A watershed-scale multi-approach assessment of design flood discharge estimates used in hydrologic risk analyses for forest road stream crossings and culverts

Sourav Mukherjee, Devendra M. Amatya, John L. Campbell, Landon Gryczkowski, Sudhanshu Panda, Sherri L. Johnson, Kelly Elder, Anna M. Jalowska, Peter Caldwell, Johnny M. Grace, Dariusz Mlynski, Andrzej Wałga

A R T I C L E  I N F O

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief

Keywords:
Flood risk
Nonstationary regional frequency analysis
Hydrologic vulnerability of culvert
Forest service experimental forests
Headwater catchments
Road-stream crossing structures

A B S T R A C T

Flood peak magnitudes and frequency estimates are key components of any effective nationwide flood risk management and flood damage abatement program. In this study, we evaluated normalized peak design discharges \( Q_{90} \) for 1,387 hydrologic unit code 16 to 20 (HUC16-20) watersheds in the White Mountain National Forest (WMNF), New Hampshire and in five Experimental Forest (EF) regions across the United States managed by USDA Forest Service (USDA-FS). Nonstationary regional frequency analysis (RFA) and single site frequency analysis (FA) with long-term high-resolution observed streamflow data along with the deterministic Rational Method (RM) and semi-empirical United States Geological Survey regional regression equation (USGS-RRE) were used. Additionally, a hydrologic vulnerability assessment was performed for 194 road culverts as a result of extreme precipitation-induced flooding on gauged and ungauged watersheds in the Hubbard Brook EF (HBR) within the WMNF. The RM outperformed the USGS-RRE in predicting \( Q_{90} \) in the gauged and ungauged HUC16-20 watersheds of WMNF and in three other small, high-relief forest headwater watersheds—Coweeta Hydrologic Lab EF’s watershed-14, and watershed-27 in North Carolina and HJ Andrews EF’s watershed 8 in Oregon. However, the USGS-RRE performed better for larger watersheds, such as the Fraser EF’s St. Louis watershed in Colorado and the Santee EF’s watershed 80 in South Carolina. About 31 %, 26 %, and 56 % of the culverts at the HBR site could not accommodate the 100-yr \( Q_{90} \) estimated by RFA, RM, and USGS-RRE, respectively. Based on the chosen RIs and techniques, it is determined that except for one culvert with diameter \( \geq 0.91 \text{ m} \) (36 in.), none of the culverts with diameter of \( 0.75 \text{ m} \) (30 in.) or larger are hydrologically vulnerable. Our results suggest that the observation based RFA works best where multiple gauges are available to extrapolate information for ungauged watersheds, otherwise, RM is best-suited for smaller headwater watersheds and USGS-RRE for larger watersheds. Results from the hydrologic vulnerability analysis revealed that replacing undersized culverts with new culverts of diameter \( \geq 0.75 \text{ m} \) will improve flood resiliency, provided that the structure is geomorphologically safe (with minimal effects of debris flow, erosion, and sedimentation) and allows for both bank-full discharge and necessary fish passage within that design limit. This study has implications in managing road culverts and crossings at Forest Service and other forested lands for their resiliency to extreme precipitation and flooding hazards induced by climate change.

* Corresponding author.
E-mail address: sourav.mukherjee@usda.gov (S. Mukherjee).

https://doi.org/10.1016/j.jhydrol.2024.130698
Received 5 September 2023; Received in revised form 14 December 2023; Accepted 17 December 2023
Available online 29 January 2024
0022-1694/Published by Elsevier B.V.
1. Introduction

Extreme precipitation and floods emerge as the foremost threat to road culvert systems, surpassing other climate change-related hazards due to their capacity to cause damage and disrupt transportation networks (Pregnolato et al., 2017; USDOT, 2018; Amatya et al., 2021a–c; Darestani et al., 2021). Floods can significantly alter hydrological processes, and cause soil erosion and landslides, having devastating impacts on forest road infrastructure and ecosystems (Mishra et al., 2022; Panda et al., 2022; Goodman et al., 2023). The flooding events that occur after extreme precipitation events has been reported in regions all over the globe (Papeziou and Montanari, 2019). Over the last few decades, intensity and the frequency of rainfall events leading to floods have increased, which has affected Experimental Forest (EF) regions and National Forest lands managed by the US Department of Agriculture Forest Service (USDAFS) (Amatya et al., 2016; Glasser, 2005).

For example, the September 2013 flood that occurred in the Roosevelt and Arapaho National Forests in Colorado affected numerous roads and trails, destroyed bridges and culverts, and claimed several lives (Gochis et al., 2015). The October 2015 flooding caused by unprecedented rainfall in the Francis Marion National Forest in South Carolina damaged culverts and roads, and inundated gauging stations (Amatya et al., 2016; Marciano and Lackmann, 2017), and altered the ecology of the forest (Van Dam et al., 2018; Mishra et al., 2021). The flooding breached dams and roads causing substantial erosion and sedimentation that may have long-term effects on water quality and aquatic habitats (Magilligan et al., 2019). Tropical storm Irene in August 2011, resulted in heavy rain and flash flooding in the Hubbard Brook National Forest in New Hampshire, which damaged culverts and bridges and made some roads impassable, blocking access to some areas for weeks to months (Olson and Bent, 2013; Yellen et al., 2016). The H.J. Andrews Experimental Forest in the Cascade Range of western Oregon has also experienced periodic flooding, including a major flood in 1996 (Chang et al., 2010; Johnson et al., 2000). The increased likelihood of extreme precipitation events under a changing climate is projected to cause substantial increases in the magnitude and frequency of floods, including those caused by hurricanes and tropical storms (Berghuijs et al., 2016).

The profound effect of flooding on forest ecosystems and infrastructure has highlighted the need for greater resilience for forest land to extreme precipitation events (Keller and Sharar, 2003; Keller and Ketcheson, 2011). Road stream crossing structures (RSCS), such as road lead off structures, fords, culverts, and bridges play a crucial role in the national forests’ road network (Keller and Sharar, 2003). This system is essential for sustainable management, which involves, balancing the demands of resource management and traffic with the preservation of aquatic habitats and ecosystem health (Seddon et al., 2021). The hydrologic risk from flooding is amplified with inadequately sized RSCS, and together with soil erosion and flooding, can cause structural failures that leads to financial losses and disturb stream connectivity essential for aquatic organism passage (Heredia et al., 2016). Many RSCS, specifically culverts on the forest lands, are currently deemed undersized for accommodating bank-full flow conditions for 1- to 2-year flooding, which is one of the critical design considerations required for flood resiliency (USDOT, 2018; Amatya et al., 2021a–c).

Deterministic and empirical methods, such as, the Rational Method (RM) (Kuichling, 1889) and Natural Resource Conservation Service (NRCS) Graphical Peak Discharge method (GPDVM) (USDAs, 2021) have been widely used for estimating flood peak discharge, although, in small to medium-sized watersheds or rural areas where hydrologic data may be limited (Fleming and Franz, 1971; Walder and O’Connor, 1997; Megnounif et al., 2013; Amatya et al., 2021a–c). However, it is important to note that deterministic methods have limitations and are not appropriate in all situations, especially complex hydrologic terrain (Grimaldi and Petroselli, 2015a,b; Grimaldi et al., 2021). Their accuracy depends on the quality and quantity of data used to develop the method and requires an accurate calibration of parameters (Montanari and Koutsoyiannis, 2014), which is expensive and time-consuming task (Ke et al., 2020). Amatya and Walega (2020) suggest validating these methods, their modified versions, and their parameters with long-term hydroclimatic data from a range of watersheds.

Estimating flood peak discharge with USGS Regional Regression Equations (USGS-RREs) is a semi-empirical approach that accounts for various factors that influence flooding, such as the size of the drainage area, topography, the type of soil, and precipitation magnitude (e.g., average 24-hr maximum precipitation) in some cases (Vogel et al., 1999; Capesius and Stephens, 2005). As such, simple linear and multivariate regressions of the record peak discharges as a function of drainage area (and additional regional predictors) are widely used for estimating flood frequency as well as predicting, ranking, and defining expected flood magnitudes for medium to large watersheds (Yochum et al., 2019; Levin and Sanocki, 2023). Although this method has proved useful, there are several limitations including: 1) poor performance when there is limited data available for developing the regression equations, 2) limited applicability to regions with complex hydrological conditions, 3) lack of consideration of factors such as land use change, antecedent moisture conditions and climate variability to predict flood peaks, and 4) high uncertainty associated with predicting flood peaks outside the range of available data, and associated assumptions, such as stationarity of the data used to develop the regression equations (Mitchell et al., 2023; Amatya et al., 2021a–c). Various studies (Amatya et al., 2021a–c; Marion, 2004; Thompson, 2006; Trommer et al., 1996) have shown that the USGS-RRE has a tendency to overestimate design discharge for small undisturbed watersheds. Additionally, extrapolating to smaller watersheds outside the boundaries of the regression data can be problematic (Genereux, 2003).

