using the DHSVM continuous-simulation method, followed by more detailed descriptions of the framework data and components.

2.1 Evaluation framework

The framework for quantitative evaluation of design flood estimates from the NG-IDF method by DHSVM is summarized in Figure 1. First, at each SNOTEL location, local meteorological forcing data were used to drive DHSVM for continuous streamflow simulations at the experimental hillslope. Second, the annual maximum water available for runoff from the DHSVM simulations ($W_{dhsvm} = P - \Delta SWE_{dhsvm}$) were validated against the annual maximum water available for runoff from the SNOTEL observations ($W_{obs} = P - \Delta SWE_{obs}$). After successful validation, the NG-IDF curves were developed based on the annual maximum W_{dhsvm}. The PREC-IDF curves were also developed based on annual maximum precipitation from the DHSVM meteorological forcing data (i.e., the SNOTEL observation). Next, flood frequency statistics (i.e., 50-year flood) were derived from the DHSVM simulated annual maximum streamflow and considered as the benchmark for later assessment. Third, the DHSVM derived NG-IDF curves and PREC-IDF curves were both used to drive the TR-55 event-based model to estimate the associated design floods. The PREC-IDF curves were derived from the observed SNOTEL precipitation used to drive DHSVM. 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. Note that in this evaluation framework, the NG-IDF curves were developed based on the DHSVM simulated annual maximum W_{dhsvm} rather than the SNOTEL annual maximum W_{obs}. This way, the design flood differences between the two methods should be exclusively due to the implicit assumptions underlying the IDF event-based method. In addition, we used W_{dhsvm} to construct NG-IDF curves rather than W_{obs} because we want to isolate the effects of using the event-based model assumptions and avoid confounding factors resulting from DHSVM model calibrations and SNOTEL observation biases (e.g., precipitation undercatch and snow drifting). Because DHSVM was still driven by the observed precipitation (which were used to construct PREC-IDF curves), the only difference between the design flood estimates from PREC-IDF and NG-IDF curves was with and without including snow process, leading to no artificial bias in the assessment of the NG-IDF method.

[Place Figure 1 here]

2.2 Study area and data sources

In this study, we chose the snow-dominated sites represented by the SNOTEL stations across the western United States as our study locations (Figure 2). We acquired daily SWE and meteorological measurements, including daily minimum and maximum air temperature and precipitation from 805 active SNOTEL stations across the western United States. Data from all

805 stations were screened following the rigorous three-stage SNOTEL quality control filter (Yan et al., 2018) and 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/. After the BCQC process, a total of 246 stations with the longest common period (2007–2013) of reliable records were retained for DHSVM calibration and spatial evaluation of the NG-IDF design method (Figure 2). The selected seven year records are generally representative of the long-term SWE dynamics at each station (Sun et al., 2019). Qualitatively, the selected SNOTEL stations show good spatial coverage over the different hydroclimate regimes of the western United States (i.e., the Cascade Range, Sierra Nevada, Rocky Mountains, and Arizona/New Mexico Mountains). Because wind speed is unavailable for most stations, we acquired for each SNOTEL station the daily wind speed data of the nearest 1/16° grid cell from the meteorological data set of Livneh et al. (2013), in which wind speed data were interpolated from National Centers for Environmental Prediction–National Center for Atmospheric Research reanalysis. Because frequency computations are less reliable with short records (i.e., seven annual maximum W data) (England et al., 2018); after calibrating DHSVM parameters, we further extracted all available records beyond the 7-year common period at each station to enhance frequency analysis. As a result, the final data set consisted of daily time series of SWE, precipitation, minimum and maximum air temperature, and wind speed for 246 SNOTEL stations with total length of records ranging from 7–28 water years over the period 1983–2013 (Figure 2).

[Place Figure 2 here]

2.3 DHSVM continuous simulation

2.3.1 Model structure and calibration

DHSVM (Wigmosta et al., 1994) is a physics-based, spatially distributed, hydrologic model that was initially developed for simulating hydrologic responses in mountainous regions at spatial scales ranging from 10 m to 150 m and subdaily to daily temporal steps. Subsequent development has enhanced the snow submodel of DHSVM to better represent climate-forest-snow interactions (Sun et al., 2018), and extended the model capability for simulating urban hydrology and water quality (Cao, Sun, Yearsley, Nijssen, & Lettenmaier, 2016; Sun et al., 2016), and reservoir operations (Zhao, Gao, Naz, Kao, & Voisin, 2016). At grid scale, DHSVM tracks the water movement among the atmosphere, canopy, snowpack (if present), and soil via mass and energy balance, which is influenced by the site-specific vegetation, soil, and topographic and climate characteristics. Overland and subsurface flow is routed downslope driven by topography across the landscape until reaching the stream channel network, where it is then routed using a simple one-dimensional approach. Details of DHSVM physics and structure can be found in Wigmosta et al. (1994) and Wigmosta, Nijssen, & Storck (2002).

DHSVM simulates the physical processes of snow accumulation and melt with a two-layer canopy submodel and a two-layer ground snowpack submodel. As canopy effects on snow processes are minimal at SNOTEL locations with open area conditions, here we describe the snowpack submodel only. The snowpack submodel is a two-layer model consisting of a thin surface layer and a deep pack layer. Snowpack mass and energy balance are solved at grid scale, considering all key terms of energy fluxes including radiative flux, turbulent flux, convective flux, and advective heat transfer from rainfall. Conductive heat exchange at the snow-ground interface is neglected.

In our previous study (Sun et al., 2019), we calibrated the DHSVM snowpack model to the daily SWE observations at the same 246 SNOTEL stations used here over the period of 2007–2013. The calibrated simulations agreed very well with the daily SWE observations with the Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) values greater than 0.8 at about 91% of the 246 stations and a mean NSE value of 0.9. In this study, we ran DHSVM using the calibrated snow parameter set identified for each SNOTEL station by Sun et al. (2019). In addition to our previous DHSVM calibration on daily SWE, here we further validated the DHSVM performance in reproducing annual maximum water available for runoff as will be described in section 2.5.

