https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Open Access
Discussions

Modelling flood frequency and magnitude in glacially conditioned settings: land use matters

Pamela E. Tetford[1], Joseph R. Desloges[1]

[1]Department of Geography, University of Toronto, Toronto, ON, M5S 3G3, Canada

Correspondence to: Pamela E. Tetford (pam.tetford@mail.utoronto.ca)

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

Abstract. A reliable flood frequency analysis (FFA) requires selection of an appropriate statistical distribution to model historic streamflow data and, where streamflow data are not available (ungauged sites), a regression-based regional flood frequency analysis (RFFA) often correlates well with downstream channel discharge to drainage area relations. However, the predictive strength of the accepted RFFA relies on an assumption of homogeneous watershed conditions. For glacially conditioned fluvial systems, inherited glacial landforms, sediments, and variable land use can alter flow paths and modify flow regimes. This study compares a multi-variate RFFA that considers 28 explanatory variables to characterize variable watershed conditions (i.e., surficial geology, climate, topography, and land use) to an accepted power-law relationship between discharge and drainage area. Archived gauge data from southern Ontario, Canada are used to test these ideas. Mathematical goodness-of-fit criteria best estimate flood discharge for a broad range of flood recurrence intervals, i.e., 1.25, 2, 5, 10, 25, 50, and 100 years. The LN, EV1, LP3, and GEV distributions are found most appropriate in 42.5%, 31.9%, 21.7%, and 3.9% of cases, respectively, suggesting that systematic model selection criterion is required for FFA in heterogeneous landscapes. Multi-variate regression of estimated flood quantiles with backward elimination of explanatory variables using principal component and discriminant analyses reveal that precipitation provides a greater predictive relationship for more frequent flood events, whereas surficial geology demonstrates more predictive ability for high magnitude, less frequent flood events. In this study, all seven flood quantiles identify a statistically significant two-predictor model that incorporates upstream drainage area and the percentage of naturalized landscape *with 5% improvement* in predictive power over the commonly used single-variable drainage area model ($p < 2.2e\text{-}16$). An analysis of variance (ANOVA) further supports the two-predictor model indicating a decrease in the sum of squares of residuals and an F statistic ($p < 0.001$) that demonstrates very strong evidence in favour of the two-predictor model (i.e., drainage area and land use) when estimating flood discharge in this low-relief landscape with pronounced glacial legacy effects and heterogenous land use.

Keywords: flood frequency, FFA, RFFA, multi-variate modelling, land use, glacial conditioning

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

1. Introduction

A reliable assessment of flood frequency and flood magnitude over space and time is critical for urban planning and infrastructure engineering that depends on flood probability (Basso et al., 2016). Flood magnitude, frequency, and duration are primary drivers of channel erosion and stream morphology (Taniguchi & Biggs, 2015) as a self-shaping alluvial channel entrains and transports sediment to adjust its dimensions, planform pattern, bed characteristics, and gradient in response to varying flow levels (Church & Ferguson, 2015). So reliable estimates of flood frequency are important for understanding geomorphic channel change. Flood frequency analysis (FFA) is widely used to estimate how often a specified flood event (or channel discharge) will occur (Farooq et al., 2018). Most often, an FFA uses the occurrence of extreme flood events to estimate the return period, T, of flood quantiles, *Q(T)*, based on long-term, historic flow data from a gauge station. This probabilistic approach "fits" the site-specific data to a statistical distribution to predict the likelihood of future flood events. To provide flexibility of fit, statistical probability distributions require two to four parameters (Zhang et al., 2020). The choice of the probabilistic model that best represents the observed data and the estimation of a distribution's parameters affects the reliability of flood prediction (Laio et al., 2009; Cunnane, 1973; Farooq et al., 2018). Poor model application and fit can lead to unreliable estimates (Basso et al., 2016). The Generalized Extreme Value (GEV) distribution, Gumbel Maximum or Extreme Value Type I (EV1) distribution, Log-Normal (LN) distribution, and Log-Pearson Type III (LP3) distribution have traditionally been recommended to characterize flood probability based on goodness-of-fit (Onen & Bagatur, 2017; Laio et al., 2009). *T*he LP3 and GEV distributions use three parameters, i.e., location, scale, and shape, and the EV1 and LN distributions uses two parameters, i.e., location and scale, to fit data distributions (see Appendix A). In Canada, it is recommended that FFA studies draw from the Normal, GEV, and Pearson distribution families. Distribution fitting with more than three parameters is not recommended due to the limited record lengths of Canadian gauge stations (Natural Resources Canada (NRC), 2019). Often, flood estimation will apply a fixed probabilistic model to historical gauge data (Di Baldassarre et al., 2009). For example, since 1967 the U.S. Geological Survey (USGS) Bulletin-17C, Guidelines for Determining Flood Flow Frequency, recommend the use of the log-Pearson Type III (LP3) distribution as an appropriate statistical distribution to characterize the probabilities of annual flood series. However, for regions with diverse flood characteristics, multiple distributions may apply for different catchments and site specific selections are often recommended (Zhang et al., 2020). The most recent U.S. Army Corps of Engineers Hydrologic Engineering Center Statistical Software Package (HEC-SSP, Version 2.2, June 2019) includes the ability to perform two goodness-of-fit tests for up to 19 statistical distributions (US Army Corps of Engineers, 2019). Recent research indicates that estimation of flood frequency and magnitude improves with the application of a systematic and objective model selection criteria when fitting observed flow data to a statistical probabilistic curve (Di Baldassarre et al., 2009).

