An Uncertainty Quantification Framework for Simulation-based Flood Frequency Analysis

Jonathan Romero-Cuellar[1], James R. Craig[1], Bryan A. Tolson[1], and Rezgar Arabzadeh[1]

[1]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, N2L 3G1, Canada

Correspondence to: Jonathan Romero-Cuellar (jromeroc@uwaterloo.ca)

Abstract. Flood frequency analysis (FFA) is essential for flood risk management and infrastructure design, yet flood quantile estimates remain highly uncertain when based on limited streamflow records. In data-scarce settings, conventional statistical FFA is strongly constrained by sampling uncertainty: short annual maximum flow records often yield broad credible intervals for high-return-period floods, making the resulting estimates difficult to use in practical flood-risk decisions. To address this limitation, we introduce a simulation-based uncertainty quantification framework—UQ-flood—that integrates a process-based hydrologic model, a stochastic weather generator, and a residual error model (REM). Designed for annual maxima, the REM accounts for model bias and residual variability, enabling the generation of probabilistic streamflow ensembles tailored to extreme event analysis. We apply UQ-flood to three Canadian watersheds with long streamflow records and contrasting hydroclimatic conditions. We compare its performance against traditional statistical FFA using the Generalized Extreme Value distribution and Bayesian inference. UQ-flood yields flood quantile estimates consistent with long-record statistical methods but with substantially narrower uncertainty bounds. Under short-record conditions (e.g., 30 years), UQ-flood produced quantile estimates that remained close to long-record empirical/statistical benchmarks, whereas statistical FFA produced much wider intervals. Additional experiments reveal that omitting the REM introduces systematic bias in flood magnitude estimates. UQ-flood avoids parametric assumptions about flow distributions, circumvents hydrologic model biases, and is adaptable to data-limited conditions. By explicitly propagating uncertainty from hydrologic simulation to flood quantiles, UQ-flood offers a practical alternative for robust flood risk management, including applications in infrastructure design and floodplain mapping. We recommend integrating residual error models into continuous simulation frameworks to improve bias correction and uncertainty quantification in flood risk estimation.

1 Introduction

Floods are among the most prevalent natural disasters, with both their intensity and frequency increasing due to global change (Milly et al., 2002). In this context, flood frequency analysis (FFA) plays a fundamental role in engineering hydrology, as it is essential for accurately estimating flood risks. These estimations are critical for designing flood protection infrastructure, planning land use, ensuring dam and industrial safety, conducting risk assessments, and making informed water resources management decisions. Industry-standard, at-site FFA typically involves fitting a statistical distribution to a

sample of observed flood data, often based on annual maximum flows or peaks over threshold. These distribution models estimate the probability of exceedance for extreme flood peaks and are used to extrapolate high flood quantiles beyond observed data, often following regulatory guidelines (Hamed and Ramachandra Rao, 2019).

Despite its central role in engineering practice, uncertainty quantification in FFA is often neglected, insufficiently characterised, or not transparently communicated (Hall and Solomatine, 2008; Merz and Thieken, 2005, 2009; Vidrio-Sahagún et al., 2025). . This is critical, as flood frequency estimates are affected by multiple sources of uncertainty—including sampling uncertainty, model structure uncertainty, and parameter uncertainty. Sampling uncertainty arises from the limited length of observed discharge records, which makes it difficult to reliably characterise the tail of the distribution, especially for rare events like the 100-year flood (Lucas et al., 2023). Structural uncertainty stems from the subjective selection of probability distributions. Goodness-of-fit tests often lack discriminatory power with small samples, allowing different distributions to fit the data equally well while producing vastly different flood quantiles (Cunnane, 1985; Laio et al., 2009). Parameter uncertainty further contributes to the problem, with flood estimates varying substantially depending on the parameter estimation method used (Hu et al., 2020).

Statistical FFA based on limited streamflow records, for example fewer than 30 years of annual maximum flows, can yield highly uncertain and often impractical estimates for rare events. In such cases, uncertainty intervals may be statistically defensible but so wide that they provide limited actionable guidance for infrastructure design, floodplain mapping, or emergency planning. This limitation highlights a key gap: the need for accessible and reliable uncertainty quantification approaches that remain useful under data-scarce conditions (Pappenberger and Beven, 2006; Kasiviswanathan et al., 2017)

Several strategies have been developed to reduce the limitations imposed by short flood records. A first option is to extend the flood record in time by incorporating historical information, documentary evidence, or paleoflood indicators (Benito et al., 2004). These approaches can improve the estimation of rare flood quantiles by adding information about events that occurred before systematic gauging began. However, they also require careful treatment of additional uncertainty sources, including event dating, magnitude reconstruction, threshold interpretation, and differences between historical, paleoflood, and instrumental data quality. Thus, while historical and paleoflood information can increase the effective sample size, it can also introduce new epistemic uncertainties that must be explicitly characterised.

A second option is to borrow information across space through regional flood frequency analysis, where data from hydrologically similar catchments are pooled to increase the effective flood sample (Cunnane, 1988). Regional approaches can reduce sampling uncertainty relative to single-site analysis, particularly in basins with short records. However, their reliability depends on the assumption that the selected catchments form a homogeneous region. Defining such regions is partly subjective and depends on choices regarding similarity criteria, catchment descriptors, regional boundaries, and

pooling rules. As a result, different regionalization decisions can lead to different pooled samples and, consequently, different flood quantile estimates.

A third option is to generate larger flood samples using continuous simulation, in which stochastic weather generators or other meteorological resampling approaches are coupled with hydrologic models to produce long synthetic streamflow records (Eagleson, 1972; Franchini et al., 1996). This approach is attractive because it can represent physical controls on flood generation, including interactions among rainfall, snowpack, soil moisture, antecedent storage, and runoff production, while reducing sampling uncertainty through large synthetic samples. Previous research has established the concept and potential of model-based, or continuous simulation, FFA frameworks (Arnaud and Lavabre, 2002; Blazkova and Beven, 2002; Eagleson, 1972; Franchini et al., 1996; Sivapalan et al., 2005; Paquet et al., 2013). In these frameworks, process-based hydrologic models are driven by synthetic meteorological inputs to generate long discharge time series for flood frequency estimation (Blazkova and Beven, 2009). Continuous simulation has been successfully applied to derive flood frequency distributions in a range of hydrologic settings (Haberlandt & Radtke (2014), Yu et al. (2019), Perez et al. (2019)).

Model-based approaches are particularly well-suited for this task because they capture complex interactions among flood drivers—such as rainfall, snowpack, and soil moisture—and their influence on runoff production (Sivapalan and Samuel, 2009). In contrast to design storm methods, model-based FFA avoids the assumption of equal return periods for rainfall and discharge. It also enables the integration of land use change (e.g., urbanization), management interventions, and future climate scenarios (Lamb et al., 2016). Another key advantage of this approach is its ability to generate large synthetic samples, which reduces sampling uncertainty—a common limitation in statistical methods relying on short observational records. Common methods to generate large prediction samples for this framework include stochastic weather generators using multivariate non-Gaussian time series (e.g., Yu et al. (2023)), stochastic storm transpositions (Andersen et al., 2024; Wright et al., 2020), and reforecast ensemble pooling (Brunner and Slater, 2022).

However, continuous simulation does not automatically solve the uncertainty problem in FFA. A key limitation is that hydrologic models are commonly calibrated to reproduce daily streamflow dynamics or general hydrograph behaviour, whereas FFA depends on accurately representing annual maximum flows. This creates a mismatch between the calibration target and the design variable of interest. A model that performs well for daily streamflow may still exhibit systematic bias or inadequate variability in simulated flood peaks (Cameron et al., 2001; Brunner et al., 2021; Grimaldi et al., 2019; Spellman et al., 2018). If this model error is ignored, continuous simulation can produce biased flood frequency estimates, even when very long synthetic streamflow records are generated.

Several methodological frameworks have been proposed to quantify the uncertainty of flood frequency estimates, yet none have achieved widespread adoption in practice. One class of approaches combines continuous hydrologic model simulation with stochastic rainfall generation and rainfall–runoff modelling, often using methods like Generalized Likelihood

Uncertainty Estimation (GLUE) to quantify uncertainty (Cameron et al., 2001). While GLUE is useful for exploring parameter and model structure uncertainty, it has been criticized for relying on subjective likelihood measures and for its tendency to overestimate uncertainty (Stedinger et al., 2008). Formal Bayesian inference methods offer a more robust probabilistic treatment of uncertainty, but they are often computationally demanding and require careful specification of prior distributions, limiting their accessibility for routine applications (Costa and Fernandes, 2017). Other approaches, including resampling methods, hierarchical Monte Carlo frameworks, variance decomposition, and ensemble pooling, have also been used to reduce or attribute uncertainty in flood estimates (Apel et al., 2004; Arnaud et al., 2017; Bomers et al., 2019; Burn, 2003; Mirzaei et al., 2015; Newman et al., 2021; Brunner and Slater, 2022). Despite this diversity of methods, many remain constrained by subjectivity, computational complexity, high data requirements, or limited treatment of hydrologic model error at the flood-peak scale. A more accessible, process-consistent framework for uncertainty quantification in FFA—particularly one that performs well under limited data availability—therefore remains a pressing need.

Given these limitations, simulation-based FFA requires an explicit treatment of hydrologic model error at the scale most relevant for design floods: the annual maximum flow. A key limitation of model-based flood estimation frameworks is the frequent neglect of model error and uncertainty in hydrologic predictions. To address this, we apply the Residual Error Model (REM) approach—a strategy designed to quantify and propagate residual uncertainty by embedding error model samples into calibrated deterministic simulations (Vogel, 2017). REM transforms deterministic outputs from process-based hydrological models into streamflow ensembles, offering a robust means of uncertainty quantification that is particularly valuable for hydrologic risk management. The REM framework consists of two key components:

1. A deterministic component, which represents changes in meteorological forcings, land use, and their non-linear interactions.
2. A stochastic component, which accounts for hydrologic uncertainty not captured by the deterministic model alone (Shabestanipour et al., 2023).

This dual structure enables REM to estimate predictive uncertainty—integrating data, parametric, and structural uncertainties—by comparing model simulations with observed streamflow records (Montanari and Koutsoyiannis, 2012). Various error modelling approaches have been proposed to represent this uncertainty (Jiang et al., 2019; Koutsoyiannis and Montanari, 2022; Li et al., 2016; McInerney et al., 2017, 2020; Shabestanipour et al., 2023; Todini, 2008).

In this study, we apply a REM that is additive in the transformed space, using the Box-Cox transformation, a first-order autoregressive (AR(1)) model, Gaussian innovations, and a flow-dependent error mean—following the formulation in Hunter et al., (2021). This error model is particularly attractive due to its ability to consolidate multiple sources of uncertainty into a single error term, enabling transparent, efficient, and straightforward generation of streamflow predictive distributions (Evin et al., 2014; Romero-Cuellar et al., 2024).

Here, we present UQ-flood, an uncertainty quantification framework for simulation-based flood frequency analysis that integrates a process-based hydrologic model, a stochastic weather generator, and a residual error model developed at the annual maxima scale. The framework is designed to generate large synthetic flood samples while correcting systematic biases and propagating residual uncertainty in simulated flood peaks. By modelling the relationship between observed and simulated annual maximum flows directly, UQ-flood addresses the mismatch between daily streamflow calibration and flood-quantile estimation. This allows uncertainty from hydrologic simulation to be propagated into flood frequency curves and associated quantile estimates. The novelty of this study lies not in continuous simulation alone, but in applying a residual error model directly at the annual-maximum scale and embedding it within a stochastic simulation framework for at-site FFA. This differs from daily-flow post-processing and from parameter-uncertainty approaches because it targets the design variable itself: annual maximum discharge.

Unlike traditional statistical FFA methods that rely solely on observed flood samples and assumed parametric distributions, UQ-flood is informed by hydrologic process representation and by the full information contained in the available meteorological and discharge records. To our knowledge, this is the first study to apply the REM approach within the context of at-site FFA for uncertainty quantification. In contrast to conventional approaches that can produce very wide uncertainty intervals under short-record conditions, UQ-flood offers a process-consistent and practical alternative for data-scarce environments.