In comparison to deterministic and semi-empirical methods like the RM and the USGS-RREs, probability-based flood frequency analysis (Vogel and Castellarin, 2017; Read and Vogel, 2015) is usually considered a more rigorous and robust technique for flood peak estimation (Mishra et al., 2022). This is particularly true for complex watersheds where the assumptions of hydrological models may not hold (Thompson, 2004). In addition, probability-based techniques involve statistical analysis of historically observed data and unlike the other methods, take into consideration the variability of flood events over time (Pegram and Parak, 2004). One of the major benefits of probability-based flood frequency analysis is that it can be used to analyze extreme events that might not have happened during the observed record of flood data, particularly in intermittent streams (Mishra et al., 2022). Based on the probability distribution, the likelihood of extreme events happening in the future can be estimated, which is critical in designing climate resilient infrastructure to manage flood risk. Furthermore, probability-based regional frequency analysis (RFA) is more reliable in transferring flood information from gauged to ungauged watersheds (Guillemin, 2010). RFA considers the regional variations in hydrologic characteristics that effect flood frequency estimates, and is, therefore, a more sophisticated technique for flood frequency analysis compared to flood frequency analysis considering a single site (hereafter referred to as FA) (Halbert et al., 2016). More specifically, a RFA, that incorporates non-stationary covariates such as hydro-geomorphic (Rootzen and Katz, 2013; Wasko et al., 2020; De Michele and Salvadori, 2002; Mishra et al., 2022) and climatic conditions (Mishra et al., 2022) is more holistic, and can be extended to predict flood peaks in similar but ungauged watersheds (Kim et al., 2020). Though these factors play an important role in modeling, they are currently under-represented in existing deterministic and semi-empirical models reported in the literature (Teng et al., 2017).

Precipitation-intensity–duration–frequency (PDIF) estimates of storm events can also be helpful for determining how frequently floods may harm RSCS causing traffic disruption and erosion in road networks and headwater valleys (Jakob et al., 2020; 2022; Mamo, 2015). The majority of RSCS on USDA-FS land are in small drainage areas (<1000 ha or 10 km²) with forested headwater watersheds, typically with sub-
daily time of concentrations (Tc) (USDA, 2021; Amatya et al., 2021a–c; Corbin et al., 2021; Yochum et al., 2019; Walega et al., 2020). Therefore, it is most likely that flood potential in small watersheds will be underestimated if design flow rates are calculated using PIDF estimates at a daily time scale with limited period-length integrated to an event-based approach (Viglione and Blöschl, 2009; Wright et al., 2019; Mukherjee et al., 2023; Amatya et al., 2021a–c). Synthetic rainfall generation models coupled with continuous rainfall-runoff transformation in data scarce conditions can provide a more accurate representation of rainfall events in hydrological models (Beven, 2011; Pathiraja et al., 2012; Winter et al., 2019; Beneyto et al., 2020; Grimaldi et al., 2022). However, its use remains limited due to the challenges in selecting the optimal rainfall simulation model for specific risk analysis needs (Grimaldi et al., 2022). As such, due to a lack of long-term historical precipitation records at a fine temporal resolution (sub-hourly and sub-daily timescales), the RSCL on USDA-FS lands and other similar landscapes are typically designed using coarser, daily resolution or, if available, the sub-hourly and sub-daily PIDFs provided by the National Oceanic and Atmospheric Administration (NOAA) Atlas 14 which has multiple caveats (Mukherjee et al., 2023; Perica et al., 2018). Multiple studies have shown that as an alternative to the NOAA data, the USDA-FS EF’s long-term, sub-hourly to sub-daily PIDFs that are available at some sites, can be used to determine reliable estimates of design discharge for the small EF watersheds (Amatya and Walega, 2020; Amatya et al., 2021a–c; Mukherjee et al., 2023).

In this study, we used USDA-FS EF’s long-term, high-resolution precipitation intensities and instantaneous streamflow records to determine the hydrologically vulnerable RSCLs to overtopping on national forest and multiple EF lands that represent a range of climatic conditions. We applied several approaches and methods including, probability-based stationary and nonstationary FA and RFA that requires gauged streamflow data, and deterministic and semi-empirical methods like RM, and USGS-RRE, respectively, for estimating the normalized peak design discharge (Qd). These methods are extensively used by engineers and land managers to perform flow analysis, determining design size of culverts, and identifying the most realistic hydrologic risk for the drainage crossings located on forested watersheds (Keller and Sherrar, 2003; Keller and Ketcheson, 2011). RFA, RM, and USGS-RRE were implemented for estimating Qd for micro-scale watersheds at the White Mountain National Forest (WMNF) region and the Hubbard Brook EF (HBR) within the WMNF region in New Hampshire, and the HJ Andrews EF (HJA) in Oregon. Similarly, FA, RM and USGS-RRE were used to estimate Qd for gauged watersheds in multiple other EFs including the Coweeta Hydrological Laboratory (CHL), Fraser (FRS), and Santee (SAN). The performance of the deterministic methods was evaluated against the results from the RFA and FA, which is important because of the limitations of the deterministic approaches in estimating flood discharge in the complex, heterogeneous micro watersheds. In this pilot study, we estimated the hydrological vulnerability of culverts to flooding for road culverts in the ungauged and gauged micro watersheds in the HBR EF.

The current study is the first multi-site hydrologic risk assessment of road culverts demonstrating a holistic approach using sub-daily/sub-hourly hydro-climatic data to assess flood-induced risk of culvert failure in small, forested headwater catchments across diverse US climate regions. Employing a novel multi-approach framework, it presents the first comprehensive evaluation of these risks, integrating non-stationary regional flood frequency analysis, and associated uncertainties and deviation from deterministic and empirical assessments with culvert hydraulics assessment. The findings from the study will have important implications for managing and restoring road culverts and crossings in forested lands, enhancing their resilience to extreme precipitation and climate-induced flooding.

2. Data and methodology

2.1. Data

The hydrological vulnerability assessment of the RSCLs involves two stages of analysis – (1) hydrologic analysis for estimating Qd, requiring multiple datasets including, hydrologic and meteorological dataset, and watershed geospatial dataset, and (2) estimation of culvert’s hydraulic capacity (Qc) that requires information on their georeferenced locations, dimensions, and type of culverts. Comparing the estimated flood discharge (= Qd x watershed drainage area) for a design event of interest with the Qc of the culvert allows assessment for the hydrological vulnerability of culvert to failure from overtopping. A complete step-by-step workflow of the data collection and processing, and the entire methodology used in the study is demonstrated in Fig. 1. A detailed information of the dataset used for the estimation of Qd and hydrological vulnerability assessment of forest RSCLs or culverts is given in the following sections.

2.1.1. Hydrologic and meteorological dataset

Long-term historical records of 15-min instantaneous streamflow and hourly and sub-hourly precipitation data were used to estimate design discharge for RSCL in one national forest, and five EFs: WMNF and HBR within the WMNF, New Hampshire; HJA, Oregon; SAN, South Carolina; CHL, North Carolina; and FRS, Colorado. Amatya et al. (2021a–c) and Harder et al. (2007) provide a detailed description of the study sites for SAN; Caldwell et al. (2016) and Laseter et al. (2012) for CHL; Likens (2013) and Campbell et al. (2021) for HBR; Johnson et al. (2021) and Fredriksen (1970) for HJA; and Alexander and Watkins (1977) for FRS. The USDA-FS’s EF watersheds used in this study and their characteristics are provided in Table 1 along with the corresponding gauges and periods of data record. The spatial maps demonstrating the study watershed locations, boundaries, elevations, and stream channels are shown in Figs. 2–4.

Watersheds with a larger drainage area encompass a larger distributed system that generally drains into several road culverts simultaneously during a flood event. This may increase the uncertainty associated with linking the design discharge estimated for the watershed and the discharge capacity of any single culvert selected from amongst a network of multiple road culverts draining micro watersheds in the large watershed. Therefore, to minimize such uncertainties, we delineate the larger watersheds located within the WMNF and HBR EF boundary into micro watersheds, ranging between Hydrologic Unit Code 16 (HUC16) to HUC20, hereafter referred to as HUC16-20. The HUC16-20 watershed delineation was performed with Arc Hydro Tool in ESRI ArcGIS platform using the laser imaging, detection, and ranging (LiDAR) based 1-m digital elevation model (DEM) data (available at https://datagateway.nrcs.usda.gov/) (Likens, 2013; Campbell et al., 2021). It is important to note that, hereafter in this study, the delineated gauged watershed ID is referred by appending with “–d” to the USGS and USDA-FS’s gauging station ID that drains out of the delineated watershed as shown in Table 2.

A total 1,387 road stream crossings (RSCLs) within the WMNF were selected in the study (Fig. 2c). These RSCLs were identified using the national hydrography dataset (Buto and Anderson, 2020). In the WMNF region we focused our analysis on the delineated watersheds (HUC16-20) that intersect with these 1,387 RSCLs as shown in Fig. 2. The distribution of drainage area of these 1,387 micro watersheds is presented in Fig. 2f.