A regional flood frequency analysis (RFFA) can be very important in determining the probability of extreme flood events where streamflow data are not readily available (Ahn & Palmer, 2016). An RFFA transfers observed hydrologic information from a group of gauged sites to comparative ungauged sites as a representation of flow statistics using hydrological variables (Odry & Arnaud, 2017). A common approach to RFFA consolidates data samples from many measuring sites and uses ordinary

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

least-squares (OLS) regression to identify a relationship between mean annual floods of multiple basins and some basin characteristic (e.g., drainage area). It has become an accepted practice to model discharge using a single-variable power-law relationship between discharge (Q) and drainage area (A_d) of the form

$$Q = \alpha A_d^{\beta} \tag{1}$$

where A_d is the upstream drainage area and the coefficient α and exponent β are empirically derived by statistical regression (Dunne & Leopold, 1978, p. 818). This power relationship can be rewritten as

$$\log Q = \log \alpha + \beta \log A_d \tag{2}$$

The reliability of this single-variable predictive relationship, however, relies on the relative regional homogeneity, with similar basin conditions and climate present (Ahn & Palmer, 2016; Hosking & Wallis, 1993; Phillips & Desloges, 2014). Research suggests that the spatial variability of basin attributes (i.e., topographic relief, climate, vegetation, and land use) and the identification of subsurface characteristics which influence hydrological and fluvial function are controlling factors of a fluvial system's drainage efficiency and relevant to the flow response in a catchment (Di Lazzaro et al., 2015; Oudin et al., 2008).

Landscape modifications that decrease infiltration will impose changes to river hydrology (Ghunowa et al., 2021; Ashmore, 2015; Taniguchi & Biggs, 2015; Winter, 2001) with a downstream cascading effect on flow regime (Royall, 2013). Human occupation, landscape manipulation, and the generation of impervious surfaces associated with urbanization *have* the most profound impact on hydrogeomorphic responses, particularly in smaller watersheds (Pasternack, 2013; Royall, 2013). And a fluvial system's response to human-induced land use change (or its sensitivity to change) will vary, depending on basin attributes (i.e., configuration, geomorphology, and sediment retention) (Royall, 2013). For this reason, the spatial heterogeneity across a landscape will likely produce a variation in flood response that may best be captured using an RFFA approach with multi-variate analysis that considers relevant parameterized basin characteristics (i.e., topographic relief, land use, vegetation, and subsurface geology) as a set of explanatory variables to estimate flood discharge (Ahn & Palmer, 2016).

Recent works have highlighted the impact of geomorphic spatial heterogeneity on the basin hydrologic response (Ahn & Palmer, 2016; Di Lazzaro et al., 2015; Taniguchi & Biggs, 2015). *To better understand* the link between intra-catchment variability and hydrological function, this study has four objectives:

1) To complete an FFA in a heterogeneous landscape that models a reliable estimation of discharge for a broad range of flood recurrence intervals (i.e., $Q_{1.25}$, Q_2, Q_5, Q_{10}, Q_{25}, Q_{50}, and Q_{100}). Model selection *is determined* by applying systematic and objective model selection criteria to optimize model fit to long term site-specific flow data (T > 10 years). A test sample of 207 individual gauge sites within a glacially conditioned regional setting is used.
2) To derive the commonly used single-variable RFFA (Eq. 1) for the test region that characterizes the relationship between discharge (Q) and site-specific drainage area (A_d) using optimized estimates of a broad range of flood quantiles to test the predictive power of a single hydrologic variable in a glacially conditioned landscape.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