2.3.2 DHSVM streamflow simulation at an experimental hillslope

A small experimental hillslope appropriate for mountainous topography to allow direct comparison between the DHSVM continuous simulation and TR-55 single event-based modeling is shown in Figure 3. The planar hillslope had a constant slope of 0.3 and a soil depth of 1.5 m, which was consistent with the setting in Wigmosta & Lettenmaier (1999) and represented by ten 4 m × 4 m grid cells. The hillslope had bare soil cover approximating SNOTEL open area conditions. Soil hydraulic properties were selected to produce a subsurface time of concentration (t_c) for the hillslope of 24 h (for consistence, the same t_c will be used with the TR-55 event-based method as discussed in section 2.4). The specified soil parameters are shown in Figure 3 and the hillslope configuration in DHSVM was confirmed to have a 24 h t_c using a steady-state test with constant rainfall input (i.e., the time in reaching steady state is 24 h). Note that the relatively long t_c (24 h) of the 40 m hillslope is more suitable for natural/predevelopment subsurface flow-dominated conditions, such as the soil matrix flow and flow through macropores or soil pipes (Beven & Germann, 1982; Yan, Sun, Wigmosta, Skaggs, Hou, et al., 2019). Different configurations of the experimental hillslope (e.g., slope, soil properties, and t_c) will be used in the future to test the NG-IDF performance under a range of conditions.

[Place Figure 3 here]

We evaluated the NG-IDF method at the experimental hillslope scale rather than at a watershed scale that is realistic for each SNOTEL station because we wanted to isolate the effects of just the NG-IDF curves and avoid confounding factors (e.g., areal reduction factor) resulting from calibrating a TR-55 model based on a design flood estimated from DHSVM streamflow simulations. In this case, we designed the experimental hillslope based on a "known" time of concentration and confirmed that the SNOTEL observation can represent the spatial variability. The use of the same experimental hillslope at all stations allowed us to examine and compare the performances of the NG-IDF method across the entire western United States under various hydroclimatic conditions, thereby providing a more robust benchmark report.

With the experimental hillslope used for each SNOTEL station, DHSVM was forced by 15-min meteorological inputs (consistent with the steady-state simulation resolution) consisting of precipitation, air temperature, humidity, wind speed, and downward shortwave and longwave radiation. Using the Mountain Microclimate Simulation Model (Hungerford, Nemani, Running, & Coughlan, 1989), daily meteorological observations collected at each SNOTEL were disaggregated into 15-min temporal step. The DHSVM continuous simulations were used for two purposes: 1) to derive design floods as performance benchmarks through flood frequency analysis on the continuous 15-min simulated streamflow; and 2) to derive a daily time series (i.e., by aggregating the 15-min output) of annual maximum water available for runoff (W_{dhsvm}) for NG-IDF curves development. This will be described in detail in the section 2.5.

2.4 TR-55 event-based rainfall-runoff model

The TR-55 guideline (Cronshey et al., 1986) uses the curve number (CN) method to convert a design storm depth (determined from IDF curves with a predefined design hyetograph) to a runoff depth by taking into account watershed characteristics and initial moisture conditions. Runoff is then transformed into a design hydrograph using a unit hydrograph (UH) approach (Mockus, 1957). Because the mathematical formulations of the CN and UH methods have been described by a large and sufficient body of prior literature, only a brief introduction is presented in this paper.

TR-55 is an empirical, lumped, single event-based, rainfall-runoff model. The CN is a function of land use, hydrologic soil group (HSG), and antecedent moisture condition (AMC). The CN values under the average AMC for different land use and HSGs can be determined based on Table 2-2 of the TR-55 guideline. The CN values for dry and wet AMCs can be estimated subsequently based on the CN value under average AMC following Ponce & Hawkins (1996). The dimensionless NRCS UH has two parameters, time to peak and peak streamflow, which are empirically estimated based on watershed drainage area and time of concentration (t_c). In the UH method, the duration of unit excess rainfall (ΔD) was set to 15-min to be consistent with the 15-min DHSVM streamflow simulation temporal scale. For details on the TR-55 method, readers are referred to Cronshey et al. (1986).

Based on the specified soil parameters for the experimental bare ground hillslope (Figure 3), the HSG was type A leading to CN=77 under average AMC and CN=88 under wet AMC. To estimate the potential peak flood magnitude, the wet AMC is used in this study across all stations.

In the UH method, we used a t_c of 24 h that is consistent with the soil hydraulic properties and hillslope configuration used in the DHSVM continuous simulation to allow a direct comparison between the two methods. In the event-based method, it is common to use a design storm duration equal to t_c (Levy & McCuen, 1999; McCuen, 1998) or as required by local regulatory agencies (SCDM, 2016; SMMEW, 2004). In this study, the 24 h storm duration is the finest temporal resolution limited by the daily resolution of SNOTEL observations. We chose the uniform hyetograph proposed in the rational method (Pilgrim & Cordery, 1993) to represent the design storm temporal pattern for both PREC-IDF and NG-IDF curves at each selected duration (1–30 days). We used the uniform hyetograph because the standard design storm hyetographs are all developed for rainfall of short durations, e.g. NRCS hyetographs and National Oceanic and Atmospheric Administration (NOAA) Atlas 14 hyetographs (McCuen, 1998; Perica et al., 2013); no hyetograph has been developed or even studied for snowmelt or water available for runoff, which can last for days to weeks (Yan, Sun, Wigmosta, Skaggs, Hou, et al., 2019). Also, the use of uniform hyetograph is the logical starting point for the NG-IDF curve assessment because it provides the simplest representation of extreme event temporal variability.

