



OPEN

DATA DESCRIPTOR

# Next-Generation Intensity-Duration-Frequency Curves for Diverse Land across the Continental United States

Hongxiang Yan<sup>1</sup>✉, Zhuoran Duan<sup>1,2</sup>, Mark S. Wigmosta<sup>1,2</sup>, Ning Sun<sup>1</sup>, Ethan D. Gutmann<sup>3</sup>, Bert Kruyt<sup>3</sup> & Jeffrey R. Arnold<sup>4</sup>

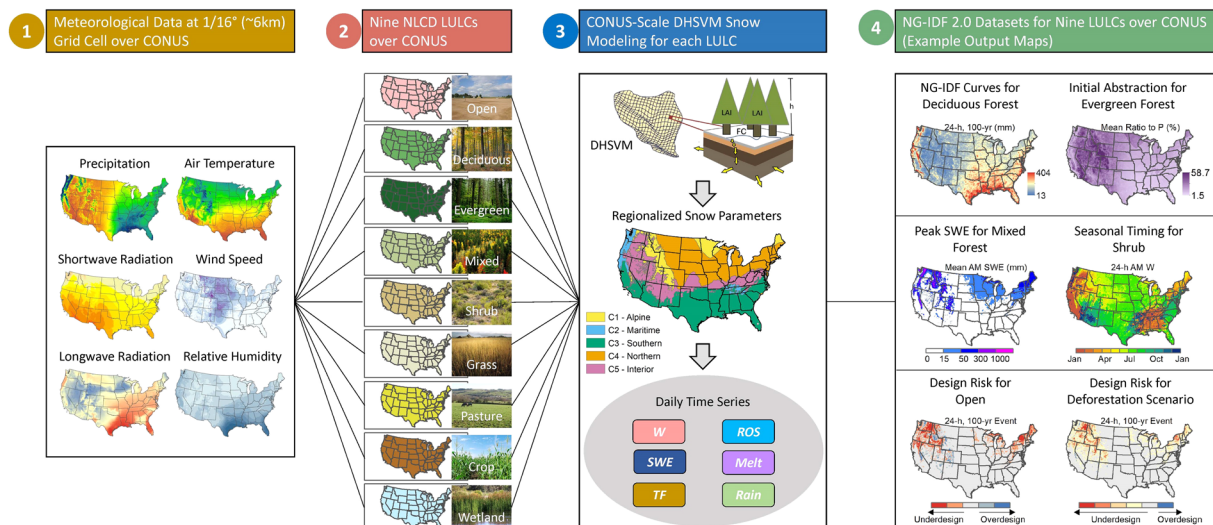
The current methods for designing hydrological infrastructure rely on precipitation-based intensity-duration-frequency curves. However, they cannot accurately predict flooding caused by snowmelt or rain-on-snow events, potentially leading to underdesigned infrastructure and property damage. To address these issues, next-generation intensity-duration-frequency (NG-IDF) curves have been developed for the open condition, characterizing water available for runoff from rainfall, snowmelt, and rain-on-snow. However, they lack consideration of land use land cover (LULC) factors, which can significantly affect runoff processes. We address this limitation by expanding open area NG-IDF dataset to include eight vegetated LULCs over the continental United States, including forest (deciduous, evergreen, mixed), shrub, grass, pasture, crop, and wetland. This NG-IDF 2.0 dataset offers a comprehensive analysis of hydrological extreme events and their associated drivers under different LULCs at a continental scale. It will serve as a useful resource for improving standard design practices and aiding in the assessment of infrastructure design risks. Additionally, it provides useful insights into how changes in LULC impact flooding magnitude, mechanisms, timing, and snow water supply.

## Background & Summary

Hydrological extreme events, such as floods and droughts, have a significant impact on human society and the natural environment<sup>1–4</sup>. In the western United States, over 50% of the water supply comes from mountain snowmelt<sup>5–7</sup>. Insufficient winter snowpack can result in water shortages and environmental strains during dry summer months<sup>8–11</sup>, whereas deep snowpack accompanied by warm temperatures and rain can result in rapid melting and consequent rain-on-snow (ROS) flooding. Without appropriate mitigative measures, these floods can cause extensive damage to infrastructure and property<sup>12–14</sup>. An instance of such a flood occurred in Yellowstone National Park in 2022, which forced the park's closure for the first time in 34 years and may cost more than \$1 billion for rebuilding damaged bridges and roads<sup>15</sup>.

At present, there is a lack of a consistent and systematic hydrological design approach for snow-dominated regions of the United States<sup>16–18</sup>. Local design manuals require or recommend the use of precipitation-based intensity-duration-frequency (PREC-IDF) curves, such as the National Oceanic and Atmospheric Administration Atlas 14<sup>19</sup>. However, the PREC-IDF method implicitly assumes that precipitation is in the form of rain and immediately available for the rainfall-runoff process, which can result in significant underestimation of flooding caused by snowmelt or ROS events (i.e., underdesign)<sup>7,20–22</sup>. For instance, the infrastructure in Yellowstone National Park was not constructed to withstand ROS flooding<sup>23</sup>. To address this need, Yan, *et al.*<sup>20</sup> proposed next-generation IDF (NG-IDF) curves, which enhances the PREC-IDF approach for hydrological design in regions dominated by both rainfall and snow. The NG-IDF curves characterize the water available for runoff (*W*) from rainfall, snowmelt, and ROS events. Yan, *et al.*<sup>24</sup> compared extreme events estimated from NG-IDF with those from PREC-IDF, using observed precipitation and snow water equivalent (SWE) data from almost 400 Snowpack Telemetry (SNOTEL) stations across the western United States. They discovered that

<sup>1</sup>Earth Systems Science Division, Pacific Northwest National Laboratory, Richland, WA, USA. <sup>2</sup>Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA. <sup>3</sup>National Center for Atmospheric Research, Boulder, CO, USA. <sup>4</sup>MITRE Corporation, McLean, VA, USA. ✉e-mail: hongxiang.yan@pnnl.gov



**Fig. 1** Diagram depicting the methodology used to generate NG-IDF 2.0 datasets for nine different LULCs across the CONUS and their resulting representation.

around 70% of these stations were subject to underdesign when using the PREC-IDF method, leading to underestimations of floods up to 324%. To expand the use of NG-IDF curves from SNOTEL stations to ungauged sites, Sun, *et al.*<sup>25</sup> employed a validated physics-based hydrological model, the Distributed Hydrology Soil Vegetation Model (DHSVM)<sup>26–28</sup>, to develop NG-IDF curves under open condition across the continental United States (CONUS) at a  $1/16^\circ$  ( $\sim 6$  km) resolution ( $>200,000$  sites).