Historical 15-min instantaneous streamflow data were obtained from 10 USGS (Kiah and Stasulis, 2018) and four HBR’s streamflow gauges (Table 2; Likens, 2013; Campbell et al., 2021) for the estimation of Qd in the WMNF. For all other EFs, historical 15-min instantaneous streamflow data were obtained only from the USDA-FS streamflow gaging stations (Table 1). The 15-min instantaneous streamflow data were normalized by drainage area of the watersheds before using them in the
RFA, and FA to estimate $Q_p$. Long-term historical hourly and sub-hourly precipitation data and design storm PIs were obtained from the USDA-FS long term onsite rain gauges. 1-hour precipitation data were also obtained from the NOAA’s rain gauges available at [https://www.ncei.noaa.gov/](https://www.ncei.noaa.gov/) that were closest to the USGS’s streamflow gauges at the WMNF (Table 2) due to the unavailability of USDA-FS onsite rain gauges in the region. The hourly and sub-hourly precipitation data were used to estimate the antecedent soil moisture conditions and runoff coefficient (Table 2; see Methods) for the gauged watersheds. Annual average precipitation data for the period 1991–2020 was obtained from Parameter elevation Regression on Independent Slopes Model (PRISM; available at [https://datagateway.nrcs.usda.gov/](https://datagateway.nrcs.usda.gov/)) for estimating $Q_p$ with
the USGS-RRE on the ungauged and gauged watersheds intersecting with RSC within the WMNF region. The 25-year, 50-year, and 100-year precipitation intensities (PI) are generally used for estimating Q, for designing forest drainage culverts and RSCs (Keller and Sherar, 2003; Keller and Ketcheson, 2011). In this study, the 25-year, 50-year, and 100-year PIs, published in Mukherjee et al., 2023, were used for estimating Q for USGS-RRE’s onsite rain gauges listed in Tables 1 and 2. In the WMNF, the 25-year, 50-year, and 100-year PI were obtained from the NOAA-Atlas14 (available at https://datagateway.nrcs.usda.gov/), Enterprise Data Warehouse of US Forest Service (available at https://www.fs.usda.gov/about-agency/enterprise-data-warehouse), and US Transit Roads (available at https://www.bts.gov/maps) for the WMNF region. The location of culverts was identified in GIS environment by using the Arc Hydro Tool and the national hydrography database for the USA (Buto and Anderson, 2020). A field survey was then conducted, and a database was developed, which included, longitude and latitude coordinates, diameter, construction material, and other related information for the 194 culverts in the HBR EF.  

### 2.2. Nonstationary regional frequency analysis

Regional frequency analysis (RFA; Hosking and Wallis, 1997) combines observations of the same variable (here, the annual peak discharge) at various gauged streams to estimate hydrologic information that can be transferred from gauged to ungauged sites. RFA is primarily used to estimate peak flood discharge using the index-flood magnitude (Dalrymple, 1960; Hosking and Wallis, 1997) as a scaling factor of the adimensional flood frequency distribution, which is assumed to be the same for various sites in a homogeneous region. The RFA was performed with the index flood method using L-moments, where the quantile function, \( Q(F) \), of the cumulative distribution function \( F \) at the \( i^{th} \) site (where \( i = 1, 2, \ldots, N \), is given as (Hosking and Wallis, 1997),

\[
Q(F) = \mu_i q(F)
\]  

where, \( \mu_i \) is the index-flood (scaling factor) calculated as the mean of the site-specific annual maximum series (AMS) of peak discharge, and \( q(F) \) is adimensionless quantile function also known as the regional growth curve (or pooled growth function) estimated using the RFA. Once the regional growth curve was obtained for a region using the L-moment statistics, the site-specific quantiles were estimated with Eq. (1). Fig. 1 illustrates the workflow of the methodology used in the RFA.

#### 2.2.1. Inclusion of non-stationary information

We employed the Block Maxima (BM) method to isolate the annual peak discharges. This method is generally applied in extreme value analysis to study the upper tail characteristics of hydrologic extremes (Coles et al., 2001) using a generalized extreme value distribution (GEV). The GEV is a three-parameter distribution comprising of the location \( \mu \), scale \( \sigma \), and shape \( \xi \) parameter, and the theoretical cumulative distribution function can be written following Coles et al. (2001) as,

#### 2.1. Watershed geospatial dataset

We used LiDAR based 1-m resolution DEM data (https://datagateway.nrcs.usda.gov/) to determine the watershed boundary and characteristics, such as slope, mean elevation, longest stream channel length, and topographic wetness index (Thomas et al., 2008; Sörensen et al., 2006) for the headwater watersheds within the WMNF and HBR. The application of 1-m LiDAR based DEM is preferred for representing the micro-watershed scale spatial variability of flood magnitudes for reducing uncertainty associated with the flood discharge estimates draining into any isolated road culvert out of a network of culverts located in small, forested headwater watersheds. 10-m DEM data (available at https://datagateway.nrcs.usda.gov/) were used for other EFs because 1-m LiDAR based DEM data were not available for some of them. However, we recognize that as a future extension of this work, 1-m LiDAR based DEM should be used for delineating these watersheds as and when the 1-m elevation and road culvert data becomes available for the culvert’s hydrologic vulnerability assessment at these EF watersheds. Wetland cover data was obtained from the updated version of National Wetland Inventory (NWI) (available at https://www.fws.gov/program/national-wetlands-inventory/data-download), for calculating the percentage of wetland area within each watershed in the WMNF and HBR EF region. Hydrological soil group data was obtained from the Oak Ridge National Laboratory Distributed Active Archive Center (Ross et al., 2018) and was used to determine the average curve number (Boughton, 1989) in each watershed within the selected study regions.

#### 2.1.3. Road stream crossing and culvert data

Roads and stream crossing (RSC) locations were obtained using the LiDAR based 1-m DEM data and primary and secondary roads data from the Census Bureau’s Topologically Integrated Geographic Encoding and Referencing (TIGER) database (available at https://datagateway.nrcs.usda.gov/), Enterprise Data Warehouse of US Forest Service (available at https://www.fs.usda.gov/about-agency/enterprise-data-warehouse), and US Transit Roads (available at https://www.bts.gov/maps) for the WMNF region. The location of culverts was identified in GIS environment by using the Arc Hydro Tool and the national hydrography database for the USA (Buto and Anderson, 2020). A field survey was then conducted, and a database was developed, which included, longitude and latitude coordinates, diameter, construction material, and other related information for the 194 culverts in the HBR EF.
Fig. 2. Spatial maps showing the (a) WMNF region, major streams and rivers based on the national hydrography (NHD) dataset, and gauging stations (marked in yellow, refer to Table 2 for station ID) used in the study, (b) HUC16-20 watershed boundary (WBD) delineated based on the 1-m LiDAR DRM data, (c) 1-m LiDAR based stream network and 1,387 RSC and culvert locations (yellow circles) within the WMNF region, (d) 1,387 HUC16-20 watersheds intersecting that intersect with the RSC, and culverts (shown in c), (e) 194 culvert locations within the HBR EF, and (f) probability distribution of drainage area of the watersheds (all 14 gauged and 1387 ungauged watersheds). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
The p-quantile of the GEV distribution is then estimated as,

\[ q_p = \left( -1 + \frac{e}{\sigma} \left( x - \mu \right) \right) ^{1/e} \times \frac{\sigma}{e} + \mu, \quad e \neq 0 \]  

(3)

where, \((1-p)\) is the non-exceedance probability.

We used the same method outlined in Laio (2004) to conduct a goodness-of-fit (GOF) test. This method employs the Anderson-Darling (A-D) test statistics with reference to a composite hypothesis that a sample of observations comes from a distribution with unspecified parameters (at 5 % significance level; Laio, 2004). In this study, the GOF test is performed using the “nsRFA” R-package (Viglione et al., 2020).

By changing the model parameters, it is possible to incorporate the non-stationary climatic data in the GEV distribution (Eq. (3)) and capture the changes in the mean and variability within the distribution without changing the distribution’s shape. This was accomplished by considering linear covariates in the \(\mu\), and \(\sigma\) parameter of the GEV model. In the analysis, \(\epsilon\) parameter was kept constant, as it can be unrealistic to vary the \(\epsilon\) parameter as a smooth function (Coles et al., 2001).

Given that significant trends in the AMS of peak discharges were detected based on the Mann-Kendall (M-K) trend test (Mann, 1945; Kendall, 1975) at 95 % confidence level, non-stationary information was included in the design discharge estimation by assuming antecedent soil moisture conditions (AMC) as a covariate in the location and scale parameter of the GEV distribution fit. Based on this assumption, the location and scale parameter were considered to vary linearly with AMC, leading to the choice of three competing GEV models, M0 (stationary), M1 (non-stationary in \(\mu\)), and M2 (non-stationary in both \(\mu\), and \(\sigma\); Table S1). Among these competing models, the best model was selected using the Akaike Information Criteria (AIC) and Likelihood Ratio Test (LRT; McCullagh and Nelder, 1989).