We used the "critical design duration" method (Rogger et al., 2012; Yan, Sun, Wigmosta, Skaggs, Hou, et al., 2019) to identify potential peak design floods to make a fair comparison with results from flood frequency analysis of annual maximum DHSVM simulated streamflow. For each selected return period, the duration (ranged from 1–30 days) that produced the largest flood peak was selected to be the critical design duration and used in the following comparisons.

2.5 NG-IDF curves and DHSVM design floods

For each selected station, daily time series of water available for runoff (i.e., rainfall plus snowmelt) from SNOTEL observations and DHSVM simulations were constructed through mass balance approach as described in the introduction section. Based on these two daily time series, we extracted annual maximum W_{obs} and W_{dhsvm} at each SNOTEL station at selected durations varying from 1–30 days. We then validated the DHSVM performance in reproducing annual maximum water available for runoff by comparing the two annual maximum W data sets. For each selected duration, the annual maximum W_{dhsvm} was validated against the W_{obs} using the nonparametric Kolmogorov-Smirnov (K-S) test (Massey, 1951).

After successful model validation, the generalized extreme value (GEV) distribution, recommended by NOAA Atlas 14 (Perica et al., 2013), was fitted to the annual maximum data set of W_{dhsvm} based on L-moments statistics (Hosking & Wallis, 1997). L-moment estimators are an analogue to conventional method-of-moment estimators, but are weighted linear combinations of order statistics. Previous studies have demonstrated that L-moment method is more robust than method-of-moment in hydrologic frequency analysis (Hosking, 1990; Vogel & Fennessey, 1993). Frequency analyses were carried out on annual maximum W_{dhsvm} series for 30 durations varying from 1–30 days. For each duration, three GEV parameters (location, scale, and shape) were estimated using L-moments statistics; and then the associated quantiles for three exceedance probabilities: 0.1, 0.04, and 0.02, which correspond to extreme events with return periods of 10, 25, and 50 years, were derived to develop the NG-IDF curves. Because of the relatively short length of the SNOTEL records (Figure 2), the return period was cut off at 50-year and the annual maximum data set of W_{dhsvm} was assumed to be stationary in the frequency analysis. For a long-term data set, we acknowledge that a trend analysis is necessary before frequency analysis to verify the stationarity assumption, such as in Yan et al. (2018) and Yan, Sun, Wigmosta, Skaggs, Hou, et al. (2019). For the purpose of comparison, we also developed the standard PREC-IDF curves based on the observed annual maximum P following the same procedure.

Similar to the frequency analysis applied to develop NG-IDF curves, the GEV distribution was also fitted to the DHSVM annual maximum 15-min streamflow data (assuming stationary) extracted from the time series of DHSVM simulated streamflow using the L-moments statistics. The design floods were then estimated for the return periods of 10, 25, and 50 years and used as the performance benchmark for later assessment. In addition, a Monte Carlo (MC) simulation procedure suggested by Hosking & Wallis (1997) and NOAA Atlas 14 (Perica et al., 2013) was used to consider sample data uncertainty in frequency analysis. After estimating the parameters of GEV distribution using the L-moments statistics at each station, a total of 1,000 MC synthetic data sets were generated with the same record length and sample L-moments. We then fitted GEV distribution to each MC synthetic data set using the L-moments statistics and estimated the associated design floods of selected return periods. Therefore, a total of 1,000 ensemble members were generated to quantify the uncertainties associated with the PREC-IDF and NG-IDF curves and DHSVM design flood statistics. In this study, all L-moments and MC analyses were performed using the "lmom" package (version 2.6) (Hosking, 2017) in R (version 3.4.3).

Besides the deterministic assessment (i.e., relative differences) between the two design flood estimates, the MC results were used to determine if these differences were statistically significant with the consideration of uncertainties. After characterizing the uncertainties using the 1,000 MC ensemble members, for each return period, we used the Z statistic to quantify the statistically significant differences of the design flood estimates between the IDF event-based method (q_{event}) and DHSVM continuous-simulation method (q_{dhsvm}) (Ganguli & Coulibaly, 2017; Madsen et al., 2009; Mikkelsen, Madsen, Arnbjerg-Nielsen, Rosbjerg, & Harremoës, 2005):

$$Z = \frac{q_{dhsvm} - q_{event}}{\sqrt{0.5(s_{dhsvm}^2 + s_{event}^2)}} \tag{1}$$

where q_{dhsvm} is the design flood obtained from the DHSVM flood frequency analysis, q_{event} is the design flood estimated from the TR-55 event-based method with input from PREC-IDF curves or NG-IDF curves, s_{dhsvm} is the DHSVM design flood standard deviation obtained from the 1,000 DHSVM design flood ensemble members, and s_{event} is the standard deviation of event-based model design flood estimated from the 1,000 design flood ensemble members generated by applying the 1,000 PREC-IDF or NG-IDF ensemble members to the TR-55 event-based model. Under the 5% significance level, $|Z| \leq 1.96$ represents the associated 95% confidence interval and suggests accepting the null hypothesis that there is no significant difference between the two methods.

3 RESULTS AND DISCUSSION

In the following, the analyses performed for both PREC-IDF and NG-IDF methods and DHSVM continuous simulation are reported. For each station, the validation of the annual maximum water available for runoff from DHSVM (W_{dhsvm}) is first described in section 3.1. Second, for each selected return period, the peak design floods derived from both PREC-IDF and NG-IDF methods are compared to the corresponding DHSVM benchmark. By comparing the results, we benchmark the performance of the NG-IDF curves in design flood estimates for infrastructure design in snow-dominated regions (sections 3.2 and 3.3).