Despite these advancements, NG-IDF research to date has been applied to open condition without accounting for the influence of land use land cover (LULC) on *W* such as canopy interception and interactions with snow, etc. For instance, forest canopy can enhance peak SWE levels and prolong the duration of snowpack<sup>29</sup>, consequently leading to an elevation in the occurrence of ROS events<sup>30</sup>. The *W* response to LULC is specific to each location and influenced by the local climate and vegetation conditions. Furthermore, at a particular location, the *W* response may vary from year to year based on the prevailing meteorological conditions. For water resources planning under nonstationarity, changes in LULC, such as deforestation, have the potential to increase peak SWE and subsequently summer water supply<sup>29,31,32</sup>. However, they can also contribute to more intense occurrences of flooding<sup>33,34</sup>.

To address this gap and build upon the work of Sun, *et al.*<sup>25</sup>, we have extended the NG-IDF datasets from open condition to include eight vegetated LULCs over the CONUS, namely deciduous forest, evergreen forest, mixed forest, shrub, grass, pasture, crop, and wetland. These LULCs were chosen in alignment with the National Land Cover Database (NLCD)<sup>35</sup> and National Resource Conservation Service (NRCS) Technical Release 55 (TR-55)<sup>36</sup>. The updated NG-IDF (NG-IDF 2.0) datasets cover more than 200,000 sites across the CONUS with a resolution of approximately 6 km. The NG-IDF 2.0 dataset spans the years 1951–2013 and includes a total of nine LULCs, including eight vegetated LULCs and the open condition. For each LULC, the datasets encompass daily time series data of *W* and SWE. They provide comprehensive information on hydrological extreme events and their associated hydrometeorological drivers. This information also proves useful in assessing the risk of infrastructure design and examining the impact of changes in LULC on the annual peak SWE, which is an indicator of potential summer water supply.

## Methods

**Water available for runoff modeling.** Figure 1 illustrates the methodology used to create NG-IDF 2.0 datasets for nine different LULCs across the CONUS. In this study, we adopt an approach where we assume a uniform LULC (i.e., 100% canopy fractional coverage) across the entire CONUS area, instead of representing the actual variation of LULC across the landscape. For each specific LULC, we separately estimate the time series of *W*. This approach is better suited for planning evaluations, by facilitating comparisons across locations. For instance, comparisons of *W* for the same locations with different LULCs will reveal the effects of LULC change on *W* (e.g., deforestations vs. afforestation), and cross-location comparisons will yield insights into the climate control on *W* under different LULC conditions. In this study, DHSVM is utilized to simulate the interaction between rainfall/snowfall and canopy at the point scale. The model incorporates a two-layer canopy model, an overstory canopy snow model, and a two-layer below-canopy energy and mass balance snowpack model. Comprehensive documentation of DHSVM can be found in numerous literature sources<sup>26–28,37</sup>, and as such, we only provide brief model descriptions here.

With the presence of a canopy, throughfall (TF) is generated when canopy interception storage exceeds the maximum interception storage capacity of overstory and understory, respectively. The model first calculates intercepted rainfall and snowfall by overstory if present. Maximum canopy interception of rain and snow is determined as a function of leaf area index (LAI). The water intercepted by the canopy is subject to evapotranspiration (ET). Potential evaporation is first calculated and represents the maximum rate at which water can be

removed from the canopy. Water intercepted by the overstory is removed from the wet fraction at the potential rate, while transpiration from the dry fraction is modeled using the Penman-Monteith approach. The understory ET is then calculated as the difference between the potential evaporation and calculated overstory ET (both wet and dry fractions). When a ground snowpack is present, excess snowfall and rainfall after interception combined with the mass release and drip from the overstory will contribute energy and mass to the ground snowpack. The snowpack model is a two-layer snow model with a thin surface layer and a lower pack layer. Energy balance at the snow surface layer is driven by net radiation, sensible and latent heat, and advected heat by rain. The surface layer receives attenuated shortwave radiation below the overstory and direct shortwave radiation in the open. Energy and mass exchange between the surface layer and the pack layer occurs only via the exchange of meltwater. Any liquid water remaining in the pack layer above its liquid water-holding capacity is released into the soil.

When simulating the open area condition without a canopy,  $W$  is estimated using mass balance as described by Sun, *et al.*<sup>25</sup>:

$$\text{Open area: } W = P - \Delta\text{SWE} + S \quad (1)$$

where  $P$  is precipitation,  $\Delta\text{SWE}$  is the change in ground SWE, and  $S$  indicates condensation (positive) or evaporation/sublimation (negative) of the snowpack. Taking into account the canopy impact on runoff,  $W$  under the canopy is estimated using:

$$\text{Under canopy: } W = \text{TF} - \Delta\text{SWE} + S \quad (2)$$

where  $\text{TF}$  is the throughfall after canopy rain/snow interception, subsequent evaporation/sublimation, and in the case of snow, melt and mass release through sloughing.  $\text{TF}$  explicitly quantifies the initial abstraction of vegetation, in contrast to  $P$  used for the open condition.