### 2.2.2. Regional parameter estimation

In the RFA, the site-specific parameters of the GEV distribution were calculated using the maximum likelihood estimation (MLE) methodology (El Adlouni et al., 2007). It is important to note that in the FA the site-specific parameters estimated with MLE were used to calculate the flood quantiles and their CIs. However, for the RFA, pooled growth curve estimates were computed after the at-site nonstationary parameter were estimated (O’Brien and Burn, 2014). To produce regional parameter estimates, at-site MLE based parameter estimations were first transformed into their corresponding L-moments of the GEV distribution using the transformation functions proposed by Hosking and Wallis (1997):

\[ \theta^R_k = \frac{\sum_{i=1}^{N} n \theta^{(i)}_k}{\sum_{i=1}^{N} n} \]

(4)

where, \(\theta^R_k\) is the regional estimate of parameter \(k\), \(N\) is the total number
Fig. 4. Spatial maps showing the (a) location of CHL and SAN EF, (b–d) watershed boundary, 10 m DEM elevation, and stream channels for (b) CHL-WS14, (c) CHL-WS27, (d) SAN-WS80, and (e–d) same as in (a–d) but for FRS-ELOUI.
of gauging stations in the region, \( n_i \) is the number of observations at the gauging station, \( i \), and \( \theta^k_i \) is the at-site estimate of parameter \( k \). The transformation functions used in this study were originally proposed by Hosking (1990) and are summarized in the Supplementary Text S1. The calculated values of the site-specific L-moments of the GEV distribution (Hosking, 1990) were substituted to obtain the regional growth curve, \( q(F) \). Finally, the regional growth curve function was substituted in Eq. (1) to calculate the specific quantiles for the specified return intervals (or exceedance probabilities). The confidence intervals (CIs) associated with the quantile estimates were defined by 95 \% confidence level and calculated with Monte Carlo simulations (a 1,000 repeated samplings) using the \textit{lmomRFA} \textit{R}-package (Hosking, 2022). Hosking and Wallis (1997) provide a detailed discussion on the procedure followed for the estimation of the population parameters of the fitted distribution, regional and site-specific precipitation quantiles, and associated CIs.

### 2.2.2. Transferring information from gauged to ungauged watersheds

The design discharge estimates with the FA, RFA, and RM, from the gauged watersheds were mapped to the ungaged watersheds in accordance with the region of influence (ROI) criteria. The ROI was calculated based on the formation of clusters using 10 classification variables or parameters of the gauged and ungauged watersheds located within the WMNF and HBR EF boundary (Table 3). First the gauged sites that form a homogeneous region based on the watershed characteristics were identified using a cluster analysis followed by a heterogeneity test proposed by Hosking and Wallis (1997) and used in previous studies (Bonnin et al., 2006; Hosking and Wallis, 1997; Ngongondo et al., 2011; Srivastava et al., 2019). Finally, the ROI was formed by measuring the similarity between the homogeneous gauged, and ungaged watersheds based on the Euclidean distances in the space, defined by parameters of the gauged and ungaged watersheds, given by,

\[
d_i = \sqrt{\sum_{j=1}^{p}(x_{ij} - x_{ij})^2}
\]

(5)

where, \( d_{ij} \) is the distance between two elements \( i \) and \( j \) given \( p \) different classification variables, and \( x_{ij} \) is the value of the \( h \)-th variable of the \( i \)-th element.

It should be noted that only three gauged watersheds among the 14 selected gauged watersheds in the WMNF region have wetland cover (Table 2). Therefore, to avoid information bias the wetland cover was not used as one of the watershed parameters for the identification of ROI in this study.

### 2.3. Rational method (RM)

The RM (Kuichling 1889) has been commonly used to estimate peak runoff rates in engineering projects in both urban and rural settings. The method has been applied in small watersheds (<80 ha or 200 acres) (Thompson, 2006) and in watersheds up to 260 ha (ODOT, 2014; Weaver and Hagans, 1994). The RM calculates peak flow rates based on the watershed area, precipitation intensity, and a runoff coefficient that indicates the proportion of precipitation that becomes runoff. The main assumptions of this method include (1) rainfall intensity is constant over the course of the storm, (2) rainfall distribution is uniform throughout the drainage area, (3) exceedance probability of peak discharge is same as that of precipitation intensity, (4) the drainage basin has no significant storage, and (5) the runoff coefficient is constant for a given watershed draining a culvert.

The RM can be formulated as,

\[
Q = C_n A
\]

(6)

where \( Q \) is the peak flow rate, \( C_n \) is the unit conversion coefficient, \( C \) is the runoff coefficient, or the percentage of precipitation converted to

### Table 2

Streamflow gauging stations and nearest rain gauges selected in the study for the WMNF and HBR region (within the WMNF region).

<table>
<thead>
<tr>
<th>Site No.</th>
<th>Name of Gauging Station</th>
<th>Station ID</th>
<th>USGS-ID</th>
<th>HUC16-20 WS ID</th>
<th>HUC16-20 WS Area (km²)</th>
<th>HUC16-20 WS Mean Elevation (m)</th>
<th>HUC16-HCU20 WS Slope (%)</th>
<th>HUC16-HCU20 WS Wetland Cover (%)</th>
<th>Data Period Used</th>
<th>Nearest RG Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bearcamp River at South Tamworth</td>
<td>USGS-01064801</td>
<td>USGS-01064801-d</td>
<td>0.4343</td>
<td>258.4</td>
<td>13.7</td>
<td>0.0</td>
<td>1993-2023</td>
<td>RG01</td>
<td>38.1</td>
</tr>
<tr>
<td>2</td>
<td>Rattlesnake Mountain-Baker River</td>
<td>USGS-01076000</td>
<td>USGS-01076000-d</td>
<td>0.0174</td>
<td>165.0</td>
<td>6.0</td>
<td>0.0</td>
<td>2001-2023</td>
<td>RG10</td>
<td>15.9</td>
</tr>
<tr>
<td>3</td>
<td>Haystack Brook-Ammonoosuc River</td>
<td>USGS-01137500</td>
<td>USGS-01137500-d</td>
<td>0.0168</td>
<td>367.1</td>
<td>6.3</td>
<td>0.0</td>
<td>1990-2023</td>
<td>RG01</td>
<td>20.1</td>
</tr>
<tr>
<td>4</td>
<td>Lower Wild River</td>
<td>USGS-01054200</td>
<td>USGS-01054200-d</td>
<td>0.0074</td>
<td>226.9</td>
<td>14.6</td>
<td>0.0</td>
<td>1989-2023</td>
<td>RG61</td>
<td>19.9</td>
</tr>
<tr>
<td>5</td>
<td>Lower Peabody River</td>
<td>USGS-01054114</td>
<td>USGS-01054114-d</td>
<td>0.0860</td>
<td>252.8</td>
<td>7.0</td>
<td>0.0</td>
<td>2012-2023</td>
<td>RG01</td>
<td>5.0</td>
</tr>
<tr>
<td>6</td>
<td>Mile Swamp-Upper Ammonoosuc River</td>
<td>USGS-01100000</td>
<td>USGS-01100000-d</td>
<td>0.0046</td>
<td>291.2</td>
<td>21.3</td>
<td>0.0</td>
<td>1991-2023</td>
<td>RG01</td>
<td>5.0</td>
</tr>
<tr>
<td>7</td>
<td>HBR-W7</td>
<td>HBR-W7</td>
<td>HBR-W7-d</td>
<td>0.1036</td>
<td>734.5</td>
<td>30.8</td>
<td>0.0</td>
<td>1965-2021</td>
<td>RG01</td>
<td>3.9</td>
</tr>
<tr>
<td>8</td>
<td>HBR-W8</td>
<td>HBR-W8</td>
<td>HBR-W8-d</td>
<td>0.2524</td>
<td>826.5</td>
<td>20.8</td>
<td>0.1</td>
<td>1968-2021</td>
<td>RG10</td>
<td>4.8</td>
</tr>
<tr>
<td>9</td>
<td>HBR-W9</td>
<td>HBR-W9</td>
<td>HBR-W9-d</td>
<td>0.1667</td>
<td>665.7</td>
<td>21.9</td>
<td>2.7</td>
<td>1963-2021</td>
<td>RG01</td>
<td>1.1</td>
</tr>
<tr>
<td>10</td>
<td>HBR-W3</td>
<td>HBR-W3</td>
<td>HBR-W3-d</td>
<td>0.0911</td>
<td>670.9</td>
<td>19.8</td>
<td>0.0</td>
<td>1957-2021</td>
<td>RG01</td>
<td>1.1</td>
</tr>
<tr>
<td>11</td>
<td>Beaver Brook-Pemigewasset River</td>
<td>USGS-01075000</td>
<td>USGS-01075000-d</td>
<td>0.0573</td>
<td>199.4</td>
<td>5.5</td>
<td>0.0</td>
<td>2001-2023</td>
<td>RG01</td>
<td>4.6</td>
</tr>
<tr>
<td>12</td>
<td>Mason Brook-Saco River</td>
<td>USGS-01064500</td>
<td>USGS-01064500-d</td>
<td>0.0238</td>
<td>141.3</td>
<td>6.5</td>
<td>0.0</td>
<td>1987-2023</td>
<td>RG01</td>
<td>32.5</td>
</tr>
<tr>
<td>13</td>
<td>Outlet East Branch Pemigewasset River</td>
<td>USGS-01074520</td>
<td>USGS-01074520-d</td>
<td>0.0169</td>
<td>262.1</td>
<td>13.7</td>
<td>32.6</td>
<td>1993-2023</td>
<td>RG01</td>
<td>11.8</td>
</tr>
<tr>
<td>14</td>
<td>Bartlett Brook-Saco River</td>
<td>USGS-01064250</td>
<td>USGS-01064250-d</td>
<td>0.0313</td>
<td>209.8</td>
<td>4.5</td>
<td>0.0</td>
<td>2009-2023</td>
<td>RG01</td>
<td>19.5</td>
</tr>
</tbody>
</table>
runoff, \( i \) is the design rainfall intensity based on the watershed time of concentration \((T_c)\), and \( A \) is the watershed drainage area. \( T_c \) was calculated with the NRCS time-lag method that uses the longest stream channel length and average curve number of a small or medium sized watershed (Folmar et al., 2007). As we were interested in the peak discharge response to extreme storm events, the mean \( C \) value was generated by back-calculation for any given gauged watershed by applying the RM to the top 1000 extreme storm events with their PIs charges. The mean \( C \) value was further adjusted by factors of 1.1, 1.2, and 1.25 for return intervals of 25, 50, and 100 years, respectively, as used in Hayes and Young (2006) and advised by ODOT (2014) to design for extremely rare flood events. The Supplemental Material Text S2 and S3 include a thorough explanation of how \( T_c \) and \( C \) were derived in this study.