3.1 Validation of DHSVM annual maximum water available for runoff

Figure 4a presents the nonparametric K-S test results between the daily annual maximum water available runoff from DHSVM simulations (W_{dhsvm}) and SNOTEL observations (W_{obs}). About 82% of 246 stations were found to accept the null hypothesis that the two data sets are drawn from the same distribution at 5% significance level, indicating no statistically significant differences between their derived NG-IDF values. Figure 4b shows the mean relative errors between the two daily annual maximum data sets ($(W_{dhsvm} - W_{obs})/W_{obs} \times 100$). It is observed that the SNOTEL stations with greater errors (more than 20%) are consistent to some extent with the stations showing statistically significant differences ($p \leq 5\%$) in the K-S test. Calver et al. (2009) recommended that errors within the 20% threshold are very good in practice given this hydrologically challenging context. Our results showed that about 94% of the 246 stations were found to have errors within 20% and about 60% of the 246 stations had errors within 10%, indicating a good performance of DHSVM in simulating annual maximum water available for runoff across the western United States. Due to the page limitation, we only showed the validation result for 1-day duration in this paper; similar or better results were found for other storm durations ranging from 2–30 days. For example, about 94% of the 246 stations showed no statistically significant differences ($p > 5\%$) in the K-S test of the 30-day annual maximum water available for runoff values. In summary, these results are consistent with Sun et al. (2019), which confirmed the good performance of the calibrated DHSVM in SWE predictions across the western United States.

[Place Figure 4 here]

Figure S1 further showed the relative errors between the two NG-IDF curves for the three selected return periods derived based on annual maximum W_{dhsvm} and W_{obs} data set. It is observed that the relative errors between NG-IDF curves were larger than the errors between the two annual maximum W data sets. This is because besides the hydrological modeling uncertainties (model structure and model parameter) as shown in Figure 4, NG-IDF curves also include data sample uncertainty in frequency analysis. For instance, the 50-year event presented larger errors than the 10-year event because all selected SNOTEL annual maximum records were less than 30 years. Calver et al. (2009) suggested that errors within 20% are currently very good and errors up to about 35% may have to be contended with in practice. Our results showed that, on average of the three return periods, about 90% and 91% of the 246 stations were found to have errors within 35%, and about 74% and 72% of the 246 stations had errors within 20% for the 1-day and 2-day duration, respectively. These results further confirmed the good performance of the calibrated DHSVM in its modeling snow process. Due to the page limitation, we only showed the comparisons for 1-day and 2-day duration in this paper; similar results were found for other storm durations ranging from 3–30 days. These results also emphasized that data sample uncertainties should be considered in the NG-IDF method assessment. In the following NG-IDF method assessment, we treated DHSVM estimates as synthetic truths and therefore the hydrologic model uncertainty can be ignored. The only uncertainty associated with NG-IDF curves were the data sample uncertainties in frequency analysis. We also examined the elevation distribution of the relative errors in annual maximum W time series (Figure S2). It is observed that errors ranging from -30% to +30% occurred at all elevations from 1000 m to 3500 m, which suggests DHSVM showed no systematic biases in W estimations across the mountain ranges of the western United

States. Sun et al. (2019) suggested lower model skills could be attributed to the underrepresented wind-related snow processes and model simulated energy balance bias.

3.2 IDF-based modeling vs. DHSVM continuous simulation

The DHSVM design flood estimates (q_{dhsvm}) were used as benchmark performance to evaluate the peak design flood estimates from both PREC-IDF (q_{PREC}) and NG-IDF (q_{NG}) curves. Figure 5 presents the scatterplots of q_{PREC} or q_{NG} versus q_{dhsvm} for the 10-, 25-, and 50-year return periods. Figure 6 presents spatially the relative errors between q_{PREC} or q_{NG} and q_{dhsvm} across the western United States.

[Place Figures 5, 6 here]

The results showed that for all three selected return periods, q_{PREC} produced larger errors than q_{NG} in reference to q_{dhsvm} at most of the 246 stations. About 73%, 69%, and 66% of the 246 stations indicated the relative errors between q_{PREC} and q_{dhsvm} larger than ±20% for the 10-, 25-, and 50-year events, respectively. Using the NG-IDF curves, the percentage of stations with relative errors larger than ±20% is reduced substantially to 18% for the 10-year event, 15% for the 25-year event, and 13% for the 50-year event. These differences are mainly attributed to the exclusion of snowmelt process in the PREC-IDF curves. For example, the considerable underestimation of design floods in the continental regime (Colorado Front Range, Wasatch Mountains, and Rocky Mountains) and intermountain regime (Blue Mountains and Northern Basin and Range) is because the magnitude of extreme precipitation events are smaller than the extreme snowmelt events (Yan et al., 2018). Note that a few stations showed overestimation when using the PREC-IDF curves in both figures. These stations are primarily located in the southern Sierra Nevada and Arizona/New Mexico Mountains, where the PREC-IDF values were significantly higher than the NG-IDF values (Yan, Sun, Wigmosta, Skaggs, Hou, et al., 2019). In the southern Sierra Nevada, the higher values of PREC-IDF than NG-IDF is caused by heavy winter precipitation (due to atmospheric rivers) and slower melt (Yan, Sun, Wigmosta, Skaggs, Leung, et al., 2019). In the Arizona/New Mexico Mountains, the overall warm conditions in winter result in shallow snowpack conditions and small snowmelt events in mid-winter to early spring (Yan et al., 2018).

In Figure 6, it is also observed that compared to q_{dhsvm}, q_{NG} showed a consistent underestimation of design floods across the western United States. As the wet AMC and critical design duration were used to avoid the potential underestimation of design flood, the tendency of underestimation is most likely attributed to the choice of uniform hyetograph. In general, a design hyetograph with a pronounced peak tends to produce a larger flood peak (Alfieri, Laio, & Claps, 2008; Grimaldi et al., 2012; Hettiarachchi, Wasko, & Sharma, 2018; Wasko & Sharma, 2015). These results suggest the importance of the choice of design hyetograph for peak flood estimates, and the potential bias from using the uniform hyetograph in NG-IDF approach for the reasons detailed in section 2.4. The limited choice of hyetographs for melt-driven events with a long duration presents an opportunity of future improvement in the NG-IDF design approach.