**DHSVM input and parameterization.** DHSVM is set up to model runoff processes at the point scale, covering the period from 1950 to 2013 with a 3-hour time resolution. The simulations are carried out on grid cells that correspond to the center of the Livneh 1/16° meteorological grid<sup>38</sup>. DHSVM's meteorological inputs consist of 3-hourly precipitation, air temperature, wind speed, relative humidity, as well as downward shortwave and longwave radiation. The 3-hourly meteorological forcing data were generated by disaggregating the daily Livneh meteorological data<sup>38</sup> using the Mountain Microclimate Simulation Model<sup>39</sup>. For ground snow processes, we used the same, validated regionalized snow parameters documented in Sun, *et al.*<sup>25</sup> for modeling ground snow accumulation and melt (Supplementary Table 1). Specifically, they used the  $k$ -means clustering technique based on the grid-level winter (November–March) precipitation, air temperature, and wind speed to classify the CONUS into five homogenous regions for snow parameterization: C1-Alpine, C2-Maritime, C3-Southern, C4-Northern, and C5-Interior (Fig. 1).

For a given LULC type, canopy parameters such as monthly LAI and height can be high variably across the CONUS, let alone account for dynamic vegetation and LULC changes due to human activities. To facilitate cross-location comparisons, we use the developed five clusters across the CONUS to represent the spatial variability of canopy parameters, and uniform canopy parameters are employed within each cluster. The cluster-specific canopy parameters represent the average canopy condition over the cluster, determined by averaging the LAI and canopy height measurements obtained from locations within the cluster. Further details about data sources are described in the following paragraph.

A series of datasets are utilized to determine canopy parameters in each cluster. First, the latest 2016 NLCD dataset<sup>35</sup> and the Landscape Fire and Resource Management Planning Tools Project (LANDFIRE) Existing Vegetation Height (EVH) database<sup>40,41</sup> are used in tandem to determine the canopy height for each of the eight vegetated LULCs at a 30-m resolution across the CONUS. Second, LAI values are obtained from combining field measurements and remotely sensed products, including the Moderate Resolution Imaging Spectroradiometer (MODIS) Version 6 LAI products MCD15A2H<sup>42</sup>, North American Carbon Program Terrestrial Ecosystem Research and Regional Analysis-Pacific Northwest (NACP TERRA-PNW) Forest Plant Traits<sup>43</sup>, and A Global Database of Field-observed Leaf Area Index in Woody Plant Species (LAI\_WOODY\_PLANTS\_1231)<sup>44</sup>. The NACP TERRA-PNW datasets provide LAI measurements for overstory trees in Oregon and Northern California, while the LAI\_WOODY\_PLANTS\_1231 datasets provide global measurements of overstory and understory LAI from 1,216 locations based on literature sources published between 1932 and 2011. Specifically, we use the 8-day, 500-m MODIS data to characterize the spatial variability and sub-seasonal changes of LAI; field data is chosen over MODIS to derive the maximum LAI values, because MODIS tends to underestimate LAI values at a local scale<sup>45,46</sup>. In line with the NLCD land cover classification, the forest types (deciduous, evergreen, mixed) and wetland land cover categories exhibit both overstory trees and understory vegetation, whereas the shrub, grass, pasture, and crop land cover categories exclusively feature understory vegetation. For example, Table 1 shows the cluster-average parameter values of maximum LAI and height for overstory and understory of deciduous forest, as well as the monthly LAI ratios to the maximum LAI values. Canopy parameter values for the other seven LULCs are presented in Supplementary Tables 2–8.

**NG-IDF curves.** We aggregated the 3-hourly  $W$  time series to create NG-IDF curves for selected durations ranging from 24 to 72 hours due to the absence of diurnal variability in the input precipitation data. For each duration, we determined the water year and calendar year annual maximum (AM)  $W$  and followed the NOAA Atlas 14<sup>19</sup> to fit a generalized extreme value (GEV) distribution to the 1951–2013 AM  $W$  datasets based on L-moments statistics<sup>47</sup>, excluding the first year to avoid initial condition uncertainty. We used the same GEV distribution to compare frequency estimates across durations and locations. We tested the stationarity assumption of the AM  $W$  data using the nonparametric Mann-Kendall test<sup>48,49</sup> and provided trends for statistically significant cases at

Canopy Parameter		Cluster				
		C1-Alpine	C2-Maritime	C3-Southern	C4-Northern	C5-Interior
Overstory Max LAI		4.5	5.3	4.9	5.2	5.5
Understory Max LAI		0.5	1.6	1.2	0.7	0.4
Overstory Height (m)		16.1	21.2	19.0	18.9	21.0
Understory Height (m)		0.5	0.9	0.7	0.5	0.4
Monthly LAI Ratio to Max LAI	January	0.10	0.27	0.19	0.22	0.21
	February	0.12	0.32	0.24	0.23	0.25
	March	0.19	0.41	0.33	0.32	0.34
	April	0.26	0.67	0.63	0.43	0.55
	May	0.57	0.99	0.96	0.68	0.80
	June	1.00	0.97	1.00	0.94	0.97
	July	0.91	1.00	0.97	1.00	1.00
	August	0.74	1.00	0.84	0.97	0.95
	September	0.55	0.82	0.70	0.79	0.84
	October	0.23	0.67	0.51	0.49	0.57
	November	0.18	0.35	0.27	0.27	0.28
	December	0.11	0.26	0.19	0.22	0.22

**Table 1.** Cluster canopy parameters of deciduous forest developed for the CONUS runs.

the 95% confidence level. For significant trends, we detrended the AM W time series using Sen's slope<sup>50</sup> while maintaining the time series average. In total, we created four NG-IDF curves for each location and LULC, using 1) water year AM W, 2) detrended water year AM W, 3) calendar year AM W, and 4) detrended calendar year AM W for average recurrence interval (ARIs) of 2, 5, 10, 25, 50, 100, and 500 years. A Monte Carlo (MC) simulation method<sup>19,47</sup> was used to consider sample data uncertainty in frequency analysis. After estimating the parameters of GEV distribution using the L-moments statistics, a total of 1,000 MC synthetic data sets were generated with the same record length. We then fitted GEV distribution to each MC synthetic data set using the L-moments statistics and estimated the associated values of the selected ARIs. Therefore, a total of 1,000 ensemble members were generated to quantify the uncertainties associated with NG-IDF curves. We provided the 90% confidence intervals using the 5% and 95% quantiles of the ensemble members. To provide a reference for comparison, we utilized the same methodology to create PREC-IDF curves with water year AM P and calendar year AM P. All trend and MC analyses were performed using the "trend"<sup>51</sup> and "lmom"<sup>52</sup> package in R, an open-source software environment.