2 Study Watersheds and Data

To evaluate the proposed UQ-flood framework for at-site application across diverse hydrologic conditions, we selected three Canadian case study watersheds based on two main criteria: (1) the availability of long daily streamflow records—each with more than 80 years of data—to enable validation of flood frequency estimates, and (2) minimal anthropogenic regulation, as the hydrologic models used in this study do not simulate human interventions such as reservoir operations or flow diversions. In addition to these criteria, we deliberately chose three watersheds that represent contrasting physiographic and climatic regions. The Lahave River in Nova Scotia was selected to represent coastal catchments influenced by maritime climate conditions; the Moyie River in British Columbia serves as a representative mountain watershed with a snowmelt-dominated regime; and the Black River in Ontario captures characteristics of the Great Lakes region, including wetland and lake-dominated hydrology.

2.1 Black River Watershed

The Black River near Washago is located in central Ontario, Canada, and drains a watershed of approximately 1,506 km^2 upstream of the Water Survey of Canada gauge 02EC002, which has over 100 years of streamflow records. The watershed is

characterised by a humid continental climate, with cold, snowy winters and warm, humid summers. The climate is strongly influenced by the Great Lakes, which moderate seasonal temperatures and enhance precipitation (Whitfield and Cannon, 2000). Hydrologically, the Black River is classified as a mixed-regime system, with a dominant nival (snowmelt) component; the nival fraction is approximately 0.7 (Burn and Whitfield, 2023). The primary flood-generating mechanisms are spring snowmelt and rain-on-snow events, which have historically produced the highest flows (Buttle et al., 2016). The watershed contains a high proportion of wetlands, which play a key role in sustaining groundwater recharge, and maintaining baseflow during dry periods.

2.2 Moyie River Watershed

The Moyie River is in southeastern British Columbia within the South BC Mountains region and drains an area of approximately 1,480 km² upstream of the Water Survey of Canada gauge 08NH006, which provides a 90-year observational record. The watershed lies within a complex mountainous terrain characterised by steep slopes, dense forest cover, and variable elevations. The climate is strongly influenced by elevation and orographic effects, resulting in a continental to subalpine climate with distinct seasonal patterns. Winters are cold with substantial snowfall, while summers are warm and relatively dry (Whitfield and Cannon, 2000). The hydrologic regime is snowmelt-dominated, with peak flows typically occurring in late spring to early summer (May–June). Rain-on-snow events in late winter or early spring also contribute to flood risk, particularly when rapid warming occurs on an established snowpack (Buttle et al., 2016).

2.3 Lahave River Watershed

The Lahave River at West Northfield is in Nova Scotia and drains an area of approximately 1,250 km² upstream of the Water Survey of Canada gauge 01EF001, providing 108 years of continuous daily flow records. The watershed is situated within a coastal, lowland physiographic region and is characterised by rolling hills, mixed forest, and numerous lakes and wetlands. The region experiences a humid maritime climate, with relatively mild winters, cool summers, and high precipitation distributed evenly throughout the year. This climate is strongly influenced by proximity to the Atlantic Ocean, which moderates seasonal temperature fluctuations and contributes to frequent rainfall. The area is also regularly affected by storms, including tropical cyclones and hurricanes, due to its exposure to Atlantic storm tracks (Whitfield and Cannon, 2000). Hydrologically, the Lahave River exhibits a rainfall-dominated regime, with flood events occurring year-round, but most frequently in the fall and winter when intense rainfall coincides with saturated soil conditions. Rain-on-snow events may also contribute to elevated flows during transitional seasons; however, snow accumulation is generally lower than in inland or higher-elevation watersheds (Buttle et al., 2016).

During a review of the annual maximum flow series, we identified an extreme event in 1956 that appeared as a statistical outlier in the historical record. Further investigation using provincial archives confirmed that this event was caused by an ice-jam flood, a process not represented in the HBV-EC model used here. Because the hydrologic model does not simulate ice-related hydraulic dynamics, inclusion of this event would have introduced bias into the flood frequency estimates.

Consequently, the 1956 event was excluded from model calibration and analysis to preserve the consistency and physical relevance of the simulation-based framework.

Figure 1 shows the location of three study watersheds.

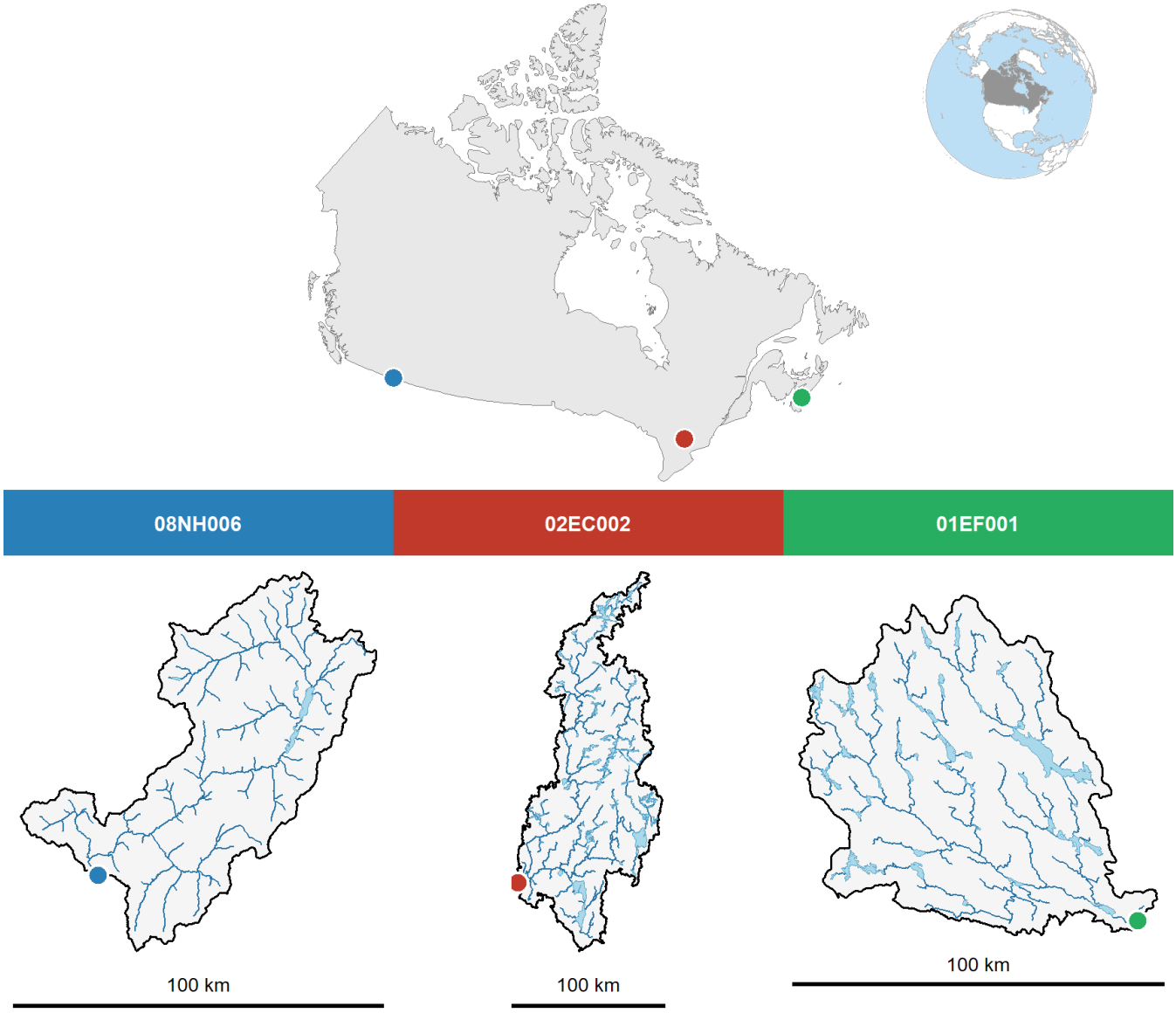


Figure 1: Geographic location of the three hydrological case studies analysed in this study. The figure shows the spatial distribution and watershed boundaries of the selected study areas.

3 Methodology

The UQ-flood framework proposed in this study integrates three key components: a process-based hydrological model, a residual error model (REM), and a stochastic weather generator (SWG). These components are used in conjunction to quantify the model uncertainty in flood estimates. The framework consists of two main phases: model setup and parameter estimation and long-term continuous simulation, as schematically illustrated in Fig. 2. The key steps are outlined as follows:

Phase 1: Model Setup and Parameter Estimation

1. Build a process-based hydrological model. The first step involves developing a hydrological model capable of simulating river discharge based on the physical processes governing the watershed. This model serves as the backbone for simulating runoff and streamflow under given climatic conditions.
2. Develop a residual error model (REM) based on annual maxima. The residual error model is developed to characterise and quantify uncertainties in hydrological predictions, specifically for extreme events. It is constructed using the annual maximum flow series, by analysing the differences between observed and simulated annual maxima. This approach captures the statistical properties of model errors—including their magnitude and variability at the scale most relevant for flood frequency analysis.
3. Set up a stochastic weather generator (SWG). The SWG is configured to generate synthetic weather data, including precipitation and maximum and minimum temperature, based on historical observations. This tool accounts for the variability in climatic conditions by preserving key statistical properties of precipitation and temperature such as marginal distributions, autocorrelation structures, and intermittency patterns.

Phase 2: Long-term Continuous Simulation

4. Generate long-term forcings using the SWG, which produces synthetic time series of daily meteorological variables (e.g., precipitation and max/min daily temperature) for an extended period, spanning greater than 1,000 years. This step ensures sufficient data for a robust non-parametric flood frequency analysis.
5. Simulate long-term continuous discharges using the hydrological model. The synthetic weather data generated by the SWG is used to drive the process-based hydrological model, simulating long-term streamflow for the watershed of interest.
6. Generate corresponding error realisations using the REM. The REM is applied to introduce stochastic error realisations into the simulated streamflow. This step adds probabilistic uncertainty to the hydrological model predictions, reflecting the inherent errors in model output.
7. Apply flood estimate uncertainty analysis. The resulting synthetic annual maxima, enhanced by uncertainty bounds, are ranked and used to compute flood frequency without assumptions or need of a specific parametric distribution. The annual exceedance probabilities are calculated using the Weibull plotting position formula, and the multi-realisation flood frequency curves illustrate the relationship between flood magnitude and return period, along with the associated uncertainties.

Figure 2 provides a schematic representation of the UQ-flood framework. The figure is divided into two phases: the Model Setup and Parameter Estimation phase (Steps 1–3) and the Long-Term Continuous Simulation phase (Steps 4–7). These steps collectively allow the computation of low-exceedance-probability flow magnitudes that explicitly account for weather variability and uncertainty in hydrological modelling. This workflow ensures that uncertainties at all stages—model predictions, weather variability, and statistical error—are propagated into the final flood frequency analysis.