In summary, the mean absolute errors across all 246 SNOTEL stations between q_{PREC} and q_{dhsvm} were about 33%, 31%, and 31% for the 10-, 25-, and 50-year events, respectively. In contrast, the average absolute errors between q_{NG} and q_{dhsvm} were about 15% for the 10-year event, 14% for the 25-year event, and 12% for the 50-year event. Despite of the use of uniform hyetograph in the NG-IDF modeling, the average error across the western United States still showed promising performance (less than the 20% threshold) for all three selected flood return periods. In contrast, the PREC-IDF curves can significantly underestimate the design floods in snow-dominated areas, posing a serious risk for infrastructure design in these locations.

Besides the uniform hyetograph, here we further examined the sensitivity of W hyetograph selection on design flood estimates. Specifically, we tested the standard rainfall NRCS Type II hyetograph and the commonly used triangular hyetograph in hydrologic design (McCuen, 1998). The NRCS Type II rainfall hyetograph uses a center-loaded pattern and is recommended for hydrologic design in the western United States. The NRCS hyetograph is developed up to 24 h duration, therefore only the 24 h W event can be used in TR-55 modeling, which may underestimate the potential peak design flood. In the triangular hyetograph, the W intensity is equal to zero at the beginning of the event, increases linearly to the middle of the event, and then decreases linearly to the end of the event when it is again equal to zero. The top and bottom panel of Figure S3 presents the scatterplots of q_{NG} with the use of the NRCS Type II rainfall hyetograph and the triangular hyetograph versus q_{dhsvm} for the 10, 25, and 50-year return periods, respectively. It is observed that with the use of standard NRCS Type II hyetograph, q_{NG} showed either underestimation or overestimation of DHSVM design floods. The q_{NG} tended to show underestimation for small W events and overestimation for large W events, suggesting standard rainfall hyetograph is inappropriate for modeling snowmelt and ROS events. Contrary to the consistent underestimation with the use of uniform hyetograph, consistent overestimation is observed for q_{NG} with the use of the triangular hyetograph, which indicates the center-loaded triangular hyetograph intensifies the W temporal patterns. In summary, these results further demonstrate the need to develop hyetographs for water available for runoff to improve NG-IDF performances.

3.3 Evaluation of NG-IDF curves with uncertainty quantification

Based on the MC method described in section 2.5, we quantified the sample data uncertainties associated with q_{PREC}, q_{NG}, and q_{dhsvm} with 1,000 ensemble members. Figure 7 displays the boxplot of the Z statistics showing the statistical significance of the relative differences between q_{PREC} and q_{dhsvm} (red boxplot), and q_{NG} and q_{dhsvm} (blue boxplot). A positive (negative) Z value indicates q_{PREC} or q_{NG} underestimated (overestimated) q_{dhsvm}. The grey bound in the figure shows the 95% confidence interval of the Z statistics ($|Z| \leq 1.96$), and estimated Z statistics within the grey bound suggests accepting the null hypothesis that there is no significant difference (5% significance level) between q_{PREC} or q_{NG} and q_{dhsvm}. In Figure 7, it can be observed that about 29%, 48%, and 65% of stations showed no statistically significant differences between q_{PREC} and q_{dhsvm} for the 10-, 25-, and 50-year events, respectively. While using the NG-IDF curves, the percentage of stations showing no significant differences increased to 56% for the 10-year event, 82% for the 25-year event, and 91% for the 50-year event. The higher percentage of stations showing no significant differences at larger return period may be due to the larger uncertainties within these estimates using a relatively short length of record (i.e., ≤28 years). Note that Figure 7 shows the Z statistics instead of design flood uncertainty. The design flood uncertainty (e.g., s_{dhsvm} and s_{event}) increased with longer return periods, leading to smaller Z values. These results demonstrate the potential of the NG-IDF design method, with significant improvement expected as snowmelt hyetographs are developed. These results also reinforce the role of standard IDF event-based method in future infrastructure design. However, we acknowledge that these results are at the experimental hillslope scale where point estimates of water available for runoff (W) can represent its spatial variability.

[Place Figure 7 here]

4 CONCLUSIONS AND FUTURE WORK

This study benchmarked the performances of NG-IDF curves in flood estimates for infrastructure design in snow-dominated regions. We used DHSVM, a physics-based hydrologic model, to simulate the snow accumulation and ablation process and derive the continuous streamflow, which were used as performance benchmarks to evaluate the design flood estimates from the TR-55 event-based model with input from NG-IDF curves. These comparisons were performed through an experimental hillslope at 246 SNOTEL stations across different hydroclimate regimes of the western United States. The NG-IDF curves were developed at each station based on the DHSVM simulated water available for runoff forced by SNOTEL meteorological data. Standard PREC-IDF curves were also developed based on SNOTEL precipitation observations for comparison purposes. Sample data uncertainties in both DHSVM flood and NG-IDF frequency analysis were also included in the evaluation.

Based on the results in this paper, there are three major findings. 1) Compared to the design flood estimates from DHSVM continuous simulation, the PREC-IDF method provided consistently biased design flood estimates and showed large relative errors due to the neglect of snow process. The NG-IDF method generated much smaller errors in the design flood estimates for return periods between 10 and 50 years. For example, the average absolute error for the 50-year flood across all 246 stations was reduced from 31% with the use of PREC-IDF curves to 12% with the use of NG-IDF curves. 2) Out of the total 246 stations, only about 31% of the stations showed errors less than 20% with the use of PREC-IDF curves, while about 85% locations showed errors less than 20% with the use of NG-IDF curves, suggesting a satisfactory performance of NG-

IDF curves in practical design in snow-dominated regions. 3) Although advanced physics-based hydrologic models can better represent the hydrological process than simple event-based models, it does not necessarily lead to a statistically significant difference in design flood estimates at the tested hillslope scale.