**Driving mechanism and seasonality.** The driving mechanism of W was identified for each location, duration, and LULC. The classification of the driving mechanism was based on the P/TF and  $\Delta$ SWE, as per Sun, *et al.*<sup>25</sup>, and included three categories:

- 1) Rainfall only, which refers to precipitation or throughfall on snow-free ground;
- 2) Snowmelt only, which refers to decreasing SWE with no concurrent precipitation or throughfall; and
- 3) ROS, which refers to decreasing SWE with concurrent precipitation or throughfall. To further refine the focus on flood potential, a ROS event was defined as having at least 10 mm of precipitation or throughfall per day falling on a snowpack with at least 10 mm SWE over the selected duration. Additionally, the sum of rain and snowmelt had to contain at least 20% snowmelt<sup>13,25,53,54</sup>.

Precipitation was used for open area while throughfall was used for vegetated LULCs. In addition to the AM W time series, the AM Rain, AM Melt, and AM ROS time series were also provided for each location, and design events were constructed using the same approach as described above. For each location, the driving mechanism that produced the largest design event was identified as the dominant mechanism of hydrological extreme events. The degree of seasonality exhibited by AM W at different durations and LULCs was quantified using circular statistics to calculate the seasonality index (SI) and mean date (MD). While the mathematical details of circular statistics have been extensively covered in previous literature<sup>55–57</sup>, interested readers may refer to Sun, *et al.*<sup>25</sup> for more information. The SI is a value between 0 and 1, with higher values indicating a greater degree of seasonality, while the MD provides insight into the average timing of AM W events.

### Data Records

The NG-IDF 2.0 datasets covering the CONUS are publicly available in ASCII format through nine Zenodo repositories. These repositories host the datasets for various land cover types, including: (1) Evergreen forest<sup>58</sup>. (2) Deciduous forest<sup>59</sup>. (3) Mixed forest<sup>60</sup>. (4) Grassland<sup>61</sup>. (5) Crop<sup>62</sup>. (6) Open area<sup>63</sup>. (7) Pasture<sup>64</sup>. (8) Shrub<sup>65</sup>. (9) Wetland<sup>66</sup>. All repositories follow identical data structures, and Table 2 presents a summary of the data structures, data files, and variables specifically for the evergreen forest as an illustrative example.