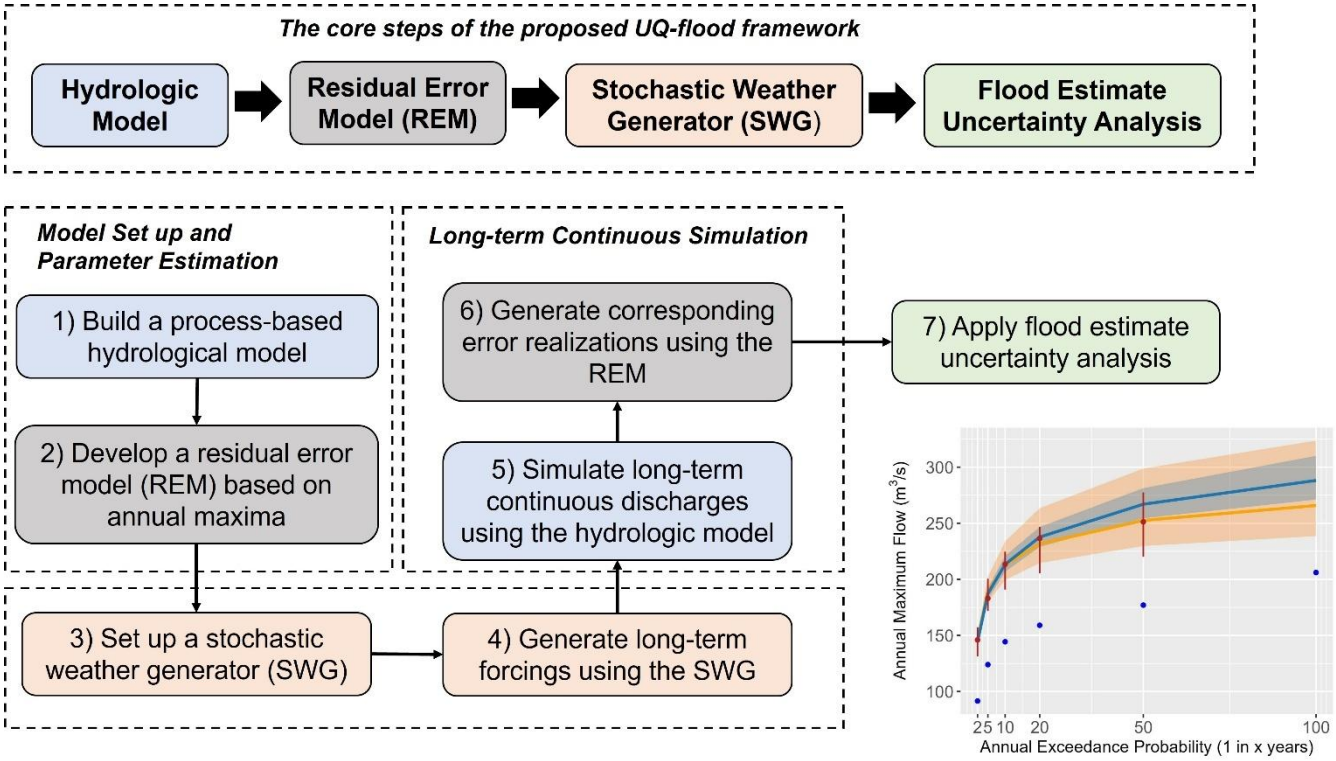


Figure 2: Conceptual diagram of the proposed UQ-flood (Residual Error Model - Flood Frequency Analysis) framework.

3.1 Hydrologic Model Development

The Hydrologiska Byråns Vattenbalansavdelning – Environment Canada (HBV-EC) model (Hamilton et al., 2000) was used to simulate watershed streamflow. A lumped representation of HBV-EC was implemented within the Raven Hydrological Modelling Framework version 3.8 (Craig et al., 2020). The HBV-EC model conceptualises catchments as a soil moisture zone atop two stacked buckets, representing upper and lower storage layers. Its structure includes three primary routines: a snow routine, soil routine, and runoff generation routine (Hamilton et al., 2000). The snow routine uses a classic degree-day approach to estimate snowmelt rates. The model partitions snowmelt and precipitation into evapotranspiration and runoff based on water and energy balance principles. Runoff is divided into surface and subsurface flow, which is further partitioned between the upper and lower buckets according to soil moisture dynamics (Seibert and Vis, 2012). A lumped

version of HBV-EC was used to simulate the average temporal dynamic of catchment behaviour without requiring detailed spatial information.

HBV-EC was selected due to its demonstrated strong performance in Canadian landscape (Beck et al., 2016; Stahl et al., 2008; Zégre et al., 2010), its widespread use for flood simulations (Fischer and Schumann, 2021; Sikorska et al., 2015), its parsimony, ease of implementation, and its computational efficiency (Seibert and Bergström, 2022). The model was run using daily time series of measured discharge, precipitation, maximum and minimum temperature, which were sourced from Environment Canada. The HBV-EC models were calibrated to a truncated discharge dataset using a combination of manual and automated approaches. Manual calibration followed a stepwise process: (1) parameters controlling the long-term water balance were adjusted to achieve volumetric flow simulations; (2) parameters governing snowpack process were calibrated to accurately simulate the freshet; (3) parameters influencing soil moisture dynamics were refined to improve recession and baseflow representation. Once an acceptable parameter set was identified through manual calibration, the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2008) was employed for further optimisation. The objective function minimised the daily discharge Kling-Gupta Efficiency (Gupta et al., 2009) often the last 30 years to ensure the best fit between simulated and observed hydrographs. The final set of parameters was evaluated against the long-term water balance, hydrographs, flow duration curves, and peak flows. Across all experiments, HBV calibration used only a 30-year daily streamflow subset. The same calibrated parameters were used when comparing 30-year vs. 108-year estimation settings.

3.2 Residual Error Modelling

The REM approach is a tool for propagating uncertainty into hydrological model predictions. It achieves this by incorporating samples of model error into deterministic predictions (Farmer and Vogel, 2016). REM combines deterministic predictions from process-based hydrologic models with stochastic components that characterise the uncertainties in model predictions (Montanari and Koutsoyiannis, 2012). This integration is accomplished by leveraging the statistical properties of the observed residual error, defined as the difference between observed and simulated streamflow (Vogel, 2017). By effectively characterising and modelling these residuals, the REM approach enables the generation of probabilistic streamflow estimates, accounting for uncertainties inherent in deterministic model outputs (Honti et al., 2013). REM characterises residual uncertainty as the discrepancy between simulations and observations. This residual term implicitly reflects the combined effects of forcing uncertainty, parameter uncertainty, and structural model error, without separating their individual contributions. This formulation is consistent with established error-modelling frameworks in hydrology (Li et al., 2016; McInerney et al., 2017).

In this study, the residual error model is developed for the annual maxima, aligning with the requirements of flood frequency analysis. While much of the literature has focused on REM applications at the daily streamflow—where model errors are often non-normal, heteroscedastic, and autocorrelated —our approach focuses on modelling these properties for annual

maxima. The statistical properties of annual maxima residuals require careful treatment to avoid biased estimates of extremes.

The selection of the REM structure was guided by diagnostic analyses of the statistical properties of the raw residuals, including variance behaviour, temporal dependence, and distributional characteristics. To address the heteroscedasticity—where errors tend to increase with larger streamflow—we apply a Box-Cox transformation (Box and Cox, 1964) with a power parameter fixed at 0.2, following the recommendations of McInerney et al. (2017). To account for autocorrelation in the error time series—where errors exhibit similar sign and magnitude over consecutive time steps—we fit a first-order autoregressive AR(1) model (Evin et al., 2014; McInerney et al., 2017). The innovation, or random component of the REM, follows a Gaussian distribution with a flow-dependent residual mean (Li et al., 2016). The parameters of REM were inferred using the method-of-moments (Hazelton, 2011). For the detailed REM formulation, including the model equations, readers are referred to Hunter et al. (2021) and Romero-Cuellar et al. (2024).

3.3 Stochastic Weather Generator (SWG)

In this study, we employed CoSMoS (Complete Stochastic Modelling Solution) (Papalexiou, 2018) as the stochastic weather generator (SWG). CoSMoS is an R package designed for the stochastic simulation of univariate, multivariate, or random field non-Gaussian time series. It implements a flexible framework that unified, extends, and improves the theory of stochastic modelling for hydroclimatic process, preserving key statistical characteristics such as marginal distributions, autocorrelation structures, and intermittency (Papalexiou, 2018). The core assumption underlying CoSMoS is that any process can be generated by transforming a parent Gaussian process.

For this study, we inferred the parameters of the generalized gamma distribution, autocorrelation structure, and dry-period distribution of daily precipitation using 30 years of gauge-based observations, with statistics estimated separately for each calendar month. Similarly, for maximum and minimum temperatures, we estimated the monthly marginal distributions and autocorrelation structures. Using the inferred parameters, CoSMoS was used to generate long-term (e.g., 10,000 synthetic years) daily time series of precipitation, maximum temperature, and minimum temperature. These synthetic meteorological forcings were then used to drive the hydrologic model, enabling the simulation of annual maximum flows and the estimation of flood frequency distributions.

To adapt CoSMoS for flood frequency purposes, we implemented two key modifications: (1) a bootstrap-based smoothing strategy across monthly boundaries to address abrupt discontinuities in statistical properties, and (2) the introduction of physical constraints on extreme precipitation to prevent unrealistically long-tailed behaviour in the generalized gamma distribution. These constraints help avoid infinite or implausible values in simulated extremes. Full details of these modifications are provided in the Supplements.

3.4 Simulation-based Flood Frequency Analysis

Following the calibration of hydrologic model parameters, inference of residual error model (REM) parameters, and fitting of the stochastic weather generator (SWG), the methodology proceeds with simulation-based flood frequency analysis. For each realisation of the REM, a synthetic sequence of 10,000 years of annual maximum streamflow is generated using climate inputs from the SWG and the calibrated hydrologic model. This yields one estimate of the flood quantile distribution. The process is repeated 1,000 times, each with a different REM realisation, resulting in 1,000 samples of flood quantile estimates. These samples are then used to construct multi-realisation flood frequency curves, illustrating the relationship between flood magnitude and annual exceedance probability, along with associated uncertainty intervals. Return periods are computed using the Weibull plotting position formula.

3.5 At-site Statistical Flood Frequency Analysis

To provide a benchmark for evaluating the proposed simulation-based approach, we applied an at-site statistical flood frequency analysis (FFA) using the Generalized Extreme Value (GEV) distribution (Jenkinson, 1955). The GEV distribution is the standard model for fitting annual maximum flow (AMF) data and is widely used in hydrologic design due to its foundation in extreme value theory and its ability to represent light, moderate, and heavy tails (Coles, 2001; Vogel and Wilson, 1996; Zaghloul et al., 2020).

In this study, we fit the GEV distribution to observed annual maxima using a Bayesian inference framework to quantify parameter uncertainty and generate credible intervals for flood quantiles (Parkes and Demeritt, 2016). This approach offers a probabilistic characterisation of uncertainty, in contrast to classical methods such as maximum likelihood estimation (Liang et al., 2012). Specifically, we implemented Markov Chain Monte Carlo (MCMC) sampling to draw from the joint posterior distribution of the GEV parameters (location, scale, and shape). Posterior mean estimates were used for point estimation, whereas the posterior predictive distributions were used to estimate flood quantiles and their 95% credible intervals.

The analysis was performed using the extRemes R package (Gilleland and Katz, 2016), which provides a Bayesian GEV fitting module with built-in MCMC capabilities. Convergence diagnostics—including trace plots, effective sample size, and Gelman–Rubin statistics—were used to assess sampling reliability and ensure robustness of the posterior estimates.

This statistical approach serves as a benchmark method against which the performance of the REM-based flood frequency estimates was compared, particularly in terms of uncertainty characterisation.

We compare simulated to observed flood quantiles corresponding to return periods of $T = 2,5,10,20,50,100$ years by computing the uncertainty range (UR), defined as the difference between the 2.5th and 97.5th percentiles of the ensemble outcomes (q). The uncertainty range is calculated at each computational time step t in the hydrographs, as expressed in the following Eq. (1):

$$UR_t = q_t^{97.5} - q_t^{2.5}, t = 1, \dots, T \quad (1)$$

3.6 Confidence Interval for Discharge Quantiles

This study quantifies uncertainty in discharge quantiles associated with specified annual exceedance probabilities (AEPs), rather than uncertainty in plotting positions or order statistics. The focus is therefore on vertical uncertainty—i.e., uncertainty in discharge for a fixed return period—which is the quantity most relevant for flood-risk assessment and engineering design. To estimate confidence intervals for discharge quantiles, we adopt an analytical approach based on fitting parametric probability distributions to annual maximum flow (AMF) data. Several candidate distributions (e.g., Weibull, exponential, gamma, and lognormal) are fitted using maximum likelihood estimation. Model selection is guided by a combination of visual diagnostics (density plots, Q–Q plots, and cumulative distribution function comparisons) and statistical criteria, including the Akaike Information Criterion (AIC) and goodness-of-fit statistics. The selected distribution provides the basis for estimating uncertainty in discharge quantiles corresponding to target AEPs.