This benchmark report also provides two main suggestions that can benefit practicing engineers and local jurisdictions responsible for surface water design. 1) The "water available for runoff" concept should be further evaluated (e.g., hyetographs for W) as a potential replacement for the current PREC-IDF based approach. 2) Systematic and consistent surface water design manuals are not available at present for snow environments of the United States and the NG-IDF curves approach can fill this gap. It is expected that the IDF-based technique will continue to play a crucial role in infrastructure design (especially for small-scale infrastructure) in the foreseeable future. Adaptations of PREC-IDF to NG-IDF design methods, rather than a complete technological change to a physics-based hydrologic model method, is more easily implemented by agencies and practicing engineers in the near-term because the technology transfer process is straightforward.

Finally, we acknowledge that further investigations are necessary to enhance the NG-IDF method in practicing design. Specifically, we outline the following three directions. 1) Develop an appropriate hyetograph shape for water available for runoff (W). All proposed hyetograph shapes in literature are for rainfall only within short duration (i.e., up to 96 h), while the W from snowmelt lasts much longer and is further complicated by the day-night energy cycle. We plan to use the DHSVM 15-min W time series to construct regional hyetographs for melt-driven events and test them against difference hillslope configurations in the evaluation framework. 2) Provide W data products for different vegetation cover (i.e., consistent with the TR-55 land cover type). Vegetation impacts snow process through canopy interception and sublimation, and changes snowpack energy balance by attenuating incoming shortwave radiation, enhancing longwave radiation, and reducing wind speed, and thus the turbulent heat fluxes. This evaluation study was based on bare ground approximating the SNOTEL open condition. Infrastructure design at the small-basin scale may have multiple types of land cover; the average W for the basin is a weighted combination of W for different land cover. We plan to use the well-validated DHSVM to derive W data products for different vegetation cover (e.g., forest, shrub) that is consistent with the TR-55 land cover classification and can be adapted to the standard IDF design workflow. 3) Examination of areal reduction factor (ARF) for W. For small basins (e.g., the hillslope used in this study), it is acceptable to assume the point W rate falls uniformly across the basin; however, it is unrealistic to assume uniform W for larger basins. In design at larger basin scale, ARF is commonly used to convert point depths to basin depths by accounting for the spatial variability of storm intensity in a basin. We plan to run the well-validated DHSVM over the conterminous United States (CONUS) at 1/16° spatial resolution and estimate the ARFs for different basins (e.g., HUC10) by calculating the ratio between the average W for the basin and the areal maximum W for different storm durations over CONUS.

ACKNOWLEDGEMENTS

All daily SNOTEL data used in this study are available from the Natural Resources Conservation Service and National Water and Climate Center at https://www.wcc.nrcs.usda.gov/snow/ (last access 7 January 2019). The United States boundaries shown in the figures were acquired from United States Geological Survey Small-scale Dataset State Boundaries of the United States 200506 Shapefile at https://catalog.data.gov/dataset/usgs-small-scale-dataset-state-boundaries-of-the-united-states-200506-shapefile (last access 17 February 2017). This material is based upon work supported by the Strategic Environmental Research and Development Program under Contract No. RC-2546. Battelle Memorial Institute operates the Pacific Northwest National Laboratory for the U.S. Department of Energy under contract DE-AC06-76RLO-1830. The DHSVM source code and BCQC SNOTEL data used in this paper are available at https://dhsvm.pnnl.gov/.

Data Availability Statement

All daily SNOTEL data used in this study are available from the Natural Resources Conservation Service and National Water and Climate Center at https://www.wcc.nrcs.usda.gov/snow/ (last access 7 January 2019). The United States boundaries shown in the figures were acquired from United States Geological Survey Small-scale Dataset State Boundaries of the United States 200506 Shapefile at https://catalog.data.gov/dataset/usgs-small-scale-dataset-state-boundaries-of-the-united-states-200506-shapefile (last access 17 February 2017). The DHSVM source code and BCQC SNOTEL data used in this paper are available at https://dhsvm.pnnl.gov/.

CONFLICT OF INTEREST

The authors state no conflicts of interest.

ORCID

Hongxiang Yan: https://orcid.org/0000-0002-2387-403X

Ning Sun: https://orcid.org/0000-0002-4094-4482

Mark Wigmosta: https://orcid.org/0000-0002-2918-8284

L. Ruby Leung: https://orcid.org/0000-0002-3221-9467

Zhangshuan Hou: https://orcid.org/0000-0002-9388-6060

Andre Coleman: https://orcid.org/0000-0002-3221-3727

Figure Captions

Figure 1. The framework for evaluating the NG-IDF method in design flood estimates using DHSVM continuous simulation method.

Figure 2. (a) The length of record in water years of the selected 246 SNOTEL stations in the snow-dominated western United States. (b) The mean annual peak SWE value in centimeters for each SNOTEL station over the total length of record.

Figure 3. DHSVM experimental hillslope and the specified soil parameters. A snowpack layer lying on the hillslope and the rain falling on the hillslope represent the snowmelt and ROS conditions. Water available for runoff from rain, snowmelt, and/or ROS infiltrated into the soil and generates subsurface flow to the outlet.

Figure 4. (a) The nonparametric K-S test results for the two daily annual maximum data sets of water available for runoff from DHSVM simulations (W_{dhsvm}) and SNOTEL observations (W_{obs}). A p-value greater than 5% suggests accepting the null hypothesis that the two data sets are drawn from the same distribution (5% significance level); while a p-value less or equal to 5% indicates rejecting the null hypothesis. (b) The mean relative errors between the two daily annual maximum data sets of water available for runoff ($(W_{dhsvm} - W_{obs})/W_{obs} \times 100$).

Figure 5. Top panel: scatterplots of the (a) 10-year, (b) 25-year, and (c) 50-year normalized peak design floods derived from PREC-IDF curves versus normalized design floods from DHSVM continuous simulations for the 246 SNOTEL stations across the western United States. Bottom panel: scatterplots of the (d) 10-year, (e) 25-year, and (f) 50-year normalized peak design floods derived from NG-IDF curves versus normalized design floods from DHSVM continuous simulations. For both panels, the straight line indicates the 1:1 line. All design floods were normalized by the maximum DHSVM design flood across all 246 stations, separately for each return period.