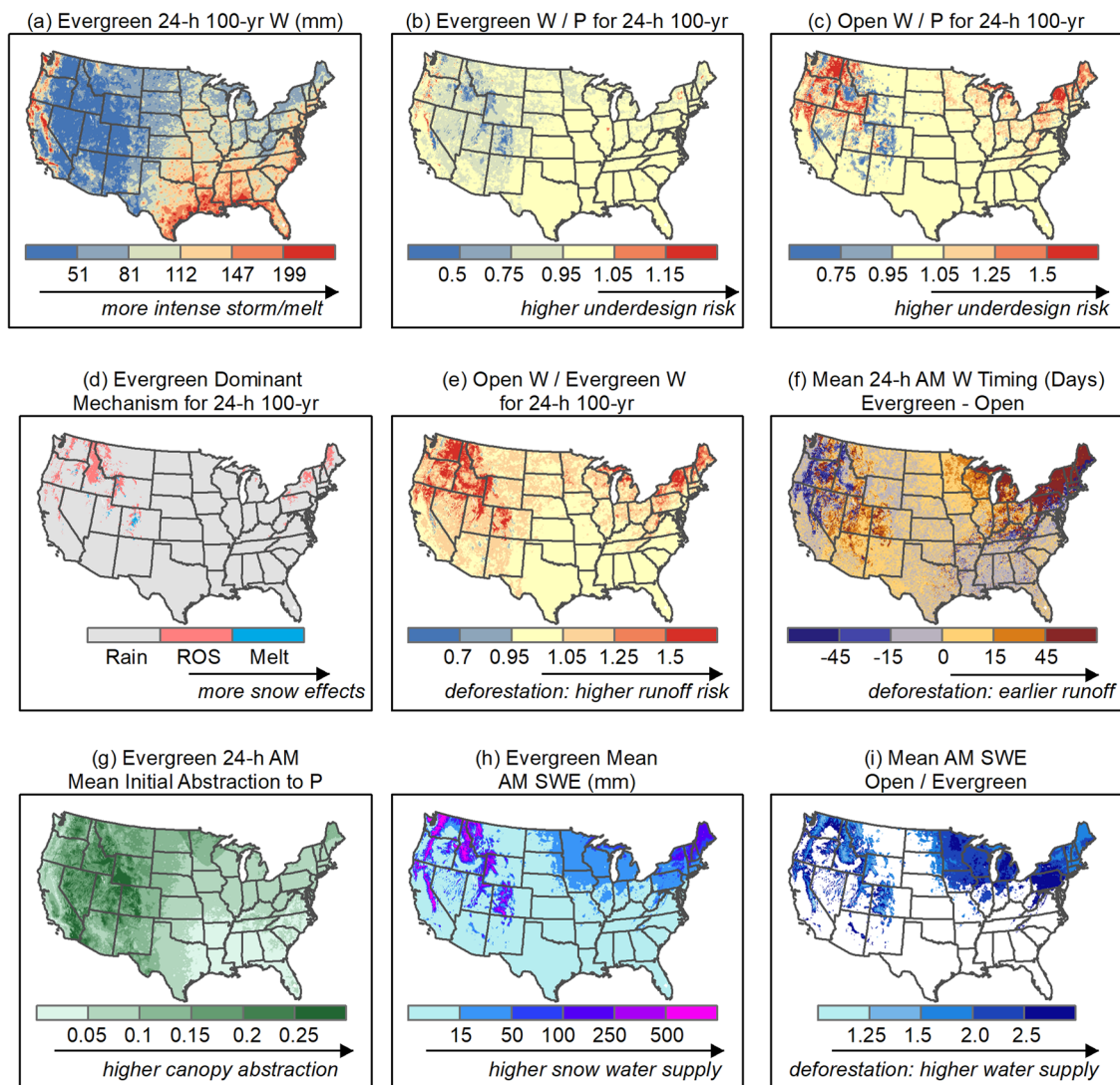
Main Folder	Naming Convention	Data File Description*
/24-h_time_series/	daily time series from 1950/1/1 to 2013/12/30	
	data_[lat]_[lon] e.g., data_25.15625_-80.71875	<b>Data Dimension:</b> 23,375 (R) × 4 (C). <b>C1:</b> W; <b>C2:</b> P; <b>C3:</b> TF; <b>C4:</b> SWE, all in mm.
/AM_time_series_[Year]/e.g., /AM_time_series_CY/	AM series with durations of 24 h, 48 h, and 72 h driven by different mechanisms from CY 1950–2013, including melt, rain, ROS, TF, and W. AM SWE series is also included.	
	[duration]_AM_time_series/[mechanism]/data_[lat]_[lon] e.g., 24-h_AM_time_series/W/data_25.15625_-80.71875	<b>Data Dimension:</b> 64 (R) × 4 (C) <b>C1:</b> Year; <b>C2:</b> Month; <b>C3:</b> Day; <b>C4:</b> AM value in mm.
/d_AM_time_series_[Year]/e.g., /d_AM_time_series_CY/	Detrended AM series with durations of 24 h, 48 h, and 72 h driven by different mechanisms from CY 1951–2013.	
	[duration]_AM_time_series/[mechanism]/data_[lat]_[lon] e.g., 24-h_AM_time_series/W/data_25.15625_-80.71875	<b>Data Dimension:</b> 63 (R) × 4 (C) <b>C1:</b> Year; <b>C2:</b> Month; <b>C3:</b> Day; <b>C4:</b> AM value in mm.
/IDF_curves_[Year]/e.g., /IDF_curves_CY/	IDF values and their 90% confidence intervals with durations of 24 h, 48 h, and 72 h driven by different mechanisms from 1951–2013, including melt, rain, ROS, TF, and W. Dominant mechanism and mean W and SWE dates from seasonality analysis are also included.	
	[duration]_[mechanism] e.g., 24-h_W	<b>Data Dimension:</b> 207,173 (R) × 9 (C) <b>C1:</b> Latitude; <b>C2:</b> Longitude; <b>C3–C9:</b> IDF values with ARIs of 2, 5, 10, 25, 50, 100, and 500 years, in mm.
/IDF_curves_[Year]_detrend/e.g., /IDF_curves_CY_detrend/	IDF values and their 90% confidence intervals using detrended AM data with durations of 24 h, 48 h, and 72 h driven by different mechanisms from 1951–2013.	
	[duration]_[mechanism] e.g., 24-h_W	<b>Data Dimension:</b> 207,173 (R) × 9 (C) <b>C1:</b> Latitude; <b>C2:</b> Longitude; <b>C3–C9:</b> IDF values with ARIs of 2, 5, 10, 25, 50, 100, and 500 years, in mm.
/trend_results_[year] e.g., /trend_results_CY/	Sen's slope of Mann-Kendall trend in AM series with different mechanisms in 1951–2013.	
	[duration]_[mechanism] e.g., 24-h_W	<b>Data Dimension:</b> 207,173 (R) × 3 (C) <b>C1:</b> Latitude; <b>C2:</b> Longitude; <b>C3:</b> Sen's slope in mm/yr.

**Table 2.** Description of the NG-IDF 2.0 datasets for evergreen forest. \*Note: In “Data Description”, C = column, R = Row. C[i] indicates the *i*th column of a data file.

## Technical Validation

Due to the absence of data for direct NG-IDF curve evaluation, the evaluation focuses on *W*, which serves as the source data for deriving NG-IDF curves. Our specific focus was on evaluating the model's simulated daily SWE, which is a key variable used in the *W* calculation besides precipitation data from the climate dataset. Currently, the NRCS SNOTEL network provides daily SWE measurements under open condition at approximately 800 stations in the western United States. To ensure quality, SNOTEL data was subject to a rigorous three-stage quality control filter<sup>20</sup> and is subsequently corrected for snowfall undercatch<sup>27</sup>. The resulting data set, called bias-corrected quality-controlled (BCQC) SNOTEL data, can be accessed at <https://climate.pnnl.gov/?category=Hydrology>. Sun, *et al.*<sup>25</sup> provided a detailed description of the comprehensive validation of DHSVM SWE simulation against SNOTEL data, and only a brief overview is presented here. Briefly, they selected 246 SNOTEL stations that shared the longest common period (2007–2013) of BCQC daily SWE records and evaluated the SWE simulation skill using three metrics: Nash-Sutcliffe Efficiency (NSE), bias in mean annual peak SWE (PEAK.ERR), and bias in the timing of peak SWE (PDATE.ERR). The results indicated that NSE of daily SWE was greater than 0.6 at 75% of the stations, absolute PEAK.ERR was less than or equal to 25% at 67% of the stations, and absolute PDATE.ERR was less than or equal to 14 days at 67% of the stations. These findings suggest that the calibrated DHSVM is capable of replicating the observed SWE dynamics at most of the stations using the regionalized snow parameters and can support large-domain hydrological applications. For more information, readers are directed to Sun, *et al.*<sup>25</sup>.