For a given AEP, the corresponding non-exceedance probability is defined as:

$$p = \left(1 - \frac{1}{AEP}\right)(N + 1) \quad (2)$$

Where N is the sample size and AEP is expressed in years. The uncertainty in this probability estimate due to finite sample size is modelled using the beta distribution, consistent with the analytical framework of Baâ et al. (2001).. The lower and upper bounds of the 95% confidence interval are computed as:

$$x_{0.05} = qbeta(0.025, p, N - p + 1) \quad (3)$$

$$x_{0.95} = qbeta(0.975, p, N - p + 1)$$

These bounds are then transformed into flow quantiles using the cumulative distribution function (CDF) of the selected parametric model:

$$Q_{0.05} = F^{-1}(x_{0.05}) \quad (4)$$

$$Q_{0.95} = F^{-1}(x_{0.95})$$

where F^{-1} is the inverse CDF (quantile function) of the fitted distribution. This procedure yields confidence intervals for discharge quantiles that explicitly reflect sampling uncertainty conditional on the assumed statistical model. It provides a transparent, closed-form alternative to resampling-based methods and is well suited for comparative analyses across multiple basins. We emphasize that this approach differs conceptually from methods that quantify uncertainty in exceedance probabilities or order statistics (horizontal uncertainty), such as fractional order statistics (e.g., Serinaldi, 2009). While such approaches are valuable—particularly for estimating uncertainty beyond the observed sample maximum—the present study prioritizes uncertainty in discharge for fixed design return periods.

3.7 Numerical Experiment Configuration

To evaluate the performance and added value of the proposed simulation-based flood frequency analysis (UQ-flood), we designed a numerical experiment that compares it against a benchmark statistical FFA approach based on the Generalized

Extreme Value (GEV) distribution, as described in Sect. 3.5. The objective of this experiment is to assess the ability of each method to produce robust flood quantile estimates and quantify associated uncertainty under different data availability conditions and modelling assumptions.

The evaluation consists of two main components:

1. Comparison under Varying Record Lengths:

We compared the UQ-flood method—applied using only 30 years of observed meteorological and hydrological data—to the statistical GEV-based FFA under two observational record lengths:

- A long-record scenario using the full historical record (≥100 years), and
- A short-record scenario using the same 30 years of observed annual maximum flows.

This setup allows us to isolate and evaluate the impact of sampling uncertainty on statistical FFA, while keeping the simulation-based method fixed in its data requirements. By comparing these scenarios, we assess whether the UQ-flood can provide reliable flood estimates even when observational records are limited, where statistical methods tend to suffer from high uncertainty.

2. Evaluation of the Continuous Simulation Approach Without Error Modelling:

To evaluate the role of the residual error model (REM) in improving simulation-based flood estimates, we conducted an additional experiment in which the continuous simulation method was applied without the error model correction. In this case, flood frequency estimates were derived directly from the annual maximum flows produced by the hydrologic model driven by stochastic meteorological inputs (referred to as AMF_{sim}). This configuration helps assess the systematic bias and underestimation issues that may arise when simulation-based FFA is used without explicitly accounting for model errors.

Together, these experiments provide a comprehensive assessment of the REM framework's performance relative to traditional methods and explore the consequences of omitting explicit error modelling in continuous simulation-based FFA.

4 Results

This section presents three case studies to evaluate the performance of the proposed UQ-flood framework for flood estimates. Section 4.1 provides an in-depth assessment of the Black River watershed, where we compare UQ-flood with the traditional statistical approach based on the Generalized Extreme Value (GEV) distribution using both long and short observational records. To assess the generalisability of our findings across different hydrological conditions, Section 4.2 applies the same analysis to two additional watersheds—Moyie River (BC) and Lahave River (NS)—focusing on the consistency of the results rather than a detailed breakdown. Together, these case studies demonstrate the utility of the UQ-flood framework under diverse conditions.

The uncertainty intervals compared here do not represent identical uncertainty decompositions. The GEV intervals represent parameter/posterior uncertainty conditional on the selected statistical model and observed annual maxima sample, whereas UQ-flood intervals represent predictive uncertainty induced by residual-error realisations conditional on the calibrated

hydrologic model and fitted weather generator. Therefore, comparisons of interval width are interpreted in terms of decision informativeness and consistency with long-record benchmarks, not as evidence that one method captures all uncertainty more completely.

4.1 Flood Frequency Estimates in the Black River Watershed

4.1.1 UQ-flood Compared to At-site Statistical FFA

The comparison of flood frequency estimates for the Black River watershed in Fig. 3 highlights notable differences between the traditional statistical approach using the GEV distribution and the proposed UQ-flood. The UQ-flood framework, which relies on only 30 years of precipitation and streamflow records, produces flood quantile estimates that align more closely with the observed annual maximum flows and their associated confidence interval for discharge quantiles, particularly for high return periods such as the 50- and 100-year floods. In contrast, the GEV-based estimates—derived from the full 108-year record—tend to underestimate extreme flows relative to the observational bounds (Fig. 3, top panel).

A key advantage of the UQ-flood framework lies in its ability to produce more constrained and informative uncertainty estimates. Across the full range of annual exceedance probabilities, UQ-flood yields narrower uncertainty bands compared to the GEV approach, demonstrating improved precision (Fig. 3, top panel). Moreover, the simulated annual maximum flows generated by UQ-flood closely follow the central tendency of the 108-years observed flood distribution, despite being calibrated with only 30 years of data. This close agreement indicates that the simulation-based approach, when corrected using an error model, effectively captures both the magnitude and variability of extreme hydrologic events.

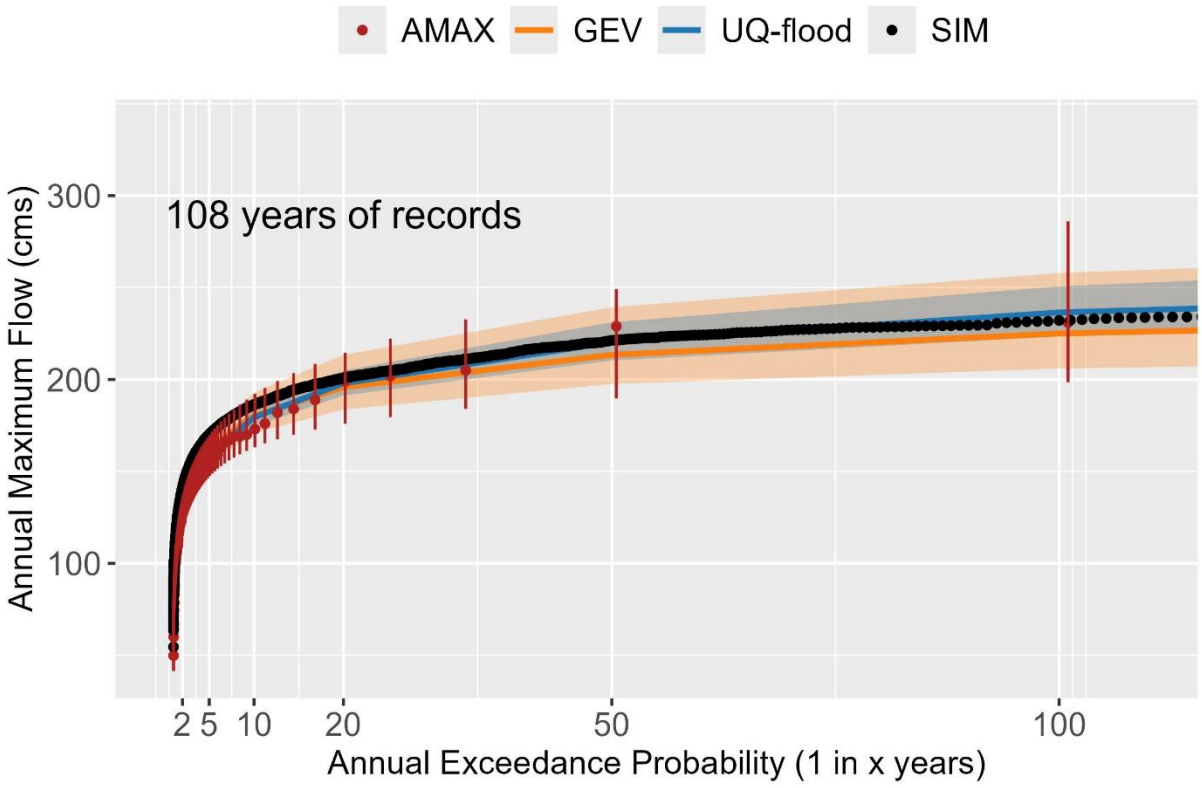
The bottom panel of Fig. 3 presents a critical and commonly encountered scenario in hydrology: applying at-site statistical FFA to short observational records (e.g., 30 years). In this case, the GEV-based estimates exhibit wide and often impractical uncertainty intervals. In contrast, the UQ-flood framework continues to provide narrow flood quantile estimates, despite relying on the same short record.

The differences in uncertainty ranges between the GEV and UQ-flood methods vary across return periods (Fig. 3). When using long observational records, the average difference in the width of the uncertainty intervals between GEV and UQ-flood is approximately 14 m^3s^{-1}. This difference increases substantially under short-record conditions, with the GEV method exhibiting an average uncertainty width approximately 198 m^3s^{-1} greater than that of UQ-flood. The orange GEV uncertainty envelope is considerably larger for the 30-year record than for the 108-year record, widening dramatically at high return periods and reaching values several times larger than the estimated quantiles, which illustrates the profound sensitivity of GEV-based flood frequency estimates to limited data availability (Figure 3). The resulting uncertainty ranges undermine the reliability of statistical estimates, thereby limiting their usefulness for practical decision-making.

For example, consider the 100-year flood estimate for the Black River watershed. The observed 100-year flood is approximately 230 m^3s^{-1}. When applying the GEV distribution to short records, the estimated 100-year flood was 322 m^3s^{-1}, with a wide uncertainty range of 531 m^3s^{-1} —offering limited practical utility. In contrast, the UQ-flood framework

estimated the 100-year flood as 236 m^3s^{-1}, with a much narrower uncertainty range of just 27 m^3s^{-1}. This result demonstrates the UQ-flood framework's ability to produce estimates that are not only closer to observed values but also more constrained, providing more reliable and actionable information for infrastructure design and flood risk planning—even when using only 30 years of data from a single site.

A critical consideration in interpreting flood frequency estimates is the treatment of uncertainty in confidence interval for discharge quantiles. Unlike traditional confidence intervals, which assess the proportion of observed data points falling within a given range, plotting positions are themselves derived from estimated exceedance probabilities—typically using formulas such as the Weibull method. These estimates are inherently uncertain, particularly when based on short observational records. Consequently, it is not statistically valid to evaluate uncertainty by simply counting the percentage of points within a 95% confidence interval on the flood frequency plot (Renard et al., 2013). To address this, the UQ-flood framework incorporates confidence interval for discharge quantiles (Sect. 3.6). This approach provides a more rigorous and informative representation of uncertainty in flood quantile estimation.