Figure 6. Top panel: the relative peak design flood error ($(q_{PREC} - q_{dhsvm})/q_{dhsvm} \times 100$) for the (a) 10-year event, (b) 25-year event, and (c) 50-year event between the peak design floods derived from PREC-IDF (q_{PREC}) and DHSVM continuous simulations (q_{dhsvm}). Bottom panel: the relative peak design flood error ($(q_{NG} - q_{dhsvm})/q_{dhsvm} \times 100$) for the (d) 10-year event, (e) 25-year event, and (f) 50-year event between the peak design floods derived from NG-IDF method (q_{NG}) and q_{dhsvm}. For both panels, stations with warm color indicate over-predictions; stations with blue color suggest under-predictions.

Figure 7. The box plot of the Z statistics for peak design flood estimates from PREC-IDF curves and NG-IDF curves over the 246 SNOTEL stations across the western United States. The Z statistic represents the statistical significance of differences in design floods obtained from event-based model versus DHSVM continuous simulations. A positive (negative) Z value indicates

event-based model underestimated (overestimated) the design flood from DHSVM. The Z statistics within the 95% confidence interval (shown as grey bound) suggests accepting the null hypothesis that there is no statistically significant difference (5% significance level) between the two design flood estimates.

Figure S1. The mean relative errors between NG-IDF curves based on W_{dhsvm} and W_{obs} for the selected three return periods and two durations.

Figure S2. The mean relative errors between the W_{dhsvm} and W_{obs} daily annual maximum data sets of water available for runoff versus the SNOTEL elevation.

Figure S3. Top panel: scatterplots of the (a) 10-year, (b) 25-year, and (c) 50-year normalized peak design floods derived from NG-IDF curves with the use of standard rainfall NRCS Type II hyetograph versus normalized design floods from DHSVM simulation for the 246 SNOTEL stations across the western United States. Bottom panel: similar to top panel but showing the NG-IDF curves with the use of triangle hyetograph. For both panels, the straight line indicates the 1:1 line. All design floods were normalized by the maximum DHSVM design flood across all 246 stations, separately for each return period.

REFERENCES

Alfieri, L., Laio, F., & Claps, P. (2008). A simulation experiment for optimal design hyetograph selection. *Hydrological Processes*, *22*(6), 813–820. https://doi.org/10.1002/hyp.6646

Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, *18*(5), 1311–1325. https://doi.org/10.1029/WR018i005p01311

Boughton, W., & Droop, O. (2003). Continuous simulation for design flood estimation—a review. *Environmental Modelling & Software*, *18*(4), 309–318. https://doi.org/10.1016/S1364-8152(03)00004-5

Boughton, W., Srikanthan, S., & Weinmann, E. (2002). Benchmarking a New Design Flood Estimation System. *Australasian Journal of Water Resources*, *6*(1), 45–52. https://doi.org/10.1080/13241583.2002.11465209

Calver, A., Stewart, E., & Goodsell, G. (2009). Comparative analysis of statistical and catchment modelling approaches to river flood frequency estimation. *Journal of Flood Risk Management*, *2*(1), 24–31. https://doi.org/10.1111/j.1753-318X.2009.01018.x

Camici, S., Tarpanelli, A., Brocca, L., Melone, F., & Moramarco, T. (2011). Design soil moisture estimation by comparing continuous and storm-based rainfall-runoff modeling. *Water Resources Research*, *47*(5), W05527. https://doi.org/10.1029/2010WR009298

Cao, Q., Sun, N., Yearsley, J., Nijssen, B., & Lettenmaier, D. P. (2016). Climate and land cover effects on the temperature of Puget Sound streams. *Hydrological Processes*, *30*(13), 2286–2304. https://doi.org/10.1002/hyp.10784

Carnevale, A. P., & Smith, N. (2017). *Trillion Dollar Infrastructure Proposals Could Create*

Millions of Jobs. Center on Education and the Workforce, Georgetown University.

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill, New York.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., … Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 2498–2514. https://doi.org/10.1002/2015WR017198

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.

Curran, J., Barth, N., Veilleux, A. G., & Ourso, R. (2016). *Estimating flood magnitude and frequency at gaged and ungaged sites on streams in Alaska and conterminous basins in Canada, based on data through water year 2012*. U.S. Geological Survey Scientific Investigations Report 2016–5024. https://doi.org/10.3133/sir20165024

England, J. F., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas Jr., W. O., Veilleux, A. G., … Mason Jr., R. R. (2018). *Guidelines for determining flood flow frequency—Bulletin 17C*. *Techniques and Methods*. Reston, VA. https://doi.org/10.3133/tm4B5

Ganguli, P., & Coulibaly, P. (2017). Does nonstationarity in rainfall require nonstationary intensity--duration--frequency curves? *Hydrology and Earth System Sciences*, *21*(12), 6461–6483. https://doi.org/10.5194/hess-21-6461-2017

Grimaldi, S., Petroselli, A., & Serinaldi, F. (2012). Design hydrograph estimation in small and ungauged watersheds: continuous simulation method versus event-based approach. *Hydrological Processes*, *26*(20), 3124–3134. https://doi.org/10.1002/hyp.8384

Hamlet, A. F. (2018). New Observed Data Sets for the Validation of Hydrology and Land Surface Models in Cold Climates. *Water Resources Research*, *54*(8), 5190–5197. https://doi.org/10.1029/2018WR023123

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. *Hydrology and Earth System Sciences*, *22*(3), 2041–2056. https://doi.org/10.5194/hess-22-2041-2018

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *52*, 105–124. https://doi.org/10.2307/2345653

Hosking, J. R. M. (2017). Package “lmom.” https://CRAN.R-project.org/package=lmom.