However, the data availability is rather limited for under-canopy SWE and most are short-term, discontinuous, point-scale measurements. Previous studies<sup>27,28,67–71</sup> have extensively validated the ability of DHSVM to simulate snow and streamflow in various vegetated watersheds across the CONUS. For instance, in an extensive evaluation of 30 hydrological models, Beckers, *et al.*<sup>69</sup> determined that DHSVM was the most suitable for hydrological modeling in forested environments. Du, *et al.*<sup>67</sup> demonstrated that DHSVM effectively replicates the dynamics of snowpack, soil water content, and streamflow patterns in the forested Mica Creek Experimental Watershed in northern Idaho. Cristea, *et al.*<sup>70</sup> confirmed that DHSVM accurately reproduces the dynamics of snow and streamflow in the forested Tuolumne basin of the Sierra Nevada, California. Sun, *et al.*<sup>28</sup> further improved the DHSVM canopy model by incorporating canopy gap structure and verified its high accuracy in



**Fig. 2** Example uses of NG-IDF datasets for nine LULCs, illustrated with calendar year AM data.

simulating snow under various canopy conditions, ranging from open to dense forest to canopy gaps, using field data from the University of Idaho Experimental Forest near Moscow, Idaho.

### Usage Notes

The NG-IDF 2.0 datasets provided in Table 2 are readily applicable for diverse hydrological applications across the CONUS without requiring further modifications. We present five data usage applications here as shown in Fig. 2, but our choices are not exhaustive.

1. *Assess design risk with the use of standard PREC-IDF curves.* To conduct hydrological design and analyses, users can obtain information on the magnitude of extreme W events and their associated P events for each LULC at any desired location (Fig. 2a,b,c).
2. *Plan various scenarios of LULC change and assess their effects on water supply and the risk of flooding.* Effective management of water resources requires careful planning for changes in LULC, which takes into account potential impact on both water supply and the risk of flooding. Here the NG-IDF 2.0 datasets that provide insights into the potential effects of LULC change on both snow water supply and flood risk are essential for effective LULC planning and sustainable water management (Fig. 2e,h,i).
3. *Offer physical insights into changes in runoff timing and mechanisms resulting from modifications in LULC.* The NG-IDF datasets offer runoff mechanisms (such as rain, snowmelt, and ROS) for each W event and seasonality for every LULC, allowing users to not only quantify changes in extreme W events but also comprehend the reasons for these changes (Fig. 2d,f). Moreover, the classification of the dominant runoff mechanism enables the development of mixed populations in the frequency of W events<sup>72,73</sup>.
4. *Enhance the standard hydrological design method by quantifying the spatial heterogeneity in LULC initial abstractions.* The TR-55, which is a commonly used IDF design method, employs a fixed ratio across the

CONUS to represent canopy initial abstraction regardless of LULC variations. The NG-IDF 2.0 datasets can systematically investigate the canopy initial abstraction ratio for nine LULCs across the CONUS, providing an opportunity to enhance runoff prediction accuracy when using the standard IDF design method (Fig. 2g).

5. *Provide spatial runoff data for to support downstream modeling applications, such as flood inundation modeling.* In addition to integrating with rainfall-runoff models like TR-55 for assessing flood risk in hydrological design, the 6 km W datasets can be employed as inputs for hydrodynamic models like Rapid Infrastructure Flood Tool (RIFT)<sup>74</sup> to enhance the accuracy of flood depth predictions (Fig. 2a), particularly for events triggered by snowmelt or ROS flooding, thus improving the existing Federal Emergency Management Agency (FEMA) Special Flood Hazard Area (SFHA) maps<sup>75</sup>.

Finally, it is important to acknowledge that the datasets are derived from a singular canopy condition, representing the average of a cluster. If the local canopy features vary significantly from this cluster average, the resulting information will differ. Moving forward, our intention is to create a cloud NG-IDF computing tool that allows users to input their specific local canopy attributes, enabling the generation of NG-IDF curves for any canopy condition related to each LULC.

### Code availability

The DHSVM source code is available at <https://github.com/pnnl/DHSVM-PNNL> Source codes that are used to develop and analyze the data are available at <https://github.com/hydro-yan/NG-IDF> The BCQC SNOTEL data are available at <https://climate.pnnl.gov/?category=Hydrology>.

Received: 16 June 2023; Accepted: 24 October 2023;