AMAX
GEV
UQ-flood
SIM
108 years of records
Annual Maximum Flow (cms)
300
200
100
2
5
10
20
50
100
Annual Exceedance Probability (1 in x years)

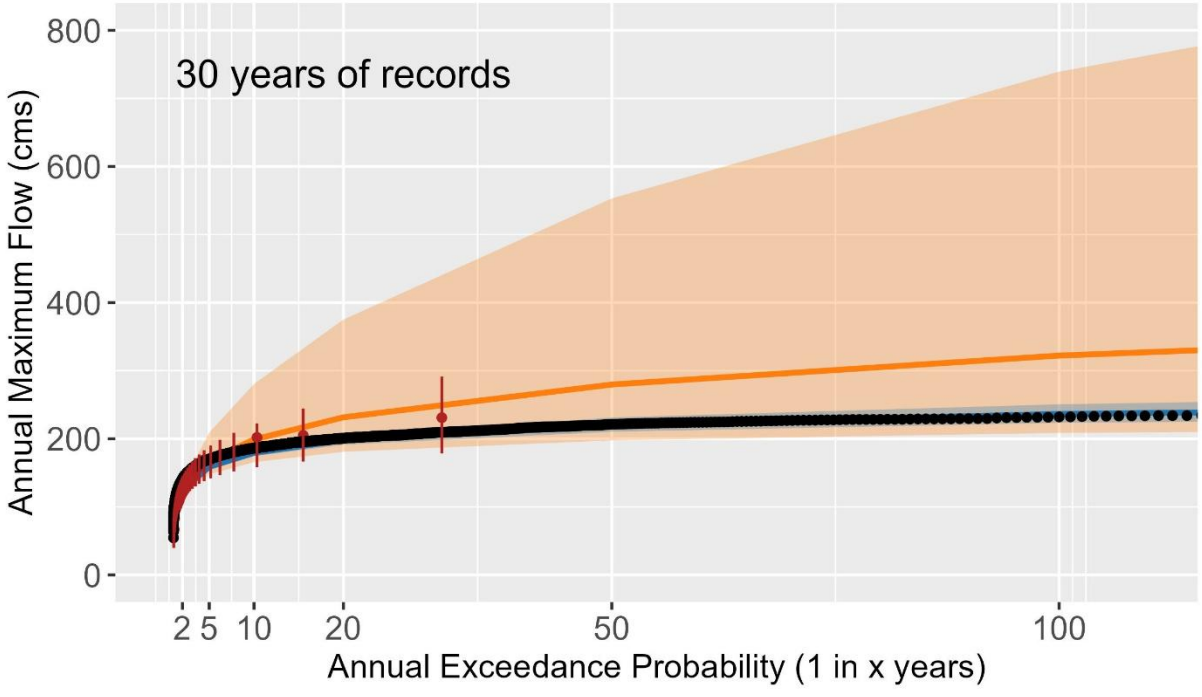

30 years of records
Annual Maximum Flow (cms)
800
600
400
200
0
2
5
10
20
50
100
Annual Exceedance Probability (1 in x years)

Figure 3: Flood frequency estimates and associated 95% uncertainty intervals for the Black River watershed derived using the statistical GEV approach and the UQ-flood framework. The observed annual maximum discharge (AMAX; red) distribution, shown with confidence intervals for discharge quantiles, and the deterministic simulation output (SIM; black points) are included for reference.. The top panel shows GEV results using long-term records (108 years); the bottom panel uses short records (30 years). The UQ-flood framework uses only the 30-year record in both cases.

4.2 Application to Additional Watersheds

To assess the generalisability of the proposed framework, we applied the same analysis to two additional watersheds: the Moyie River (left panels in Fig. 4) and the Lahave River (right panels in Fig. 4). As in the Black River analysis, flood frequency estimates were generated using a confidence interval for discharge quantiles, traditional statistical approach based on the Generalized Extreme Value (GEV) distribution, and the simulation-based method incorporating a residual error model (UQ-flood). For the GEV method, both long-term and short-term observational records were used, while the UQ-flood framework consistently relied on only 30 years of data. The confidence interval for discharge quantiles results with sampling uncertainty use the long records.

The results for both watersheds are consistent with those observed in the Black River. In the top panels (A and B), where GEV is applied using long records, and in the bottom panels (C and D), where GEV uses only 30 years of data, UQ-flood consistently produces narrower uncertainty intervals while remaining consistent with long-record empirical benchmarks. These narrower bands are more informative and actionable for risk-based decision-making. In contrast, the GEV-based estimates—especially when derived from short records—exhibit substantially wider uncertainty bounds, limiting their practical utility in planning and design applications.

Across all panels and watersheds, the simulated annual maximum flows without error correction (AMF_{sim}, shown as black points) consistently underestimate flood magnitudes. This systematic bias illustrates a key limitation of using continuous simulation in isolation. The integration of a residual error model in the UQ-flood framework is essential to correct any model bias and to propagate predictive uncertainty into flood quantile estimates. As a result, UQ-flood not only improves the reliability of simulation-based flood frequency analysis but also ensures that uncertainty estimates remain robust and interpretable, even in data-scarce environments.

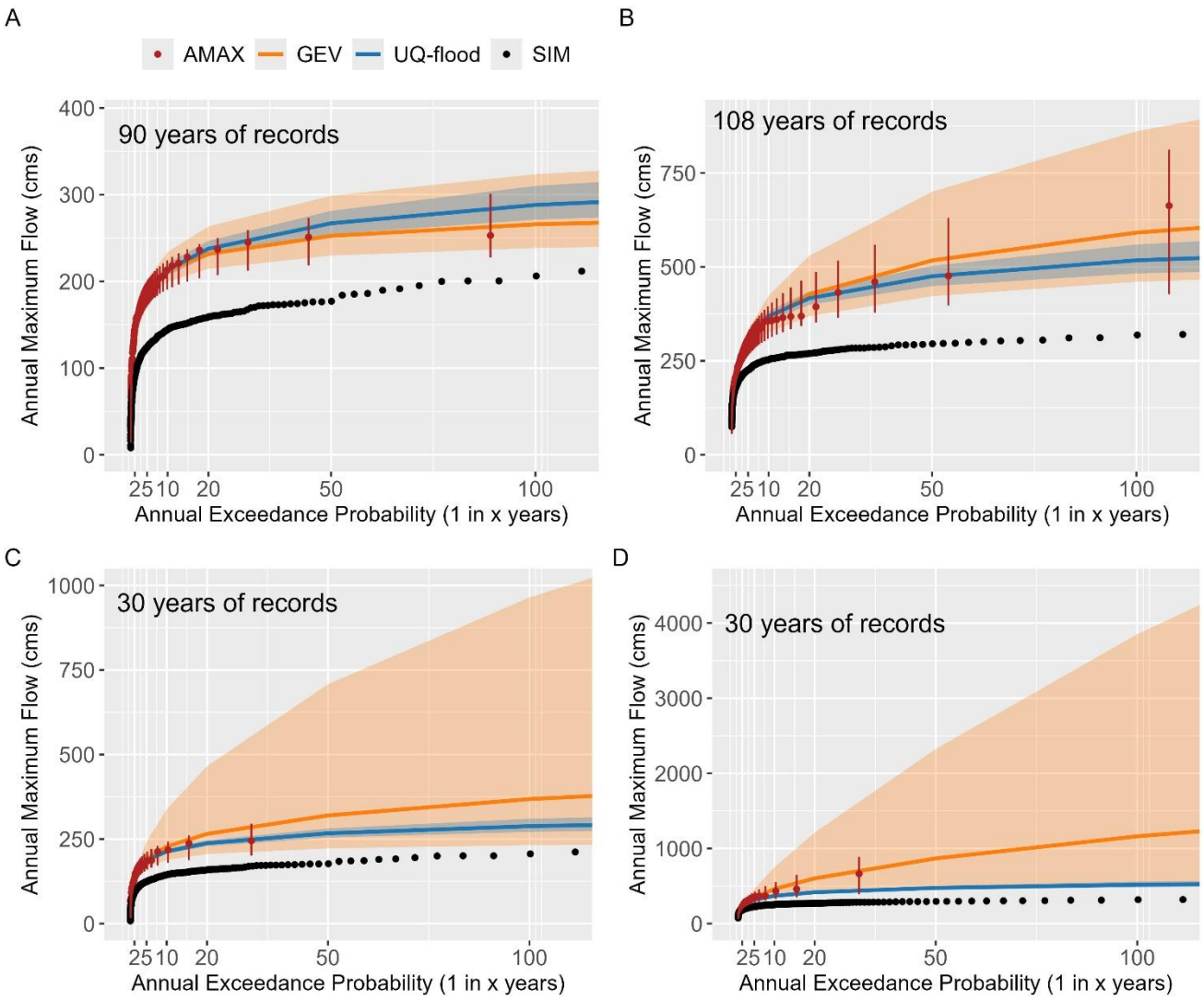


Figure 4: Flood frequency estimates and associated 95% uncertainty intervals for the Moyie River (left panels: A and C) and Lahave River (right panels: B and D) derived using the statistical GEV approach and the UQ-flood framework. The observed annual maximum discharge (AMAX; red) distribution, shown with confidence intervals for discharge quantiles, and the deterministic simulation output (SIM; black points) are included for reference. The top panel shows GEV results using long-term records (108 years); the bottom panel uses short records (30 years). The UQ-flood framework uses only the 30-year record in both cases

5 Discussion

The results of this study should be interpreted in the context of the broader short-sample problem in flood frequency analysis. As discussed in the Introduction, limited annual maximum flow records can produce uncertainty intervals for rare

flood quantiles that are too wide to support practical design and planning decisions. Existing strategies address this problem by adding information in time through historical or paleoflood data, in space through regional flood frequency analysis, or synthetically through continuous simulation. UQ-flood contributes to the third pathway by using process-based simulation to generate large flood samples, while explicitly correcting hydrologic model bias and propagating residual uncertainty at the annual maxima scale. In this sense, the framework is not intended to replace historical-data or regional approaches, but to provide a complementary, process-consistent alternative when local records are short and uncertainty bounds from conventional statistical FFA become impractical.

5.1 Placing the UQ-flood Framework in Context

Uncertainty is a fundamental aspect of flood frequency analysis (FFA), yet it has historically received limited attention in practical hydrology (Blöschl et al., 2019; Merz et al., 2015; Pappenberger and Beven, 2006). Traditional statistical methods—particularly those based on the Generalized Extreme Value (GEV) distribution—have long served as the foundation for regulatory and engineering applications (Vogel and Wilson, 1996). However, these approaches can be sensitive to sample size, rely on parametric assumptions, and often provide point estimates without a formal treatment of uncertainty (Vidrio-Sahagún et al., 2025). This can be especially limiting when working with short streamflow records, which are increasingly common in contemporary hydrologic design practice.

The UQ-flood framework presented in this study offers a complementary approach to existing methods by integrating process-based hydrologic simulation with a residual error model (REM) and a stochastic weather generator. This integration supports a shift towards simulation-based uncertainty quantification that is consistent with hydrologic processes. Rather than replacing existing methods, UQ-flood aims to enhance uncertainty characterisation by propagating uncertainty from model inputs through to flood quantile estimates. In this context, incorporating error models into continuous simulation frameworks may help improve bias correction and support more transparent flood risk estimation.

5.2 Advancing Uncertainty Quantification in FFA

A key contribution of this study is the implementation of an integrated simulation-based framework that combines a process-based hydrologic model, a stochastic weather generator, and a residual error model (REM) to support uncertainty characterisation in flood estimates. To our knowledge, this is among the first applications of the REM approach within the context of at-site flood frequency analysis (FFA), allowing for direct quantification of uncertainty in flood estimates. The REM is applied at the annual maxima scale, which aligns with the requirements of engineering design and regulatory assessments. This scale-specific focus may offer a more targeted way to represent uncertainty compared to conventional daily-scale error modelling.

The REM contributes to the UQ-flood framework by accounting for systematic biases in hydrologic simulations and representing residual variability not captured by deterministic model structures. While components such as continuous simulation, stochastic weather generation, and error modelling have been explored individually in previous FFA studies, the

integration of these elements into a unified framework is a distinguishing feature of UQ-flood (Fig. 2). This integration enables consistent propagation of uncertainty from meteorological inputs to flood quantiles.

A central motivation for applying the REM at the annual maxima scale is the mismatch between the calibration target of hydrologic models and the target variable in FFA. Hydrologic models are commonly calibrated using daily streamflow metrics that reward overall hydrograph performance, water balance, or seasonal dynamics. However, strong daily-flow performance does not guarantee unbiased representation of annual maximum flows, which are controlled by threshold processes, antecedent conditions, snowmelt–rainfall interactions, and tail behaviour. The experiments without the REM demonstrate this issue directly: deterministic continuous simulation tended to underestimate flood magnitudes across the study watersheds. By estimating residual errors between observed and simulated annual maxima, UQ-flood corrects the scale most relevant to design flood estimation rather than assuming that daily calibration is sufficient for flood-frequency applications.