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., … Coleman, A. (2019). Incorporating Climate Non-stationarity and Snowmelt Processes in Intensity-Duration-Frequency Analyses with Case Studies in Mountainous Areas. *Journal of Hydrometeorology*, JHM-D-19-0055.1. https://doi.org/10.1175/JHM-D-19-0055.1

Hungerford, R. D., Nemani, R. R., Running, S. W., & Coughlan, J. C. (1989). *MTCLIM: A Mountain Microclimate Simulation Model*. Ogden, Utah: U.S. Department of Agriculture.

Levy, B., & McCuen, R. (1999). Assessment of Storm Duration for Hydrologic Design. *Journal of Hydrologic Engineering*, *4*(3), 209–213. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(209)

Li, D., Wrzesien, M. L., Durand, M., Adam, J., & Lettenmaier, D. P. (2017). How much runoff originates as snow in the western United States, and how will that change in the future? *Geophysical Research Letters*, *44*(12), 6163–6172. https://doi.org/10.1002/2017GL073551

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., … Nijssen, B. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, *26*(23), 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., … Dettinger, M. D. (2016). Yosemite Hydroclimate Network: Distributed stream and atmospheric data for the Tuolumne River watershed and surroundings. *Water Resources Research*, *52*(9), 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. *Atmospheric Research*, *92*(3), 343–349. https://doi.org/https://doi.org/10.1016/j.atmosres.2009.01.013

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78. https://doi.org/10.1080/01621459.1951.10500769

McCuen, R. H. (1998). *Hydrologic analysis and design* (2nd ed.). Prentice-Hall, Upper Saddle River, N.J.

Mikkelsen, P. S., Madsen, H., Arnbjerg-Nielsen, K., Rosbjerg, D., & Harremoës, P. (2005). Selection of regional historical rainfall time series as input to urban drainage simulations at ungauged locations. *Atmospheric Research*, *77*(1), 4–17. https://doi.org/https://doi.org/10.1016/j.atmosres.2004.10.016

Mockus, V. (1957). *Use of storm and watersheds characteristics in synthetic hydrograph analysis and application*. U.S. Department of Agriculture, Soil Conservation Service.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nnadi, F. N., Kline, F. X., Wray, H. L., & Wanielista, M. P. (1999). Comparison of design storm concepts using continuous simulation with short duration storms. *Journal of the American Water Resources Association*, *35*(1), 61–72. https://doi.org/10.1111/j.1752-1688.1999.tb05452.x

Pathiraja, S., Westra, S., & Sharma, A. (2012). Why continuous simulation? The role of antecedent moisture in design flood estimation. *Water Resources Research*, *48*(6), W06534. https://doi.org/10.1029/2011WR010997

Perica, S., Martin, D., Pavlovic, S., Roy, I., Laurent, M. St., Trypaluk, C., … 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.

Pilgrim, D. H., & Cordery, I. (1993). Flood runoff. In D. R. Maidment (Ed.), *Handbook of Hydrology, vol. 9* (pp. 1–42). McGraw-Hill, New York.

Ponce, V. M., & Hawkins, R. H. (1996). Runoff Curve Number: Has It Reached Maturity? *Journal of Hydrologic Engineering*, *1*(1), 11–19. https://doi.org/10.1061/(ASCE)1084-0699(1996)1:1(11)

Rogger, M., Kohl, B., Pirkl, H., Viglione, A., Komma, J., Kirnbauer, R., … Blöschl, G. (2012). Runoff models and flood frequency statistics for design flood estimation in Austria – Do they tell a consistent story? *Journal of Hydrology*, *456–457*, 30–43. https://doi.org/10.1016/j.jhydrol.2012.05.068

SCDM. (2016). *Snohomish County Drainage Manual Volume III Hydrologic Analysis and Flow Control BMPs*. Available online at: https://snohomishcountywa.gov/DocumentCenter/View/31221.

SMMEW. (2004). *Stormwater Management Manual for Eastern Washington*. Available online at: https://fortress.wa.gov/ecy/publications/documents/0410076.pdf: Washington State Department of Ecology.

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. *Hydrological Processes*, *32*(13), 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. *Journal of Geophysical Research: Atmospheres*, *124*(10), 5296–5313. https://doi.org/10.1029/2018JD030140

Sun, N., Yearsley, J., Baptiste, M., Cao, Q., Lettenmaier, D. P., & Nijssen, B. (2016). A spatially distributed model for assessment of the effects of changing land use and climate on urban stream quality. *Hydrological Processes*, *30*(25), 4779–4798. https://doi.org/10.1002/hyp.10964

Vogel, R. M., & Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*. https://doi.org/10.1029/93WR00341

Wasko, C., & Sharma, A. (2015). Steeper temporal distribution of rain intensity at higher temperatures within Australian storms. *Nature Geoscience*, *8*(7), 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 Resources Research*, *35*(1), 255–264. https://doi.org/10.1029/1998WR900017

Wigmosta, M. S., Nijssen, B., & Storck, P. (2002). The distributed hydrology soil vegetation model. In V. P. Singh (Ed.), *Mathematical Models of Small Watershed Hydrology and Applications* (pp. 7–42). Highlands Ranch, Colorado: Water Resources Publication.

Wigmosta, M. S., Vail, L. W., & Lettenmaier, D. P. (1994). A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, *30*(6), 1665–1679. https://doi.org/10.1029/94WR00436

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-Climate/Climate Change, and Environmental Engineering* (pp. 278–286). Pittsburgh, PA: American Society of Civil Engineers. https://doi.org/10.1061/9780784482346.028

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. *Journal of Hydrologic Engineering*, *24*(7), 04019020. 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 Resources Research*, *54*(2), 1093–1108. https://doi.org/10.1002/2017WR021290

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. *Geophysical Research Letters*, *46*(2), 767–775. https://doi.org/10.1029/2018GL080260

Zhao, G., Gao, H., Naz, B. S., Kao, S.-C., & Voisin, N. (2016). Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in Water Resources*, *98*, 16–31. https://doi.org/10.1016/j.advwatres.2016.10.014