3) To present a multi-variate, regression based RFFA that considers the spatially variability of hydrologic controls in the context of inherited glacial landforms, sediments, and land use. To achieve this goal, *t*wenty-eight (28) explanatory variables are explored to represent basin characteristics (i.e., topographic relief, climate, land use, vegetation, and subsurface geology). To identify the most parsimonious discharge models for recurrence intervals of 1.25, 2, 5, 10, 25, 50, and 100 years where backwards elimination of explanatory variables is employed in principal component and discriminant analyses.
4) To compare the predictive power of a multi-variate derived RFFA that considers multiple basin hydrologic controls to a generally accepted single-variable RFFA in a glacially conditioned setting.

2. **Regional Setting**

This flood frequency study focuses on a test region of peninsular southern Ontario, Canada (Figure 1) that is bounded by the Canadian Shield to the north, the three lower Great Lakes, Huron, Erie, and Ontario to the southwest and the Ottawa and St. Lawrence Rivers to the east. Located within the North American Great Lakes watershed, it is a region of modest relief, with elevation ranging from 544 m asl near Lake Huron draining by way of the St. Lawrence River lowlands at less than 70 m to the Atlantic Ocean (Larson & Schaetzl, 2001). Convective, synoptic, and tropical systems that influence the humid, continental climate of the region are enhanced by local, regional, and topographic conditions (Paixao et al., 2011). Moisture and temperature associated with the Great Lakes influence inland precipitation for up to 50 km. Consequently, the mean annual precipitation varies regionally from 800 mm to 1200 mm (Paixao et al., 2011). During winter months, precipitation typically accumulates in the form of snow, generating spring snowmelt floods that dominate river flow regimes (Javelle et al., 2003). The surficial geology of the region, and the hydrologic controls exerted by the parent materials, are the product of the region's glacial history (Chapman & Putnam, 1984). Recurring continental glaciations over the last ~2 million years have topographically influenced the fluvial drainage networks of southern Ontario (Desloges et al., 2020; Fulton et al., 1986).

Deglaciation, approximately 12 to 13 thousand years ago, has left pronounced glacial legacy effects with complex sequences of subglacial, ice-contact, and proglacial sediments deposited during the final retreat of the Laurentide Ice Sheet (Phillips & Desloges, 2014, 2015; Larson & Schaetzl, 2001). The most common physiographic features include sheets of till, finer glaciolacustrine plains of sand or clay, glaciofluvial outwash deposits of sand, gravel, silts and clays, and a configuration of moraines (Thayer et al., 2016). Two significant post-glacial geomorphic features are the Niagara Escarpment and the Oak Ridges Moraine (Figure 1). The Niagara Escarpment is a Paleozoic limestone bedrock ridge resulting from differential glacial erosion and weathering of harder and softer rock that arches from the region between Lakes Ontario and Erie, bypassing Lake Ontario and extending northward to Georgian Bay (Chapman & Putnam, 1984; Phillips & Desloges, 2014).