Compared to other uncertainty quantification approaches, UQ-flood seeks to balance methodological rigour with practical implementation. For instance, unlike GLUE-based methods, it does not rely on subjective likelihood thresholds or behavioural criteria (Blazkova and Beven, 2009; Cameron et al., 2000). Similarly, while full Bayesian approaches offer comprehensive uncertainty treatment, they can be computationally intensive. Costa and Fernandes (2017), for example, proposed a Bayesian method that jointly infers parameters of the weather generator, hydrologic model, and a skew exponential power error distribution using MCMC proposed a Bayesian indirect flood estimation method that jointly infers parameters of the weather generator, hydrologic model, and a complex skew exponential power error distribution using MCMC. While robust, such methods may be less accessible for operational use (Costa and Fernandes, 2017). In contrast, UQ-flood applies a simplified error model, offering a faster and more accessible alternative without compromising statistical validity.

UQ-flood also offers an alternative to bootstrap-based single-site approaches, which depend on resampling the available data (Arnaud et al., 2017; Bomers et al., 2019; Burn, 2003; Mirzaei et al., 2015). In small-sample contexts—such as 30 years of annual maxima—bootstrap methods may struggle to capture uncertainty adequately, especially when the sample is unrepresentative. Resampling from limited datasets can lead to narrow confidence intervals and potentially biased estimates (Hesterberg, 2015). The commonly used percentile bootstrap interval is particularly sensitive to sample size. In contrast, UQ-flood appears to perform reasonably well under data-scarce conditions (Fig. 3), providing interpretable quantile estimates even with limited records.

Importantly, UQ-flood does not require fitting a parametric distribution directly to the annual maximum flow sample. Instead, flood quantiles are estimated empirically from large synthetic ensembles generated through process-based simulation and residual error modelling. This flexibility may improve interpretability and help avoid tail sensitivity issues associated with parametric distribution fitting (Macdonald et al., 2022; Merz et al., 2022).

Although this study does not explicitly address non-stationarity, the UQ-flood framework is well-suited to exploring such conditions. By perturbing the system with alternative land use scenarios, climate projections, or water management

strategies, users can assess the influence of changing drivers on flood hazard while maintaining a consistent uncertainty quantification approach.

Overall, UQ-flood contributes to the broader movement towards physics-informed uncertainty quantification in flood risk assessment. By embedding uncertainty within the modelling chain—from climate inputs to flood quantiles—it offers a scalable and decision-relevant methodology that complements existing statistical approaches.

5.3 Limitations and Future Directions

While the UQ-flood framework offers a robust approach for quantifying uncertainty in flood frequency analysis, several limitations and areas for future development should be acknowledged. First, caution is warranted when directly comparing the uncertainty ranges produced by UQ-flood and traditional statistical FFA. These methods quantify different sources of uncertainty. UQ-flood captures the predictive uncertainty of the hydrologic model, including uncertainties in inputs, structure, and parameters. In contrast, statistical FFA primarily characterises parametric uncertainty associated with fitting the GEV distribution. Despite these theoretical differences, both approaches provide valuable representations of uncertainty in flood quantile estimation and can be regarded as broadly comparable for practical decision-making (Abbasian et al., 2025).

The accuracy and reliability of the resulting flood estimates are also inherently influenced by the choices made across the modelling chain and the methods used to represent uncertainty. Decisions related to hydrologic model structure, calibration strategy, weather generator configuration, and residual error modelling all affect the resulting flood quantiles. As highlighted by Kritidou et al., (2026), these modelling options play a critical role in determining the reliability of flood estimates and should therefore be carefully considered when applying simulation-based flood frequency analysis frameworks.A key enabler of the UQ-flood framework is the stochastic weather generator (SWG), which provides the synthetic meteorological inputs necessary for long-term continuous simulation. The SWG is essential for producing the large sample sizes required to support robust, nonparametric flood frequency estimation and effective uncertainty quantification. However, selecting and adjusting a reliable SWG is not a trivial task (Grimaldi et al., 2022). Currently, there are no standardised guidelines for choosing an appropriate SWG for FFA, and each available method carries distinct assumptions and limitations (Fischer et al., 2025). Successful implementation may require advanced expertise in hydroclimatology and statistical modelling, as the choice of SWG can significantly influence the accuracy and realism of simulated flood extremes.

In this study, we used the CoSMoS weather generator (Papalexiou, 2018) due to its ability to simulate non-Gaussian time series. However, we identified and corrected a bias in the generation of extreme precipitation values, which initially exhibited unrealistically long-tailed behaviour when used to produce very long meteorological time series. This issue arises in part because SWGs tend to be fit to the full data distribution rather than specifically to the tail. To address this, we implemented a constraint on extreme precipitation to preserve physical plausibility. Importantly, UQ-flood is not tied to CoSMoS or any single weather generator. It can be informed by any meteorological sampling approaches, such as stochastic storm transposition (Wright et al., 2020), reforecast ensemble pooling (Brunner and Slater, 2022), or convection-resolving

climate models (Dallan et al., 2024). These alternatives may be particularly useful in regions where convective storms or surprise floods dominate the hydrologic regime. This flexibility ensures UQ-flood can be adapted to a wide range of climatic contexts and emerging flood risks. The synthetic sequences should be interpreted as stationary realisations of the climate represented by the calibration period, not as projections of future climate.

This study does not account for observational uncertainty associated with streamflow measurements. However, rating-curve uncertainty can materially alter estimated flood magnitudes, especially for long return periods. Similarly, uncertainty in the parameters of the weather generator used to simulate rainfall inputs is not considered. To some extent, we assume that rainfall records at each site are equal to or longer than the corresponding streamflow records, which may not hold in all cases. Including these additional sources of uncertainty could improve the robustness of future analyses and support more comprehensive flood risk assessments.

A core assumption of UQ-flood is that process-based hydrologic models can derive more useful information than purely statistical extrapolation from short records. That is, expert knowledge about watershed processes—such as water balance, storage, and runoff generation—enables the model to constrain future flood behavior more effectively than sampling-based uncertainty alone (Su and Tung, 2013). In this sense, UQ-flood acts as a more intelligent flood sampler, embedding physical understanding directly into the generation of extreme event ensembles.

Another strength of UQ-flood is its modular structure, which makes it readily extendable to different modeling environments. While this study employed a lumped, conceptual hydrologic model (HBV-EC), which was appropriate given the scale and data availability of the selected watersheds, the framework is equally compatible with physically based distributed models. Applying UQ-flood within a distributed modeling context could enhance the spatial realism of simulated flood responses, extend the approach to ungauged watersheds, and improve the representativeness of predictive ensembles (Perez et al., 2019). However, the performance and added value of UQ-flood in such models remain to be systematically evaluated, particularly in terms of how well the residual error model captures localized hydrologic variability and structural uncertainty across heterogeneous landscapes.

This study focused on three Canadian watersheds with long streamflow records to enable evaluation. While this provided validation of the UQ-flood framework under diverse hydroclimatic conditions, broader testing across data-scarce and climatically varied regions is desirable to establish its generalizability. Additionally, comparison with regional FFA methods is necessary to further assess its relative performance and applicability.

5.4 Implications for Design and Decision-Making

The findings from the three case studies highlight the practical value of UQ-flood for flood risk management. When applied to watersheds with long records, UQ-flood produced flood quantile estimates consistent with traditional statistical methods, but with significantly narrower uncertainty intervals. This has direct implications for design safety, regulatory compliance, and investment decision-making. Critically, when applied under short-record conditions (e.g., 30 years), UQ-flood

maintained robust and actionable flood estimates, while statistical FFA produced wide, often impractical uncertainty bounds. These results demonstrate the method's resilience to data scarcity, a common challenge in real-world applications.

From a practical standpoint, narrower uncertainty bounds enable engineers and planners to avoid both overdesign (leading to excessive costs) and under design (resulting in heightened risk). In fields such as insurance, urban planning, and floodplain mapping, proper uncertainty quantification supports more accurate risk pricing, prioritization of investments, and the design of climate-resilient infrastructure.

Importantly, UQ-flood also enhances the interpretability and credibility of flood estimates. Unlike statistical methods that rely heavily on distributional assumptions, the UQ-flood framework is built on a process-based hydrologic model and observed model performance, making it more transparent and intuitive for engineers, scientists, and decision-makers to communicate and justify design choices.

Finally, the UQ-flood framework aligns with the emerging concept of the design simulation approach (Fischer et al., 2025), which advocates for simulation-driven flood estimates as the foundation for engineering design. In this context, UQ-flood offers a pathway to inform key design decisions—such as the selection of safety factors—based on quantifiable uncertainty. Rather than relying on fixed safety margins, decision-makers can assess how confident they are in their design flood estimates and adjust accordingly, thereby improving both cost-efficiency and risk robustness.

6 Conclusions

Flood estimates are inherently uncertain due to a range of contributing factors, including model structural errors, parameter uncertainty, data limitations, and forcing variability. Failure to account for these uncertainties can result in either overly conservative designs or underestimated flood risks. Accurately characterizing and transparently communicating this uncertainty is essential to ensure the reliability and usability of flood frequency estimates in engineering design and flood risk management.

In this study, we proposed and tested an uncertainty quantification framework for simulation-based flood frequency analysis (UQ-flood), which explicitly accounts for hydrologic model errors through the integration of a statistical residual error model (REM). The framework corrects systematic biases in simulated flows and propagates predictive uncertainty into flood quantile estimates. It combines a process-based hydrologic model, a stochastic weather generator, and a residual error model to characterize uncertainty of extreme events. Central to this framework is the error model, which captures both systematic biases and unexplained variability in hydrologic simulations.

As a proof of concept, we applied the proposed UQ-flood framework to three Canadian watersheds with contrasting climatic and physiographic characteristics. To evaluate its performance, we compared UQ-flood against traditional statistical flood frequency analysis (FFA) methods based on the Generalized Extreme Value (GEV) distribution and Bayesian inference. The key findings are as follows:

1. The proposed UQ-flood framework produces flood magnitude estimates that are consistent with those obtained from statistical methods applied to long observational records (e.g., 100 years), but with considerably narrower and more informative uncertainty bounds, even when using shorter 30-year records.
2. When applied to short observational records (e.g., 30 years), statistical FFA yields wide and often impractical uncertainty ranges. In contrast, the UQ-flood framework provides robust and actionable estimates.
3. Continuous simulation without incorporating an explicit error model typically leads to systematic bias flood quantiles, underscoring the critical importance of accounting for model error in simulation-based frequency analysis.

These results suggest that explicitly modelling hydrologic simulation error at the annual maxima scale can make continuous simulation more useful for engineering flood-frequency applications. UQ-flood offers a process-consistent and practical alternative for at-site flood risk estimation, particularly where observed records are short and conventional uncertainty intervals are too broad to guide decisions. This capability is especially valuable in regions where Regional Flood Frequency Analysis (RFFA) is limited by sparse gauging networks, heterogeneous catchment characteristics, or a lack of suitable donor catchments. Future work should evaluate the framework across a wider range of hydroclimatic regimes, incorporate additional uncertainty sources such as rating-curve and weather-generator parameter uncertainty, and explore mechanism-aware residual error models conditioned on flood-generating processes.

Code and data availability

The stochastic weather generator CoSMoS is available for download from its Github repository: https://github.com/COSMOS. The Raven hydrological modelling framework can be accessed at https://raven.uwaterloo.ca/. Historical weather data were obtained from Environment and Climate Change Canada (ECCC) and are publicly available at https://climate.weather.gc.ca/. Streamflow data were sourced from the Water Survey of Canada and can be downloaded via https://dd.weather.gc.ca/hydrometric/csv/. The UQ-flood scripts and datasets used in the analyses presented in this manuscript are openly available at: https://github.com/rarabzad/REM_UQ_EXTREME/tree/main.