### 2.4. USGS regional regression equation (RRE)

The USGS RREs are generated with generalized least squares regression based on regionalized flood frequency data for a collection of gauged watersheds within a hydrologically homogeneous region. These predictive equations are frequently used to estimate design discharge at various return intervals and illustrate the statistical link between flood frequency analysis-based estimations and basin variables (e.g., basin area, impervious percentage, slope, maximum rainfall intensity, etc.) (Table S2). Following Amatya et al. (2021a–c), we used the equations reported by Weaver et al. (2009) for the CHL EF rather than the most recent Feaster et al. (2014) equations because the later did not include equations for region 2 (Blue Ridge ecoregion) in North Carolina, where the CHL is located.

### 2.5. Evaluation of the design discharges

In this study, percentage bias (PB; Mizukami et al., 2019) was calculated to evaluate how well \( Q_p \) estimates from the deterministic and semi-empirical methods like the RM, and statistical methods like the USGS RRE agreed with the probability based \( Q_p \) estimates derived from observed streamflow records. PB is calculated as,

\[
P_B = \frac{(Q_p - Q_e)}{Q_p} \times 100
\]

where:

- \( Q_p \) = design discharge, calculated using the probability-based frequency analysis,
- \( Q_e \) = design discharge, calculated using the deterministic methods, such as, the RM, and USGS RREs.

### 2.6. Hydrologic vulnerability assessment for road culverts

Hydrologic vulnerability of the road culverts in this study was assessed based on the ability of the culvert to accommodate the design discharge of interest without a failure due to overtopping. This was achieved by comparing the \( Q_c \) of a culvert draining a specific watershed with the design peak discharge estimate for that watershed at its outlet. The \( Q_c \) is defined as the allowable passage flow rate through a culvert and was calculated using the orifice flow equation assuming a submerged inlet control. Mathematically, the orifice flow equation (Malcom, 1989) gives the mass flow rate (or \( Q_c \)) through an orifice, given the area of the orifice, \( A \), and can be written as:

\[
Q_c = C_d \times A \times \sqrt{2gH}
\]

where \( C_d \) is the coefficient of discharge, \( g \) is the acceleration due to the gravity of Earth (in m/s²), and \( H \) is the driving head or the mean center line (the distance between the water surface level and the centroid of the culvert, in m). The \( C_d \), a dimensionless coefficient of discharge, is equal to 0.62 for a square-edged entrance and when the flow approaches a well-rounded entrance with a submerged inlet (bank-full) condition (Franzini and Finnemore, 1997). While ASCE (1992) suggests equation (8) is applicable only when \( H/D \geq 2 \), Franzini and Finnemore (1997) established that the error is only 2 % when \( H/D \geq 1.2 \), where \( D \) is the diameter of the culvert. In this study, we based our calculations on the later assumption that \( H \approx 1.2D \).

For a given return interval or exceedance probability of a design discharge, a culvert is deemed hydrologically vulnerable if the design discharge for the watershed is greater than either the bank full discharge or the magnitude of \( Q_c \), representing the culvert capacity. However, it was beyond the scope of this study to quantify the probability of failure risk of these vulnerable culverts by accounting for the exceedance of design discharge over their life span used in cost-benefit analyses (Hansen, 1987).

### 3. Results and discussion

#### 3.1. Comparison of peak design discharge estimates and uncertainties for the WMNF region

The magnitude of the annual peaks of 15-min instantaneous streamflow discharges per unit drainage area and their linear trends for the 14 gauging stations selected within and adjacent to the WMNF, and HBR EF are shown in Fig. S1. Out of the 14 gauging stations, the annual peak discharges per unit drainage area for the three gauging stations in HBR, HBR-WS3, HBR-WS6, and HBR-WS7 were found to exhibit statistically significant (at 95 % confidence level) increasing trends based on the Mann Kendall (MK) Trend test (Fig. S1, Table 2; Mann, 1945; Kendall, 1975). The results of the GOF test, used for assessing the suitability

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Abbreviation</th>
<th>Description and source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LON, LAT</td>
<td>Coordinate of the centroid of watershed derived based on watershed delineation processed in Arc Hydro using the LiDAR based 1-m digital elevation model data.</td>
</tr>
<tr>
<td>2</td>
<td>TOPWET</td>
<td>Topographic wetness index, ( \ln(a/S) ); where ( \ln ) is the natural log, ‘a’ is the upslope area per unit contour length and ‘S’ is the slope at that point. (LiDAR based 1-m digital elevation model data)</td>
</tr>
<tr>
<td>3</td>
<td>HUC16-20 Longest Channel Length</td>
<td>Processed in Arc Hydro using LiDAR based 1-m digital elevation model data.</td>
</tr>
<tr>
<td>4</td>
<td>PRECIP</td>
<td>Mean annual precipitation from the PRISM dataset for 1991-2020</td>
</tr>
<tr>
<td>5</td>
<td>ELEV</td>
<td>Mean watershed elevation (meters) from LiDAR based 1-m digital elevation model data.</td>
</tr>
<tr>
<td>6</td>
<td>SLOPE</td>
<td>Average Slope (%) of the watershed derived from LiDAR based 1-m digital elevation model data.</td>
</tr>
<tr>
<td>7</td>
<td>AREA</td>
<td>Watershed drainage area, km²</td>
</tr>
<tr>
<td>8</td>
<td>HSOIL</td>
<td>Hydrological Soil group obtained from Oak Ridge National Laboratory Distributed Active Archive Center.</td>
</tr>
<tr>
<td>9</td>
<td>CN</td>
<td>Average curve number of a watershed derived from the Hydrological Soil group dataset.</td>
</tr>
<tr>
<td>10</td>
<td>TcNRCS</td>
<td>Time of concentration derived using the NRCS lag-time method.</td>
</tr>
</tbody>
</table>
of the extreme value distribution used in the analysis are summarized in Table 4. The GOF test indicated that the GEV distribution shows a statistically good fit at 95% confidence level for all gauging stations, except for one gauging station for which the Gumbel distribution is indicated as a statistically good fit.

Numerous studies (Eisenbies et al., 2007; Beier et al., 2015; Fang and Pomeroy, 2016; Ceá and Fraga, 2018) have highlighted increase in the AMC was linked to significantly increasing trends in streamflow discharges leading to more severe floods in hydrologic systems, such as in the HBR EF. This is particularly important in the context of changing climate and the increased frequency of precipitation extremes that have led to extended periods of saturated soil conditions that contribute to floodings (Eisenbies et al., 2007; Mukherjee et al., 2023), and landslides in some cases (Ward et al., 2020; Kirschbaum et al., 2020). The Pearson correlation analysis suggests a statistically significant (at 95% confidence level) relationship between AMC and the annual peak discharges per unit drainage area at the HBR-WS3, HBR-WS6, and HBR-WS8 gauging stations as shown in Fig. S2. AMC was calculated as the 5-day sum of 24-hr precipitation (obtained from the nearest rain gauge station, Table 2) prior to the day of the peak discharge event. Subsequently, non-stationary GEV modeling assumptions were applied for these three gauging stations with the AMC as a covariate, which were further evaluated to determine the best fit GEV model using the AIC and LRT test criteria as shown in Table S3. Table S4 illustrates the GEV parameters estimated with MLE, their transformed L-moments and L-moment ratio (see Text S1 for Methods), and estimated index flood magnitude used in the RFA for the gauged watersheds.

The 25-yr, 50-yr, and 100-yr Qₚ estimated by the RM and USGS-RRE were focused on the HUC16-20 watersheds corresponding to the 14 gauging stations at the WMNF and HBR EF. Tables S5 and S6 list the variables and their values used in the calculation of the peak discharges with the RM and USGS-RRE (see Methods). The Tc was found to vary from 7-mins to 3-hr for the gauged HUC16-20 watersheds (Table S5).