Published online: 04 December 2023

### References

1. AghaKouchak, A., Feldman, D., Hoerling, M., Huxman, T. & Lund, J. Water and climate: Recognize anthropogenic drought. *Nature* **524**, 409–411 (2015).
2. Zscheischler, J. *et al.* A typology of compound weather and climate events. *Nat. Rev. Earth Environ.* **1**, 333–347 (2020).
3. Yan, H., Moradkhani, H. & Zarekarizi, M. A probabilistic drought forecasting framework: A combined dynamical and statistical approach. *J. Hydrol.* **548**, 291–304 (2017).
4. Yan, H. & Moradkhani, H. Toward more robust extreme flood prediction by Bayesian hierarchical and multimodeling. *Nat. Hazards* **81**, 203–225 (2016).
5. Bales, R. C. *et al.* Mountain hydrology of the western United States. *Water Resour. Res.* **42**, W08432 (2006).
6. Li, D., Wrzesien, M. L., Durand, M., Adam, J. & Lettenmaier, D. P. How much runoff originates as snow in the western United States, and how will that change in the future? *Geophys. Res. Lett.* **44**, 6163–6172 (2017).
7. Yan, H. *et al.* Observed Spatiotemporal Changes in the Mechanisms of Extreme Water Available for Runoff in the Western United States. *Geophys. Res. Lett.* **46**, 767–775 (2019).
8. Marlier, M. E. *et al.* The 2015 drought in Washington State: a harbinger of things to come? *Environ. Res. Lett.* **12**, 114008 (2017).
9. Harpold, A., Dettinger, M. & Rajagopal, S. Defining Snow Drought and Why It Matters. *Eos (Washington, DC)*. <https://doi.org/10.1029/2017EO068775> (2017).
10. Yan, H., Sun, N., Fullerton, A. & Baerwalde, M. Greater vulnerability of snowmelt-fed river thermal regimes to a warming climate. *Environ. Res. Lett.* **16**, 054006 (2021).
11. Zarekarizi, M., Yan, H., Ahmadalipour, A. & Moradkhani, H. A Probabilistic Framework for Agricultural Drought Forecasting Using the Ensemble Data Assimilation and Bayesian Multivariate Modeling. in *Global Drought and Flood: Observation, Modeling, and Prediction* 147–164, <https://doi.org/10.1002/9781119427339.ch8> (2021).
12. McCabe, G. J., Hay, L. E. & Clark, M. P. Rain-on-Snow Events in the Western United States. *Bull. Am. Meteorol. Soc.* **88**, 319–328 (2007).
13. Musselman, K. N. *et al.* Projected increases and shifts in rain-on-snow flood risk over western North America. *Nat. Clim. Chang.* **8**, 808–812 (2018).
14. Vahedifard, F., AghaKouchak, A., Ragno, E., Shahrokhbabadi, S. & Mallakpour, I. Lessons from the Oroville dam. *Science (80-)*. **355**, 1139–1140 (2017).
15. Yale Environment 360. Post-Flood Yellowstone Rebuilding Could Cost More Than \$1 Billion. <https://e360.yale.edu/digest/yellowstone-rebuild-could-cost-more-than-1-billion> (2022).
16. Yan, H., Sun, N., Chen, X. & Wigmosta, M. S. Next-Generation Intensity-Duration-Frequency Curves for Climate-Resilient Infrastructure Design: Advances and Opportunities. *Front. Water* **2**, 545051 (2020).
17. Yan, H. *et al.* Evaluating next-generation intensity-duration-frequency curves for design flood estimates in the snow-dominated western United States. *Hydrol. Process.* **34**, 1255–1268 (2020).
18. Hamlet, A. F. New Observed Data Sets for the Validation of Hydrology and Land Surface Models in Cold Climates. *Water Resour. Res.* **54**, 5190–5197 (2018).
19. Perica, S. *et al.* *Precipitation-Frequency Atlas of the United States*, NOAA Atlas 14. (2013).
20. Yan, H. *et al.* Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design in Snow-Dominated Environments. *Water Resour. Res.* **54**, 1093–1108 (2018).
21. Cho, E. & Jacobs, J. M. Extreme Value Snow Water Equivalent and Snowmelt for Infrastructure Design Over the Contiguous United States. *Water Resour. Res.* **56** (2020).
22. Yan, H. *et al.* The Role of Snowmelt Temporal Pattern in Flood Estimation for a Small Snow-Dominated Basin in the Sierra Nevada. *Water Resources Research*, **59**(10), e2023WR034496 (2023).
23. Rose, M. Yellowstone National Park was never built to take on the rain and snow that comes with climate change. *Popular Science* <https://www.popsoci.com/environment/yellowstone-extreme-flood-park-infrastructure-climate-change/> (2022).
24. Yan, H. *et al.* Next-Generation Intensity-Duration-Frequency Curves to Reduce Errors in Peak Flood Design. *J. Hydrol. Eng.* **24**, 04019020 (2019).
25. Sun, N. *et al.* Datasets for characterizing extreme events relevant to hydrologic design over the conterminous United States. *Sci. Data* **9**, 154 (2022).
26. Wigmosta, M. S., Vail, L. W. & Lettenmaier, D. P. A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.* **30**, 1665–1679 (1994).