Author contributions

JRC: Writing original draft, investigation, visualization, software, and methodology

JRCr: Conceptualization, funding acquisition, project administration, supervision, writing, investigation and methodology

BAT: Supervision, writing, investigation and methodology

RA: Software and data curation

Competing interest

The authors declare that they have no conflict of interest.

Financial support

This work was funded under the Natural Resources Canada Flood Hazard Identification and Mapping Program (FHIMP) project FHIMP23-27P2.

References

Abbasian, M., Wright, D. B., Notaro, M., Vavrus, S., and Vimont, D. J.: Flood frequency sampling error: insights from regional analysis, stochastic storm transposition, and physics-based modeling, J. Hydrol., 662, 133802, https://doi.org/10.1016/J.JHYDROL.2025.133802, 2025.

Andersen, C. B., Wright, D. B., and Thorndahl, S.: CON-SST-RAIN: Continuous Stochastic Space–Time Rainfall generation based on Markov chains and transposition of weather radar data, J. Hydrol., 637, 131385, https://doi.org/10.1016/J.JHYDROL.2024.131385, 2024.

Apel, H., Thieken, A. H., Merz, B., and Blöschl, G.: Flood risk assessment and associated uncertainty, Nat. Hazards Earth Syst. Sci., 4, 295–308, https://doi.org/10.5194/NHESS-4-295-2004, 2004.

Arnaud, P. and Lavabre, J.: Coupled rainfall model and discharge model for flood frequency estimation, Water Resour. Res., 38, 11–1, https://doi.org/10.1029/2001WR000474, 2002.

Arnaud, P., Cantet, P., and Odry, J.: Uncertainties of flood frequency estimation approaches based on continuous simulation using data resampling, J. Hydrol., 554, https://doi.org/10.1016/j.jhydrol.2017.09.011, 2017.

Baâ, K.M., Díaz-Delgado, C. and Caârsteanu, A.: Confidence Intervals of Quantiles in Hydrology Computed by an Analytical Method, Natural Hazards, 24, 1–12, https://doi.org/10.1023/A:1011101700032, 2001.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52, 3599–3622, https://doi.org/10.1002/2015WR018247, 2016.

Benito, G., Lang, M., Barriendos, M. et al.: Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods, Natural Hazards, 31, 623–643, https://doi.org/10.1023/B:NHAZ.0000024895.48463.eb, 2004.

Blazkova, S. and Beven, K.: Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), Water Resour. Res., 38, 14–1, https://doi.org/10.1029/2001WR000500, 2002.

Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood

frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resour. Res., 45, W00B16, https://doi.org/10.1029/2007WR006726, 2009.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M., Collins, A., Croke, B., Dathe, A., David, P. C., de Barros, F. P. J., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Düthmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., et al.: Twenty-three Unsolved Problems in Hydrology (UPH) – a community perspective, Hydrol. Sci. J., 02626667.2019.1620507, https://doi.org/10.1080/02626667.2019.1620507, 2019.

Bomers, A., Schielen, R. M. J., and Hulscher, S. J. M. H.: Decreasing uncertainty in flood frequency analyses by including historic flood events in an efficient bootstrap approach, Nat. Hazards Earth Syst. Sci., 19, 1895–1908, https://doi.org/10.5194/NHESS-19-1895-2019, 2019.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, J. R. Stat. Soc., 26, 211–252, 1964.

Brunner, M. I., Melsen, L. A., Wood, A. W., Rakovec, O., Mizukami, N., Knoben, W. J. M., and Clark, M. P.: Flood spatial coherence, triggers, and performance in hydrological simulations: large-sample evaluation of four streamflow-calibrated models, Hydrol. Earth Syst. Sci., 25, 105–119, https://doi.org/10.5194/hess-25-105-2021, 2021

Brunner, M. I. and Slater, L. J.: Extreme floods in Europe: Going beyond observations using reforecast ensemble pooling, Hydrol. Earth Syst. Sci., 26, 469–482, https://doi.org/10.5194/HESS-26-469-2022, 2022.

Burn, D. H.: The use of resampling for estimating confidence intervals for single site and pooled frequency analysis / Utilisation d'un rééchantillonnage pour l'estimation des intervalles de confiance lors d'analyses fréquentielles mono et multi-site, Hydrol. Sci. J., 48, 25–38, https://doi.org/10.1623/HYSJ.48.1.25.43485, 2003.

Burn, D. H. and Whitfield, P. H.: Climate related changes to flood regimes show an increasing rainfall influence, J. Hydrol., 617, 129075, https://doi.org/10.1016/J.JHYDROL.2023.129075, 2023.

Buttle, J. M., Allen, D. M., Caissie, D., Davison, B., Hayashi, M., Peters, D. L., Pomeroy, J. W., Simonovic, S., St-Hilaire, A., and Whitfield, P. H.: Flood processes in Canada: Regional and special aspects, Can. Water Resour. J. / Rev. Can. des ressources hydriques, 41, 7–30, https://doi.org/10.1080/07011784.2015.1131629, 2016.

Cameron, D., Beven, K., Tawn, J., and Naden, P.: Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation), Hydrol. Earth Syst. Sci., 4, 23–34, https://doi.org/10.5194/HESS-4-23-2000, 2000.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics, Springer-Verlag, Berlin, https://doi.org/10.1007/978-1-4471-3675-0, 2001.

Costa, V. and Fernandes, W.: Bayesian estimation of extreme flood quantiles using a rainfall-runoff model and a stochastic daily rainfall generator, J. Hydrol., 554, 137–154, https://doi.org/10.1016/J.JHYDROL.2017.09.003, 2017.

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., Mai, J., Serrer, M., Sgro, N., Shafii, M., Snowdon, A. P., and Tolson, B. A.: Flexible watershed simulation with the Raven hydrological modelling framework, Environ. Model. Softw., 129, https://doi.org/10.1016/j.envsoft.2020.104728, 2020.

Cunnane, C.: Factors affecting choice of distribution for flood series, Hydrol. Sci. J., 30, 25–36, https://doi.org/10.1080/02626668509490969, 1985.

Cunnane, C.: Methods and merits of regional flood frequency analysis, J. Hydrol., 100, 269-290, https://doi.org/10.1016/0022-1694(88)90188-6, 1988.

Dallan, E., Borga, M., Fosser, G., Canale, A., Roghani, B., Marani, M., and Marra, F.: A Method to Assess and Explain Changes in Sub-Daily Precipitation Return Levels From Convection-Permitting Simulations, Water Resour. Res., 60, e2023WR035969, https://doi.org/10.1029/2023WR035969, 2024.

Eagleson, P. S.: Dynamics of flood frequency, Water Resour. Res., 8, 878–898, https://doi.org/10.1029/WR008I004P00878, 1972.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G.: Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity, Water Resour. Res., 50, 2350–2375, https://doi.org/10.1002/2013WR014185, 2014.

Farmer, W. H. and Vogel, R. M.: On the deterministic and stochastic use of hydrologic models, Water Resour. Res., 52, 5619–5633, https://doi.org/10.1002/2016WR019129, 2016.

Fischer, S. and Schumann, A. H.: Multivariate Flood Frequency Analysis in Large River Basins Considering Tributary Impacts and Flood Types, Water Resour. Res., 57, e2020WR029029, https://doi.org/10.1029/2020WR029029, 2021.

Fischer, S., Dallan, E., Fiori, A., Grimaldi, S., Kochanek, K., Prieto, C., Reis, D. S., and Volpi, E.: Hydrological design in the HELPING decade – inspiring the community to innovate the hydrological design concept, Hydrol. Sci. J., 70, 375–389, https://doi.org/10.1080/02626667.2024.2436634, 2025.

Fischer, S., Pahlow, M., and Kumar, S. S.: Impact of catchment and climate attributes on flood generating processes and their effect on flood statistics, J. Hydrol., 646, 132-361, https://doi.org/10.1016/j.jhydrol.2024.132361, 2025.

Franchini, M., Helmlinger, K. R., Foufoula-Georgiou, E., and Todini, E.: Stochastic storm transposition coupled with rainfall—runoff modeling for estimation of exceedance probabilities of design floods, J. Hydrol., 175, 511–532, https://doi.org/10.1016/S0022-1694(96)80022-9, 1996.

Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in R, J. Stat. Softw., 72, 1–39, https://doi.org/10.18637/JSS.V072.I08, 2016.

Grimaldi, S., Schumann, G. J.-P., Shokri, A., Walker, J. P., & Pauwels, V. R. N.: Challenges, opportunities and pitfalls for

global coupled hydrologic-hydraulic modeling of floods, Water Resour. Res., 55, 5277–5300, https://doi.org/10.1029/2018WR024289, 2019.

Grimaldi, S., Volpi, E., Langousis, A., Michael Papalexiou, S., De Luca, D. L., Piscopia, R., Nerantzaki, S. D., Papacharalampous, G., and Petroselli, A.: Continuous hydrologic modelling for small and ungauged basins: A comparison of eight rainfall models for sub-daily runoff simulations, J. Hydrol., 610, 127866, https://doi.org/10.1016/J.JHYDROL.2022.127866, 2022.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, 2009.

Haberlandt, U. and Radtke, I.: Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows, Hydrol. Earth Syst. Sci., 18, https://doi.org/10.5194/hess-18-353-2014, 2014.

Hall, J. and Solomatine, D.: A framework for uncertainty analysis in flood risk management decisions, Int. J. River Basin Manag., 6, 85–98, https://doi.org/10.1080/15715124.2008.9635339, 2008.

Hamed, K. H. and Ramachandra Rao, A.: Flood Frequency Analysis, 1st Editio., CRC Press, Boca Raton, 376 pp., https://doi.org/10.1201/9780429128813, 2019.

Hamilton, A. S., Hutchinson, D. G., and Moore, R. D.: Estimating Winter Streamflow Using Conceptual Streamflow Model, J. Cold Reg. Eng., 14, 158–175, https://doi.org/10.1061/(ASCE)0887-381X(2000)14:4(158), 2000.

Hazelton, M. .: Methods of Moments Estimation BT, in: International Encyclopedia of Statistical Science, edited by: Lovric, M., Springer Berlin Heidelberg, Berlin, 816–817, https://doi.org/10.1007/978-3-642-04898-2, 2011.

Hesterberg, T. C.: What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum, Am. Stat., 69, 371–386, https://doi.org/10.1080/00031305.2015.1089789, 2015.

Honti, M., Stamm, C., and Reichert, P.: Integrated uncertainty assessment of discharge predictions with a statistical error model, Water Resour. Res., 49, 4866–4884, https://doi.org/10.1002/wrcr.20374, 2013.

Hu, L., Nikolopoulos, E. I., Marra, F., and Anagnostou, E. N.: Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: An evaluation over the contiguous United States, J. Flood Risk Manag., 13, e12580, https://doi.org/10.1111/JFR3.12580, 2020.

Hunter, J., Thyer, M., McInerney, D., and Kavetski, D.: Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions, J. Hydrol., 603, 126578, https://doi.org/10.1016/J.JHYDROL.2021.126578, 2021.

Jenkinson, A. F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. R. Meteorol. Soc., 81, 158–171, https://doi.org/10.1002/QJ.49708134804, 1955.

Jiang, X., Gupta, H. V., Liang, Z., and Li, B.: Toward Improved Probabilistic Predictions for Flood Forecasts Generated Using Deterministic Models, Water Resour. Res., 55, 9519–9543, https://doi.org/10.1029/2019WR025477, 2019.

Kasiviswanathan, K. S. S., He, J., and Tay, J.-H. H.: Flood frequency analysis using multi-objective optimization based interval estimation approach, J. Hydrol., 545, 251–262, 2017.

Koutsoyiannis, D. and Montanari, A.: Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions, Water Resour. Res., 58, e2021WR031215, https://doi.org/10.1029/2021WR031215, 2022.

Kritidou, E., Kauzlaric, M., Staudiger, M., Evin, G., Hingray, B., Vis, M., and Viviroli, D.: Partitioning uncertainties of extreme flood estimates using long continuous simulations, J. Hydrol., 134-804, https://doi.org/10.1016/j.jhydrol.2025.134804, 2026.