The magnitudes of Qₚ (and uncertainties) for 25-yr, 50-yr, and 100-yr return periods, calculated using the three selected methods, the stationary or non-stationary RFA, RM, and USGS-RRE for the selected gauging stations, are shown in Fig. 5(a–c) and Tables S7–S9. The agreements among these three methods were evaluated based on the PE statistics (see Methods), as shown in Fig. 5(d–f) and Table S10. The Qₚ magnitudes, calculated using the RFA, are comparable with the magnitudes of observed annual peak discharges at all the gauging stations, particularly for HBR-WS8. This comparability is also noteworthy at HBR-WS7, where non-stationarity was detected in the mean as well as in the variability of the distribution of annual peak discharges (Table S3). Consequently, the 25-yr, 50-yr, and 100-yr Qₚ magnitudes at this gauging station were greater than the other gauging stations, HBR-WS3, and HBR-WS6. Importantly, HBR-WS7-d, and HBR-WS8-d are north-facing watersheds which naturally get more precipitation along with lower evapotranspiration, potentially resulting in higher AMC than other watersheds within the HBR EF (Campbell et al., 2011). This is reflected in the Qₚ magnitudes estimated by the RFA for HBR-WS7-d, and HBR-WS8-d which was greater than the values estimated for the rest of the watersheds.

Furthermore, for most of the gauged watersheds, the 25-yr, 50-yr and 100-yr Qₚ values estimated by the RFA agree well with those estimated by RM and the USGS-RRE. This is indicated by the magnitude of estimated PB which fell within the range of 0–50 % for most of the gauges. However, there was considerable disagreement among methods in some cases, especially the watersheds, USGS-01064801-d, USGS-01076000-d, USGS-01130000-d, USGS-01137500-d, and USGS-01064500-d. Among these gauging stations, the PB in the peak discharge estimates for the watersheds USGS-01076000-d, USGS-01137500-d, USGS-01130000-d, USGS-01064500-d suggest that the RM method overpredicts the 25-yr, 50-yr, and 100-yr Qₚ estimates by 296–306 %, 99–111 %, 406–440 %, 102–115 %, respectively, as compared to those estimated by the RFA. On the other hand, the USGS-RRE method overpredicts the 25-yr, 50-yr, and 100-yr Qₚ estimates for USGS-01064801-d, USGS-01130000-d, and USGS-01064500-d watersheds by 254–293 %, 394–485 %, and 319–405 %, respectively, as compared to those estimated by the RFA. The considerably large and unidentified uncertainty in Qₚ could come from extrapolating the prediction equations to smaller watersheds outside the boundaries of the regression data used to develop the USGS-RRE (Genereux, 2003).

Fig. S3 illustrates the ROI of the HUC16-20 gauged watersheds across the ungauged watersheds that intersect with the RSCS within the WMNF and HBR region. The 25-yr, 50-yr and 100-yr Qₚ values for the ungauged HUC16-20 watersheds are shown in Fig. S4, Fig. S5, and Fig. 6, respectively. The Qₚ estimated by the RFA and RM method show a better agreement for almost all the watersheds as compared to the USGS-RREs, although the RM slightly underpredicts the Qₚ in both gauged and ungauged watersheds. This is indicated by the lower magnitude of PB values for the ungauged watersheds (Fig. 6d) which are mostly within a range of 0 % to 50 % for Qₚ estimated by the RM. In contrast, the USGS-RREs overpredict the Qₚ by 25–50 % for most of the watersheds, as indicated by the magnitude of PB in Fig. 6e. Interestingly, USGS-RREs use percentage of wetland area as one of the key variables for Qₚ estimation for streams in New Hampshire. Results indicate that the USGS-RREs yielded considerably greater Qₚ estimates and a higher over-prediction (greater PB) for the study watersheds with fewer wetland areas as compared to the watersheds with larger wetland areas, as demonstrated in Fig. S6. A significantly strong negative relationship between the Qₚ and proportion of wetland area of the ungauged watersheds (Pearson Correlation coefficient, R = −0.6; p-value < 0.05) can be seen in Fig. S6 (f). The greater storage provided by wetlands and floodplain in the watershed attenuates the peak discharge in the channel (Amaty et al., 2019), which may have ultimately impacted the Qₚ.
The RM performed well in small watersheds with drainage area of less than the 120-ha limit recommended for the method in the literature (Thompson, 2006; ODOT, 2014). All the study watersheds are at a scale of HUC16-20 with a drainage area of less than 50-ha, whereas the USGS equations were derived using data from watersheds with a drainage area greater than 180-ha (Olson, 2008). This raises questions about the reliability of using the USGS-RREs as compared to the RM for estimating $Q_p$ estimates for these small watersheds. One possible explanation for the better performance of RM in small watersheds is that the RM assumes negligible water storage and, accordingly, the flood estimates respond linearly to increases in PI (Amatya et al., 2021a–c) unlike the USGS-RREs that were developed assuming a linear connection between flood estimates and the drainage area of the watersheds. This also

---

**Fig. 5.** Comparison of (a) 25-yr, (b) 50-yr, and (c) 100-yr $Q_p$ (shown by bar) and 5–95% confidence intervals (shown by error bars) for HUC16-20 gauged watersheds (shown by WS ID) selected within and adjacent to the WMNF (and HBR EF) region estimated by the RFA, RM, and USGS-RRE methodology. The corresponding PB (%) for the 25-yr, 50-yr and 100-yr peak discharge are shown in (d–f) for the RM and USGS-RRE methodology. It is important to note that for the RFA analysis, non-stationary model was selected as the best fit model for the gauged watersheds, HBR-WS3-d, HBR-WS6-d, and HBR-WS7-d, and stationary model was selected as the best fit for the rest of the watersheds (refer to Table S3).
explains why the PB for the $Q_p$ estimated by RM is greater for the higher RI, while PB for the $Q_p$ estimated with the USGS-RRE is independent of the increase in RI (Fig. 5(d–f), and Fig. 6(d–e)).

3.2. Comparison of peak design discharge estimates for other study watersheds

Fig. S7 depicts the magnitude of the yearly peaks of 15-min instantaneous discharge per unit drainage area and associated linear trends for the seven gauging stations for the reference watersheds, two within the CHL (WS ID: CHL-WS14, and CHL-WS27), one within the FRS (WS ID: FRS-ELOU1), three within the HJA (HJA-WS1, HJA-WS8, and HJA-WS9), and one within the SAN (WS ID: SAN-WS80). Given that 15-min instantaneous streamflow data was available for more than two reference watersheds only within the HJA, we performed the RFA for the gauging stations located within that HJA EF, whereas single-site
frequency analysis (FA) was performed for study watersheds in all other EFs. A storm on October 3–4, 2015, influenced by Hurricane Joaquin caused nearly 500 mm rainfall within 48-hr in South Carolina. This storm potentially raised the peak discharge measured at the SAN-WS80 gauging station for the year 2015 (St. George and Mudelsee, 2019) as shown in Fig. 57, causing high non-linearity and artifact in the annual variability of the flood peaks over the selected period. Therefore, an additional analysis was also performed for SAN-WS80 by removing the year 2015 peak discharge, hereafter denoted as SAN-WS80R.

For CHL-WS14 and CHL-WS27 watersheds, the Gumbel method was selected as the best fitted distribution per the GOF test based on AD statistics (Table 5). For all other EF watersheds, the GEV distribution was selected as the best fitted distribution. The yearly peak discharge for the HJA-WS1 gauging stations, was found to have a statistically significant (at 95 % confidence level) increasing trend, as determined by the M-K Trend test (Fig. 57, Table 5; Mann, 1945; Kendall, 1975). As such, non-stationary GEV modeling assumptions were applied for this gauging station with AMC as covariate, which was further evaluated to determine the best fit GEV model using the AIC and LRT test criteria as indicated in Table S11. Table S12 illustrates the transformed L-moments, L-moment ratio (see Text S1 for Methods), and estimated index flood magnitudes used in the RFA for the HJA’s study watersheds. Table S13 illustrates the 25-yr, 50-yr, and 100-yr Qp values estimated with MLE for both RFA and FA for all the EFs.

Large disparities between the Qp estimated by the RM using on-site PIs and the USGS-RRE are evident in Fig. 7 and Tables S13–S16. In CHL watersheds, both the RM and USGS-RRE overpredicted the Qp estimates compared to the FA. However, the RM performed better than the USGS-RRE which can be attributed to the smaller size of the watersheds (CHL-WS14: 61-ha, and CHL-WS27: 39-ha) where the assumptions of uniform rainfall and negligible storage hold. Furthermore, measurements from rural watersheds larger than 260 km² area were used to create the USGS-RRE for the Blue Ridge region of North Carolina, where the CHL-WS14 and CHL-WS27 watersheds, with substantially smaller drainage area, are located (Weaver et al., 2009). The USGS-RRE substantially overpredicted the Qp estimates for CHL-WS14 which is a relatively larger watershed compared to CHL-WS27. In contrast, the Qp estimated by the RM is greater for the smaller watershed, CHL-WS27, with overprediction of the Qp by as much as 88–124 % compared to the RFA using the observed streamflow records. However, large uncertainty in Qp can be noted for the RM method which may stem from the large uncertainty of onsite design storm PI at the CHL-RG31 (Table 1; Mukherjee et al., 2023). These findings are consistent with previous studies on the same forested watersheds (Genereux, 2003, Amatya et al., 2021a–c). For example, Genereux (2003) suggested that the RM, based on a 50-year Qp estimate, outperformed other methods such as the USGS RRE, the NRCS TR-55 method, and the North Carolina Department of Transportation methods for a 19.6-ha 66 % forested watershed in North Carolina’s Piedmont region.