27. Sun, N. *et al.* Regional Snow Parameters Estimation for Large-Domain Hydrological Applications in the Western United States. *J. Geophys. Res. Atmos.* **124**, 5296–5313 (2019).
28. Sun, N. *et al.* 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 (2018).
29. Sun, N. *et al.* Forest Canopy Density Effects on Snowpack Across the Climate Gradients of the Western United States Mountain Ranges. *Water Resour. Res.* **58**, e2020WR029194 (2022).
30. Mooney, P. A. & Lee, H. Afforestation affects Rain-On-Snow climatology over Norway. *Environ. Res. Lett.* <https://doi.org/10.1088/1748-9326/ac6684> (2022).
31. Lundquist, J. D., Dickerson-Lange, S. E., Lutz, J. A. & Cristea, N. C. Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling. *Water Resour. Res.* **49**, 6356–6370 (2013).
32. Currier, W. R., Sun, N., Wigmosta, M., Cristea, N. & Lundquist, J. D. The impact of forest-controlled snow variability on late-season streamflow varies by climatic region and forest structure. *Hydrol. Process.* **36**, e14614 (2022).
33. Yan, H. & Edwards, F. G. Effects of Land Use Change on Hydrologic Response at a Watershed Scale, Arkansas. *J. Hydrol. Eng.* **18**, 1779–1785 (2013).
34. Zhang, W., Villarini, G., Vecchi, G. A. & Smith, J. A. Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature* **563**, 384–388 (2018).
35. Yang, L. *et al.* A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens.* **146**, 108–123 (2018).
36. Cronshey, R. *et al.* *Urban Hydrology for Small Watersheds—TR-55.* (1986).
37. Wigmosta, M. S., Nijssen, B. & Storck, P. The distributed hydrology soil vegetation model. in *Mathematical Models of Small Watershed Hydrology and Applications* (ed. Singh, V. P.) 7–42 (Water Resources Publication, 2002).
38. Livneh, B. *et al.* 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 (2013).
39. Hungerford, R. D., Nemani, R. R., Running, S. W. & Coughlan, J. C. *MTCLIM: A Mountain Microclimate Simulation Model.* (U.S. Department of Agriculture, 1989).
40. Rollins, M. G. LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. *Int. J. Wildl. Fire* **18**, 235 (2009).
41. Ryan, K. C. & Opperman, T. S. LANDFIRE – A national vegetation/fuels data base for use in fuels treatment, restoration, and suppression planning. *For. Ecol. Manage.* **294**, 208–216 (2013).
42. Myneni, R., Knyazikhin, Y. & Park, T. MCD15A2H MODIS/Terra+ Aqua Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006. in (2015).
43. Law, B. E. & Berner, L. T. NACP TERRA-PNW: Forest Plant Traits, NPP, Biomass, and Soil Properties, 1999–2014. <https://doi.org/10.3334/ORNLDAAC/1292> (2015).
44. Iio, A. & Ito, A. A Global Database of Field-observed Leaf Area Index in Woody Plant Species, 1932–2011. <https://doi.org/10.3334/ORNLDAAC/1231> (2014).
45. Law, B. E., Van Tuyl, S., Cescatti, A. & Baldocchi, D. D. Estimation of leaf area index in open-canopy ponderosa pine forests at different successional stages and management regimes in Oregon. *Agric. For. Meteorol.* **108**, 1–14 (2001).
46. Bolstad, P. V. & Gower, S. T. Estimation of leaf area index in fourteen southern Wisconsin forest stands using a portable radiometer. *Tree Physiol.* **7**, 115–124 (1990).
47. Hosking, J. R. M. & Wallis, J. R. *Regional Frequency Analysis: An Approach Based on L-Moments.* (Cambridge University Press, Cambridge, U. K., 1997).
48. Mann, H. B. Nonparametric Tests Against Trend. *Econometrica* **13**, 245 (1945).
49. Kendall, M. G. *Rank Correlation Methods.* (1975).
50. Sen, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. *J. Am. Stat. Assoc.* **63**, 1379–1389 (1968).
51. Pohlert, T. Package 'trend'. <https://CRAN.R-project.org/package=trend>, (2016).
52. Hosking, J. R. M. Package 'lmom'. <https://CRAN.R-project.org/package=lmom>, (2017).
53. Li, D., Lettenmaier, D. P., Margulis, S. A. & Andreadis, K. The Role of Rain-on-Snow in Flooding Over the Conterminous United States. *Water Resour. Res.* **55**, 8492–8513 (2019).
54. Freudiger, D., Kohn, I., Stahl, K. & Weiler, M. Large-scale analysis of changing frequencies of rain-on-snow events with flood-generation potential. *Hydrol. Earth Syst. Sci.* **18**, 2695–2709 (2014).
55. Burn, D. H. Catchment similarity for regional flood frequency analysis using seasonality measures. *J. Hydrol.* **202**, 212–230 (1997).
56. Villarini, G. On the seasonality of flooding across the continental United States. *Adv. Water Resour.* **87**, 80–91 (2016).
57. Berghuijs, W. R., Woods, R. A., Hutton, C. J. & Sivapalan, M. Dominant flood generating mechanisms across the United States. *Geophys. Res. Lett.* **43**, 4382–4390 (2016).
58. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Evergreen Forest. *Zenodo* <https://doi.org/10.5281/zenodo.7976419> (2023).
59. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Deciduous Forest. *Zenodo* <https://doi.org/10.5281/zenodo.7972064> (2023).
60. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Mixed Forest. *Zenodo* <https://doi.org/10.5281/zenodo.7980090> (2023).
61. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Grassland. *Zenodo* <https://doi.org/10.5281/zenodo.7978021> (2023).
62. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Crop. *Zenodo* <https://doi.org/10.5281/zenodo.7963584> (2023).
63. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Open Area. *Zenodo* <https://doi.org/10.5281/zenodo.7982399> (2023).
64. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Pasture. *Zenodo* <https://doi.org/10.5281/zenodo.7983633> (2023).
65. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Shrub. *Zenodo* <https://doi.org/10.5281/zenodo.7992039> (2023).
66. Yan, H., Wigmosta, M. S., Duan, Z. & Sun, N. CONUS NG-IDF 2.0: Wetland. *Zenodo* <https://doi.org/10.5281/zenodo.7995995> (2023).
67. Du, E., Link, T. E., Gravelle, J. A. & Hubbart, J. A. Validation and sensitivity test of the distributed hydrology soil-vegetation model (DHSV) in a forested mountain watershed. *Hydrol. Process.* **28**, 6196–6210 (2014).
68. Perkins, W. A. *et al.* Parallel Distributed Hydrology Soil Vegetation Model (DHSV) using global arrays. *Environ. Model. Softw.* **122**, 104533 (2019).
69. Beckers, J., Smerdon, B. & Wilson, M. *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, 2009).
70. Cristea, N. C., Lundquist, J. D., Loheide, S. P., Lowry, C. S. & Moore, C. E. Modelling how vegetation cover affects climate change impacts on streamflow timing and magnitude in the snowmelt-dominated upper Tuolumne Basin, Sierra Nevada. *Hydrol. Process.* **28**, 3896–3918 (2014).
71. Storck, P. *Trees, snow, and flooding: An investigation of forest canopy effects on snow accumulation and melt at the plot and watershed scales in the Pacific Northwest.* (University of Washington, Seattle, 2000).



72. Barth, N. A., Villarini, G. & White, K. Accounting for Mixed Populations in Flood Frequency Analysis: Bulletin 17C Perspective. *J. Hydrol. Eng.* **24** (2019).
73. Yu, G., Wright, D. B. & Davenport, F. V. Diverse Physical Processes Drive Upper-Tail Flood Quantiles in the US Mountain West. *Geophys. Res. Lett.* **49** (2022).
74. Judi, D., Rakowski, C., Waichler, S., Feng, Y. & Wigmosta, M. Integrated Modeling Approach for the Development of Climate-Informed, Actionable Information. *Water* **10**, 775 (2018).
75. Federal Emergency Management Agency. *NFIP Flood Insurance Manual*. (2016).

### Acknowledgements

This material is based upon work supported by the Environmental Security Technology Certification Program under Contract No. EW21-5140. Battelle Memorial Institute operates the Pacific Northwest National Laboratory (PNNL) for the U.S. Department of Energy under contract DE-AC06-76RLO-1830. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award BER-ERCAP0023966.

### Author contributions

H.Y., Z.D. and M.S.W. designed the general approach of developing the datasets. Z.D. performed model simulations. H.Y. analyzed data and drafted the manuscript. N.S., E.D.G., B.K. and J.R.A. contributed to the final version of the manuscript. M.S.W. and H.Y. advised and managed the projects that provided funding for this research. All authors participated in discussions and reviews during the development of this manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02680-4>.

**Correspondence** and requests for materials should be addressed to H.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Battelle Memorial Institute, NCAR, MITRE Corp. 2023