Laio, F., Di Baldassarre, G., and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes, Water Resour. Res., 45, W07416, https://doi.org/10.1029/2007WR006666, 2009.

Lamb, R., Faulkner, D., Wass, P., and Cameron, D.: Have applications of continuous rainfall–runoff simulation realized the vision for process-based flood frequency analysis?, Hydrol. Process., 30, 2463–2481, https://doi.org/10.1002/HYP.10882, 2016.

Li, M., Wang, Q. J., Bennett, J. C., and Robertson, D. E.: Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting, Hydrol. Earth Syst. Sci., 20, 3561–3579, https://doi.org/10.5194/hess-20-3561-2016, 2016.

Liang, Z., Chang, W., and Li, B.: Bayesian flood frequency analysis in the light of model and parameter uncertainties, Stoch. Environ. Res. Risk Assess., 26, 721–730, https://doi.org/10.1007/S00477-011-0552-Y/TABLES/6, 2012.

Lucas, M., Renard, B., Le Coz, J., Lang, M., Bard, A., and Pierrefeu, G.: Are historical stage records useful to decrease the uncertainty of flood frequency analysis ? A 200-year long case study, J. Hydrol., 624, 129840, https://doi.org/10.1016/J.JHYDROL.2023.129840, 2023.

Macdonald, E., Merz, B., Guse, B., Wietzke, L., Ullrich, S., Kemter, M., Ahrens, B., and Vorogushyn, S.: Event and Catchment Controls of Heavy Tail Behavior of Floods, Water Resour. Res., 58, e2021WR031260, https://doi.org/10.1029/2021WR031260;JOURNAL:JOURNAL:19447973;REQUESTEDJOURNAL:JOURNAL:19447973;WGROUP:STRING:PUBLICATION, 2022.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., and Kuczera, G.: Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors, Water Resour. Res., 53, 2199–2239, https://doi.org/10.1002/2016WR019168, 2017.

McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Tuteja, N., and Kuczera, G.: Multi-temporal Hydrological Residual Error Modeling for Seamless Subseasonal Streamflow Forecasting, Water Resour. Res., 56, e2019WR026979, https://doi.org/10.1029/2019WR026979, 2020.

Merz, B. and Thieken, A. H.: Separating natural and epistemic uncertainty in flood frequency analysis, J. Hydrol., 309, 114–132, https://doi.org/10.1016/J.JHYDROL.2004.11.015, 2005.

Merz, B. and Thieken, A. H.: Flood risk curves and uncertainty bounds, Nat. Hazards, 51, 437–458, https://doi.org/10.1007/S11069-009-9452-6/FIGURES/10, 2009.

Merz, B., Vorogushyn, S., Lall, U., Viglione, A., and Blöschl, G.: Charting unknown waters—On the role of surprise in flood risk assessment and management, Water Resour. Res., 51, 6399–6416, https://doi.org/10.1002/2015WR017464, 2015.

Merz, B., Basso, S., Fischer, S., Lun, D., Blöschl, G., Merz, R., Guse, B., Viglione, A., Vorogushyn, S., Macdonald, E., Wietzke, L., and Schumann, A.: Understanding Heavy Tails of Flood Peak Distributions, Water Resour. Res., 58, e2021WR030506,
https://doi.org/10.1029/2021WR030506;JOURNAL:JOURNAL:19447973;REQUESTEDJOURNAL:JOURNAL:19447973;WGROUP:STRING:PUBLICATION, 2022.

Milly, P. C. D., Wetherald, R. T., Dunne, K. A., and Delworth, T. L.: Increasing risk of great floods in a changing climate, Nature, 415, 514–517, https://doi.org/10.1038/415514a, 2002.

Mirzaei, M., Huang, Y. F., El-Shafie, A., Chimeh, T., Lee, J., Vaizadeh, N., and Adamowski, J.: Uncertainty analysis for extreme flood events in a semi-arid region, Nat. Hazards, 78, 1947–1960, https://doi.org/10.1007/S11069-015-1812-9/FIGURES/7, 2015.

Montanari, A. and Koutsoyiannis, D.: A blueprint for process-based modeling of uncertain hydrological systems, Water Resour. Res., 48, W09555, https://doi.org/10.1029/2011WR011412, 2012.

Newman, A. J., Stone, A. G., Saharia, M., Holman, K. D., Addor, N., and Clark, M. P.: Identifying sensitivities in flood frequency analyses using a stochastic hydrologic modeling system, Hydrol. Earth Syst. Sci., 25, 5603–5621, https://doi.org/10.5194/HESS-25-5603-2021, 2021.

Papalexiou, S. M.: Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency, Adv. Water Resour., 115, 234–252, https://doi.org/10.1016/J.ADVWATRES.2018.02.013, 2018.

Pappenberger, F. and Beven, K. J.: Ignorance is bliss: Or seven reasons not to use uncertainty analysis, Water Resour. Res., 42, W05302, https://doi.org/10.1029/2005WR004820, 2006.

Paquet, E., Garavaglia, F., Garçon, R., and Gailhard, J.: The SCHADEX method: A semi-continuous rainfall–runoff simulation for extreme flood estimation, J. Hydrol., 495, 23–37, https://doi.org/10.1016/j.jhydrol.2013.04.045, 2013.

Parkes, B. and Demeritt, D.: Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates, J. Hydrol., 540, 1189–1208, https://doi.org/10.1016/J.JHYDROL.2016.07.025, 2016.

Perez, G., Mantilla, R., Krajewski, W. F., and Wright, D. B.: Using Physically Based Synthetic Peak Flows to Assess Local and Regional Flood Frequency Analysis Methods, Water Resour. Res., 55, 8384–8403, https://doi.org/10.1029/2019WR024827, 2019.

Reis, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, J. Hydrol., 313, 97–116, https://doi.org/10.1016/J.JHYDROL.2005.02.028, 2005.

Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., Najib, K., Carreau, J., Arnaud, P., Aubert, Y., Borchi, F., Soubeyroux, J.-M., Jourdain, S., Veysseire, J.-M., Sauquet, E., Cipriani, T., and Auffray, A.: Data-based comparison of frequency analysis methods: A general framework, Water Resour. Res., 49, doi:10.1002/wrcr.20087, 2013.Romero-Cuellar, J., Arabzadeh, R., Craig, J. R., Tolson, B. A., and Mai, J.: A multi-model evaluation of probabilistic

streamflow predictions via residual error modelling, J. Hydrol., 635, 131152, https://doi.org/10.1016/J.JHYDROL.2024.131152, 2024.

Seibert, J. and Bergström, S.: A retrospective on hydrological catchment modelling based on half a century with the HBV model, Hydrol. Earth Syst. Sci., 26, 1371–1388, https://doi.org/10.5194/HESS-26-1371-2022, 2022.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, https://doi.org/10.5194/hess-16-3315-2012, 2012.

Serinaldi, F.: Assessing the applicability of fractional order statistics for computing confidence intervals for extreme quantiles, J. Hydrol., 376, 528–541, https://doi.org/10.1016/J.JHYDROL.2009.07.065, 2009.

Shabestanipour, G., Brodeur, Z., Farmer, W. H., Steinschneider, S., Vogel, R. M., and Lamontagne, J. R.: Stochastic Watershed Model Ensembles for Long-Range Planning: Verification and Validation, Water Resour. Res., 59, e2022WR032201, https://doi.org/10.1029/2022WR032201, 2023.

Sikorska, A. E., Viviroli, D., and Seibert, J.: Flood-type classification in mountainous catchments using crisp and fuzzy decision trees, Water Resour. Res., 51, 7959–7976, https://doi.org/10.1002/2015WR017326, 2015.

Sivapalan, M. and Samuel, J. M.: Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment, Hydrol. Process., 23, 1671–1675, https://doi.org/10.1002/HYP.7292, 2009.

Sivapalan, M., Blöschl, G., Merz, R., and Gutknecht, D.: Linking flood frequency to long-term water balance: Incorporating effects of seasonality, Water Resour. Res., 41, 1–17, https://doi.org/10.1029/2004WR003439, 2005.

Spellman, P., Webster, V., Watkins, D.: Bias correcting instantaneous peak flows generated using a continuous, semi-distributed hydrologic model, J Flood Risk Management., 11:e12342. https://doi.org/10.1111/jfr3.12342, 2018.

Stahl, K., Moore, R. D., Shea, J. M., Hutchinson, D., and Cannon, A. J.: Coupled modelling of glacier and streamflow response to future climate scenarios, Water Resour. Res., 44, https://doi.org/10.1029/2007WR005956, 2008.

Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R.: Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, Water Resour. Res., 44, W00B06, https://doi.org/10.1029/2008WR006822, 2008.

Su, H. T. and Tung, Y. K.: Incorporating uncertainty of distribution parameters due to sampling errors in flood-damage-reduction project evaluation, Water Resour. Res., 49, 1680–1692, https://doi.org/10.1002/WRCR.20116, 2013.

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Int. J. River Basin Manag., 6, 123–137, https://doi.org/10.1080/15715124.2008.9635342, 2008.

Tolson, B. A. and Shoemaker, C. A.: Efficient prediction uncertainty approximation in the calibration of environmental simulation models, Water Resour. Res., 44, 4411, https://doi.org/10.1029/2007WR005869, 2008.

Velásquez, N., Krajewski, W.F. Effect of streamflow measurement error on flood frequency estimation. Stoch Environ Res Risk Assess 38, 2903–2910, https://doi.org/10.1007/s00477-024-02707-1, 2024.

Vidrio-Sahagún, C. T., Ruschkowski, J., He, J., Pietroniro, A., and Hairabedian, M.: Design flood estimation in flood hazard studies: a three-decade systematic review of practices in Canada, Can. Water Resour. J. / Rev. Can. des ressources hydriques, 1–20, https://doi.org/10.1080/07011784.2025.2462603, 2025.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, Water Resour. Res., 49, 675–692, https://doi.org/10.1029/2011WR010782, 2013.

Vogel, R. M.: Stochastic watershed models for hydrologic risk management, Water Secur., 1, 28–35, https://doi.org/10.1016/J.WASEC.2017.06.001, 2017.

Vogel, R. M. and Wilson, I.: Probability Distribution of Annual Maximum, Mean, and Minimum Streamflows in the United States, J. Hydrol. Eng., 1, 69–76, https://doi.org/10.1061/(ASCE)1084-0699(1996)1:2(69), 1996.

Whitfield, P. H. and Cannon, A. J.: Recent Variations in Climate and Hydrology in Canada, Can. Water Resour. J., 25, 19–65, https://doi.org/10.4296/CWRJ2501019;WGROUP:STRING:PUBLICATION, 2000.

Wright, D. B., Yu, G., and England, J. F.: Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects, J. Hydrol., 585, 124816, https://doi.org/10.1016/J.JHYDROL.2020.124816, 2020.

Yu, G., Wright, D. B., Zhu, Z., Smith, C., and Holman, K. D.: Process-based flood frequency analysis in an agricultural watershed exhibiting nonstationary flood seasonality, Hydrol. Earth Syst. Sci., 23, 2225–2243, https://doi.org/10.5194/HESS-23-2225-2019, 2019.

Yu, G., Liu, T., McGuire, L. A., Wright, D. B., Hatchett, B. J., Miller, J. J., Berli, M., Giovando, J., Bartles, M., and Floyd, I. E.: Process-Based Quantification of the Role of Wildfire in Shaping Flood Frequency, Water Resour. Res., 59, e2023WR035013, https://doi.org/10.1029/2023WR035013, 2023.

Zaghloul, M., Papalexiou, S. M., Elshorbagy, A., and Coulibaly, P.: Revisiting flood peak distributions: A pan-Canadian investigation, Adv. Water Resour., 145, 103720, https://doi.org/10.1016/J.ADVWATRES.2020.103720, 2020.

Zégre, N., Skaugset, A. E., Som, N. A., McDonnell, J. J., and Ganio, L. M.: In lieu of the paired catchment approach: Hydrologic model change detection at the catchment scale, Water Resour. Res., 46, W11544, https://doi.org/10.1029/2009WR008601, 2010.