Interestingly, both the RM and USGS-RRE performed well at FRSL-ELOUI watershed with RM and USGS-RRE underpredicting the 100-yr Qp by 61 %, and 46 %, respectively compared to the FA (Fig. 7, Table S16). Although the drainage area of the watershed is considerably large (259-ha) for the assumptions of the RM to hold, the high gradient (average slope of 72.4 %) mountainous streams likely overcome the prediction error associated with minimal storage assumption of the RM (David, 2011) that has been identified to be a source of prediction error for low-gradient, 160-ha watershed with substantial storage (Amatya et al., 2021a–c).

At the HJA there was a strong relationship between flood estimates and drainage area. For example, RM performed well compared to RFA for the small watershed, HJA-WS8 (20.4-ha), underpredicting 25-yr, 50-yr, and 100-yr Qp by 11 %, 4 % and 0 %, respectively. PB associated with the RM was relatively greater for HJA-WS1 (95.9-ha or 0.959 km²), overpredicting the 25-yr, 50-yr, and 100-yr Qp by 43 %, 39 %, and 37 %, respectively. Both USGS-RRE and RM performed poorly in predicting the 100-yr Qp for the very small watershed, HJA-WS9 (8.5-ha), overpredicting Qp by 287 %, and 377 %, respectively.

In SAN-WS80, the RM underestimated the design discharge by up to 61 % for a return interval greater than 25-years. On the other hand, the USGS-RRE underestimated it by just 8 %. This is in agreement with Amatya et al. (2021a–c) which found similar large underestimation by RM and somewhat overprediction by the USGS-RRE at that time when streamflow data, available only through 2016, was used. In addition, both FA and RM showed large uncertainties in their Qp estimates. The large uncertainties in Qp estimated by FA can be attributed to the statistical artifact imposed by the non-linearity in the variance of annual peak streamflow records at the SAN-WS80 gauging station caused by the indirect influence of the 2015 Hurricane Joaquin in South Carolina. Interestingly, the uncertainty associated with the Qp at this site significantly reduced after removing the flow event associated with this 2015 extreme event as shown in Fig. 7 for the site ID SAN-WS80R. The large uncertainties in Qp estimated by the RM, however, is most likely because it is gently sloped watershed with significant storage (Amatya et al., 2021a–c) and a large drainage area (160-ha), which exceeds the 120-ha limit specified for the approach (Thompson, 2006; ODOT, 2014; Amatya et al., 2021a–c). Overall, these findings indicate that the RM works better for smaller watersheds (WS14 and WS27 in NC, and HJA-WS8 in OR) with drainage area up to 120-ha, whereas the USGS RRE is more reliable for larger watersheds (SAN-WS80, FRSL-ELOUI) at flood return intervals ( exacerbated by 25-yr), normally of importance for design/sizing of RSCS and culverts.

However, it is important to note that with the RM there is a lack of consideration for any storage features, such as wetlands, channels, and floodplains, that do not entirely fill and reach a continuous inflow-outflow state during a storm event (ODOT, 2014). This omission can be an important source of error in deriving Qp on the watersheds with considerable portion of wetland area within the WMNF region and may be supplemented by the uncertainties associated with the estimation of the runoff coefficient stemming from uncertainties in PIs (Grimaldi and Petroselli, 2015a,b; Tedela et al., 2012). Accumulated snow cover in snow-dominated watersheds in the HJA, HBR, and FRs EF can also lead to greater magnitudes and uncertainties of the runoff coefficient (Merz and Blöschl, 2009). Previous research has shown that combining runoff coefficients and Tc obtained from field measurements, where possible

<table>
<thead>
<tr>
<th>Gaging Station / WS ID</th>
<th>Number of years analyzed (Tc)</th>
<th>MK Test (p-value)</th>
<th>Distribution</th>
<th>AD Statistics</th>
<th>AD test (p-value)</th>
<th>Type of Analysis Performed</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL-WS14</td>
<td>39</td>
<td>0.61</td>
<td>Gumbel</td>
<td>0.66</td>
<td>0.91</td>
<td>FA</td>
</tr>
<tr>
<td>CHL-WS27</td>
<td>40</td>
<td>0.27</td>
<td>Gumbel</td>
<td>0.60</td>
<td>0.88</td>
<td>FA</td>
</tr>
<tr>
<td>FRSL-ELOUI</td>
<td>78</td>
<td>0.66</td>
<td>GEV</td>
<td>0.27</td>
<td>0.39</td>
<td>FA</td>
</tr>
<tr>
<td>HJA-WS1</td>
<td>67</td>
<td>0.02</td>
<td>GEV</td>
<td>0.48</td>
<td>0.86</td>
<td>RFA</td>
</tr>
<tr>
<td>HJA-WS8</td>
<td>56</td>
<td>0.90</td>
<td>GEV</td>
<td>0.29</td>
<td>0.54</td>
<td>RFA</td>
</tr>
<tr>
<td>HJA-WS9</td>
<td>51</td>
<td>0.17</td>
<td>GEV</td>
<td>0.39</td>
<td>0.75</td>
<td>RFA</td>
</tr>
<tr>
<td>SAN-WS80R</td>
<td>40</td>
<td>0.77</td>
<td>GEV</td>
<td>0.36</td>
<td>0.73</td>
<td>FA</td>
</tr>
<tr>
<td>SAN-WS80</td>
<td>41</td>
<td>0.99</td>
<td>GEV</td>
<td>0.37</td>
<td>0.74</td>
<td>FA</td>
</tr>
</tbody>
</table>
(Hayes and Young, 2006; Thompson, 2006; Trommer et al., 1996) and varied return intervals (McEnroe et al., 2013), can greatly enhance RM estimations.

3.3. Hydrologic vulnerability of culverts within the HBR

Hydrologic vulnerability assessment of road culverts was performed only for the HBR EF region using a dataset of total 194 culverts. The culvert diameters (D) ranged from 0.38 m (15 in) to 1.83 m (72 in), as shown in Fig. 8a. Fig. 8b illustrates the estimated hydraulic capacity of each of these culverts (see Methods for estimation of hydraulic capacity). Out of the 194 culverts, 140 culverts have a diameter of 0.46 m (18 in). The minimum and maximum hydraulic capacity across these culverts with different sizes vary between 0.2 m³ s⁻¹ (for D = 0.38 m) and 10.7 m³ s⁻¹ (for D = 1.83 m).

Fig. 8c-d) illustrate the hydrologic vulnerability of the 194 culverts to various levels of flood discharge, i.e., for 25-yr, 50-yr, and 100-yr flood return periods, respectively, estimated by the RFA, FA, RM, and USGS-RRE methods. The corresponding PB values for the 25-yr, 50-yr and 100-yr peak discharge are shown in (d–f) for the RM and USGS-RRE methodology. It is important to note that for the RFA analysis, non-stationary model was selected as the best fit model for the gauged watershed, HJA-WS1, and stationary model was selected as the best fit for the rest of the watersheds (refer to Table S11).

Fig. 7. Comparison of (a) 25-yr, (b) 50-yr, and (c) 100-yr Qₚ (shown by bar) and 5–95% confidence intervals (shown by error bars) for HJA, CHL, SAN, and FRS EF watersheds estimated by the RFA, FA, RM, and USGS-RRE methods. The corresponding PB values for the 25-yr, 50-yr and 100-yr peak discharge are shown in (d–f) for the RM and USGS-RRE methodology. It is important to note that for the RFA analysis, non-stationary model was selected as the best fit model for the gauged watershed, HJA-WS1, and stationary model was selected as the best fit for the rest of the watersheds (refer to Table S11).
undersized. These culverts are considered as hydrologically vulnerable for that flood return period, as shown by red circles.

A higher percentage of road culverts was found to be vulnerable to overtopping based on the $Q_p$ estimated by the USGS-RRE compared to the RFA and RM (Fig. 8). For example, out of the total 194 culverts, 40 % (or 78 culverts, 91 % of which has $D = 0.46$ m), 42 % (or 83 culverts out of which 86 % has $D = 0.46$ m, and 12 % has $D = 0.61$ m), and 56 % (or 109 culverts out of which 1% has $D = 0.38$ m, 87 % has $D = 0.46$ m, and 11 % has $D = 0.61$ m) of the culverts are vulnerable to 25-year, 50-year, and 100-year $Q_p$, respectively, estimated with the USGS-RRE (Fig. 8f, and Table S17). On the other hand, 19 %, 25 %, and 31 % of the culverts are vulnerable based on the 25-year, 50-year, and 100-year $Q_p$ estimated by RFA. Similarly, 17 %, 24 %, and 31 % of the culverts are vulnerable to overtopping based on the 25-year, 50-year, and 100-year flood estimated by the RM. Most of these culverts have $D = 0.46$ m.

Overall, except for one culvert with diameter of 0.91 m, none of the culverts sized 0.75 m in $D$ or larger were identified as hydrologically vulnerable based on the selected RIs and methods. Therefore, our investigation reveals that replacement of undersized culverts with new ones having a $D$ larger than or equal to 0.75 m is likely to increase flood resiliency of the culverts within the HBR EF. However, these findings are only valid (1) for bank-full conditions where distance between the water level and the center of the culvert, $H = 1.2D$, (2) considering solely the modeled hydrologic vulnerability to overtopping, with no assessment of probability of risk of failure due to exceedance of design discharge.
during the intended culvert life span, and (3) assuming no vulnerability to geomorphological risk factors (e.g., sedimentation, bank erosion, siltation, caving, and debris flow), or fish passage requirements are necessary.

An increase in culvert size raises overall installation and material costs of the structure by roughly 14% on average (Piehl et al., 1988). This additional expense is minor in exchange for a large increase in culvert flow capacity and a 50% to 75% reduction in the probability of culvert failure over the physical life of the structure. Therefore, final decision of culvert sizing, and replacement or restoration is assumed to be on a forest manager or engineer who would decide on allowable risks of failure and for costs of restoration based on the culvert’s significance and intended life span using an economic cost-benefit analysis (Hansen et al., 2009; Gillespie et al., 2014).

4. Summary and conclusion

This study provides a comprehensive assessment of hydrologic risk from flooding in several gauged and ungauged forested watersheds by implementing advanced statistical and widely used deterministic approaches using the long-term high resolution (15-min) instantaneous streamflow records from multiple United States Department of Agriculture Forest Service (USDA-FS) and United States Geological Survey (USGS) stream gauges and various characteristics of the watersheds draining at those gauged outlets. The objective of this study was three-fold: (1) to estimate peak design discharge ($Q_p$) for 25-year, 50-year and 100-year flood return periods (which are mostly preferred for road culvert design in the national forest lands) based on stationary and non-stationary, probability based at-site frequency analysis (FA) and regional frequency analysis (RFA) for gauged watersheds and transfer that information to the ungauged watersheds, (2) to perform a watershed-scale comparison between the $Q_p$ estimated by FA and RFA and $Q_p$ obtained from the deterministic Rational Method (RM) and semi-empirical USGS regional regression equations (USGS-RRE), and (3) to assess the hydrologic vulnerability of 194 road culverts to overtopping by floods in a pilot EF region, HBR, for which a complete culvert dataset was curated by USDA-FS personnel. To the best of our knowledge, this is the first study that provides a comprehensive multi-approach assessment of hydrologic risks of culvert failure from floods in small, forested headwater watersheds across different climate regions of the US.

The results from this study are rich in watershed scale hydrologic-risk information and can help with an effective assessment for increasing flood resiliency of road culverts in the studied watersheds: 1387 watersheds in the White Mountain National Forest (WMNF) and Hubbard Brook (HBR), and watersheds in multiple other EFs, including three watersheds in HJ Andrews (HJA), two watersheds in Coweeta Hydrologic Lab (CHL), and one watershed each in the Santee (SAN), and Fraser (FRS). The RM outperformed the USGS-RRE in estimating $Q_p$ in the WMNF’s smaller watersheds in New Hampshire, and in three other small, high-relief headwater watersheds, including the CHL-WS14 in North Carolina and HJA-WS8 in Oregon. The USGS-RRE, on the other hand, performed better at the larger watersheds, such as Fraser’s East Louis (FRS-ELOU) watershed in Colorado and the SAN-WS80 watershed in South Carolina. Antecedent soil moisture condition was strongly correlated with the annual peak discharges per unit drainage area observed from the HBR-WS3, HBR-WS6, and HBR-WS7 gauging stations which exhibited strong non-stationarity. The USGS-RREs produced substantially higher $Q_p$ estimates and a higher overprediction (larger percentage bias (PB)) for watersheds with a lower percentage of wetland area than the watersheds with a higher percentage of wetland area within the WMNF region. The poor performance of the RM, as expected in the larger watersheds (drainage area > 120-ha), such as SAN-WS80 can be attributed to its large drainage area and high-water storage capacity, supporting earlier results reported by Amatya et al. (2021a–c) for SAN-WS80. In the HBR, a greater number of road culverts were found to be vulnerable to overtopping due to flooding based on the 100-yr flood design discharge estimated by the USGS-RRE as compared to the RFA and RM although USGS-RRE method may overestimate flood design discharge for small watersheds like the one selected in the study.

These results suggest the RFA method provides the least uncertainty in $Q_p$ estimation among the methods studied in EF watersheds for which observed streamflow data from multiple gauges are available to ensure robust information transfer from the gauged to the ungauged watersheds. The results from the hydrologic vulnerability assessment revealed that about 31%, 26% and 56% of the culverts are vulnerable to overtopping due to 100-yr $Q_p$ based on all the three methods, RFA, RM and USGS-RRE, respectively. Almost all the culverts, hydrologically vulnerable to overtopping due to flooding within the HBR EF have a diameter of less than 0.75 m. This finding indicates that replacing undersized culverts with new culverts with a D larger than or equal to 0.75 m can provide some insurance in the face of increasing extreme precipitation events within the HBR EF. However, these findings are only valid for bank-full conditions and/or where distance between the water level and the center of the culvert, $H = 1.2D$, and locations where the structure is not prone to geomorphological vulnerability (such as sedimentation, bank erosion, siltation, caving, and debris flow) and permits aquatic organism passage. It is important to note that the design discharge estimated for the rest of the watersheds with BSCs in the WMNF can be used in the near future for hydrologic vulnerability assessment of the road culverts as and when the culvert data becomes available.

There are several logical extensions to the vulnerability assessment described in this article. Evaluation of hydraulic capacity in the presence of large woody debris is an important factor in forested watersheds that is missing in the current analysis. Integrating hydrologic vulnerability with a geomorphological vulnerability assessment would provide a more comprehensive and reliable assessment. In addition, design flood exceedance probability-based failure risk assessment of culverts over their intended life span used in economic cost-benefit analysis was also left for future studies. As a result, the findings of these case studies should be evaluated with caution and by incorporating present field conditions before undertaking design initiatives or projecting them to other similar watersheds. Relief culverts primarily provides road-side ditch-drainage relief and hydrologic connectivity through cross-drainage, therefore, special care should be taken to isolate and treat the hydrologic vulnerability of relief culverts differently from the other culverts draining a specific upland area. The methodology used in the hydrologic vulnerability assessment of culverts can be replicated for other EF regions as and when culvert data becomes available. Further improvements are possible with the estimation of $Q_p$ using a continuous simulation approach that tracks the soil moisture dynamics (Winter et al., 2019) which can reduce uncertainties in $Q_p$ estimates for watersheds with high water retention capability. Similarly, model parameters determined directly on the basis of physiographic characteristics of catchments that are able to simulate the rainfall losses and the dynamics of runoff formation taking into account the spatial variability of the catchment, can lead to further improvements of the results (Petroselli and Grimoldi, 2018; Mlynski et al., 2020). Precipitation-intensity-frequency-duration estimates for future time-periods, developed using the Coupled Model Intercomparison Project Phase 6 outputs, can be used in estimating $Q_p$ for designing possible climate change resilient culverts as another future scope of this study (Jalowska et al., 2021).

Overall, the results from this study can be useful for strategically investing in road-stream crossing road design and re-design. This will foster flood resiliency that has the potential to reduce the economic and societal costs through reduced failure rates and lower maintenance costs while maintaining important ecological values of the forested watersheds. However, it is worth mentioning that engineers must carefully select the most appropriate design strategy for developing new and renovated roads depending on government criteria (USDOT, 2012, 2018) and proper engineering judgment for the specific watershed and drainage culverts. After completion of the design, careful consideration
should also be given to the proposed installation procedures for the drainage culvert, such as checking for alignment, stream dimension, and substrate composition (Hansen et al., 2009). Finally, developing and implementing effective monitoring procedures and strategies for the drainage culverts will maintain its flood resiliency throughout the life of the structure.

CRedit authorship contribution statement

Sourav Mukherjee: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. Devendra M. Amatya: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. John L. Campbell: Resources, Writing – review & editing. Landon Gryczkowski: Resources, Writing – review & editing. Sherri L. Johnson: Resources, Writing – review & editing. Kelly Elder: Resources, Writing – review & editing. Anna M. Jalowska: Writing – review & editing. Peter Caldwell: Resources, Writing – review & editing. Johnny M. Grace: Writing – review & editing. Dariusz Młynski: Writing – review & editing. Andrzej Wałęga: Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This study was supported by a funding agreement (#18AI11330155072) between the USDA Forest Service and the Oak Ridge Institute for Science and Education (ORISE). We are thankful for the streamflow data provided by USDAFS researchers, Dr. Nina Lany at the Hubbard Brook EF, Dr. Chris Oishi at the Coweeta Hydrology Laboratory, and Charles A. Harrison at the Santee EF, and Banning Starr and Laurie Poth at the Fraser Experimental Forest. We thank technicians, Ian Halm, Hannah Vollmer and Gabe Winant for providing the culvert dataset for the Hubbard Brook EF. The USGS provided streamflow data and NOAA provided hourly precipitation data and storm design precipitation intensities. We would like to gratefully acknowledge the help of Shawna Reid at the USDAFS Southern Research Station, for acquiring the geospatial dataset of watersheds and related information about road-stream crossings and road networks located across the EFs in this study. We also thank Karl Buchholz at the Francis Marion National Forest, SC, and Stephanie Laseter and Johnny Boggs, both at the USDAFS Southern Research Station, for their support. The opinions presented in this article are those of the authors and should not be construed to represent any official USDA or US Government determination or policy. Similarly, the views expressed in this article are those of the authors and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2024.130698.