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

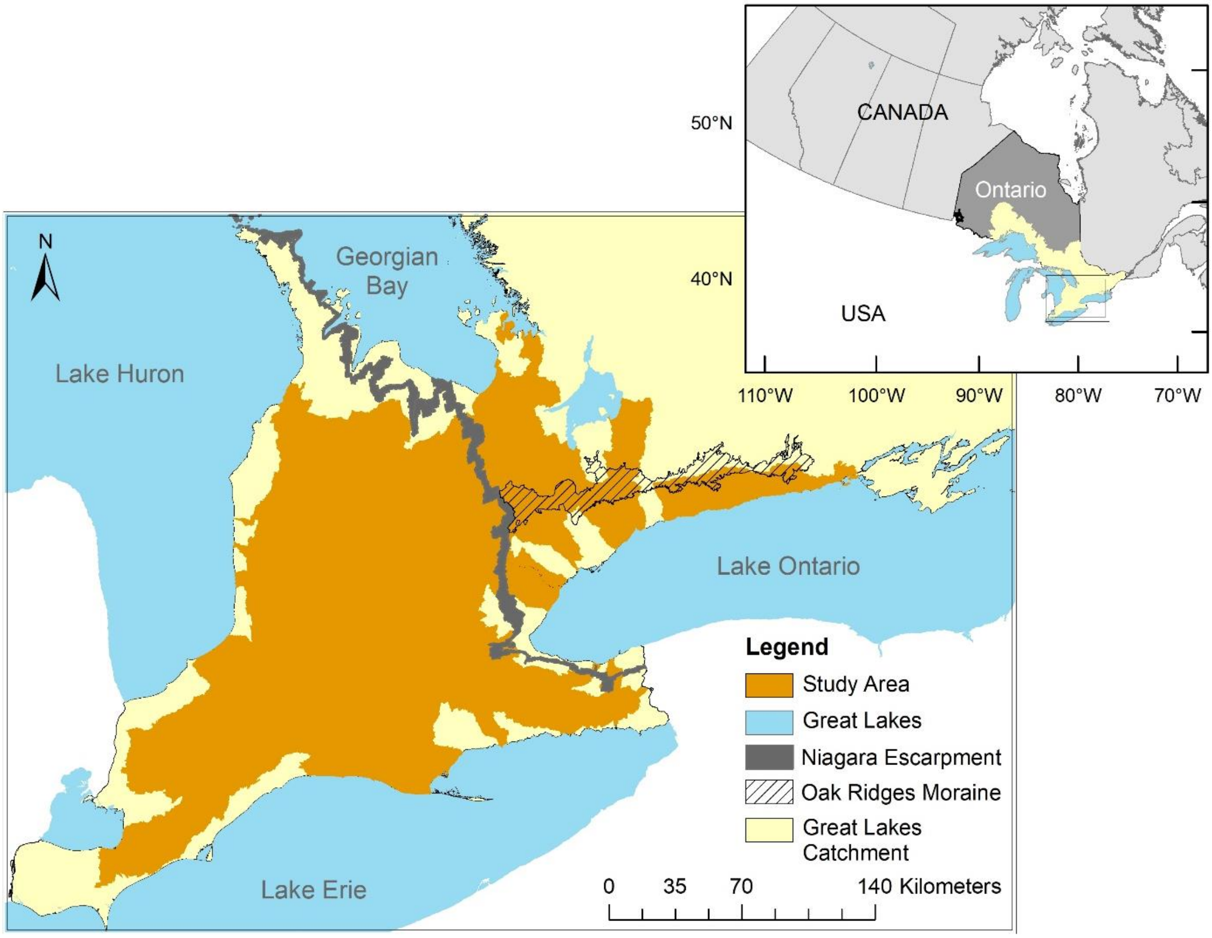

Figure 1 - Map identifying the study area (indicated in orange) and two significant post-glacial geomorphic features that influence drainage networks of southern Ontario (i.e., the Niagara Escarpment and the Oak Ridges Moraine). The inset map (upper right) indicates the study region within the Ontario portion of the Laurentian Great Lakes catchment relative to Canada.

Several preglacial rivers have carved deep valleys into the Niagara Escarpment, however, Late Pleistocene glaciations have infilled these valleys with varying thicknesses of till (Chapman & Putnam, 1984) directing catchment flow mostly away from the escarpment crest. The Oak Ridges Moraine is a stratified kame moraine of glacial drift that extends from the Niagara Escarpment 160 km eastward across south-central Ontario (Phillips & Desloges, 2014). This massive ridge forms a drainage divide, separating catchments flowing north to Georgian Bay/Lake Huron and south to Lake Ontario. Glacial sediments typically blanket the study area at a thickness of 50 m, and up to 350 m in some places (Larson & Schaetzl, 2001). In many areas, where stratified limestones and shales of the Palaeozoic age lie beneath the thick glacial overburden, fertile soils rich in calcium carbonate and clay are produced (Phillips & Desloges, 2014, 2015; Desloges et al., 2020). These support southern Ontario's widespread agricultural development.

More recent European settlement and regional expansion have resulted in differentiated land use with extensive agricultural land, natural and reforested areas, and clustered urban settlement (Chapman & Putnam, 1984). The southern Ontario region

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

continues to accommodate an increasing population. Drawn by employment, most settle in built-up cities and surrounding areas, driving clustered regional urbanization that consumes surrounding rural lands. However, a comparable demand to expand the total area of cropland has also occurred to support larger farming operations (Donnan, 2008).

3. Methods and Data Collection

140 An overview of the methodology for this study is provided in Figure 2.

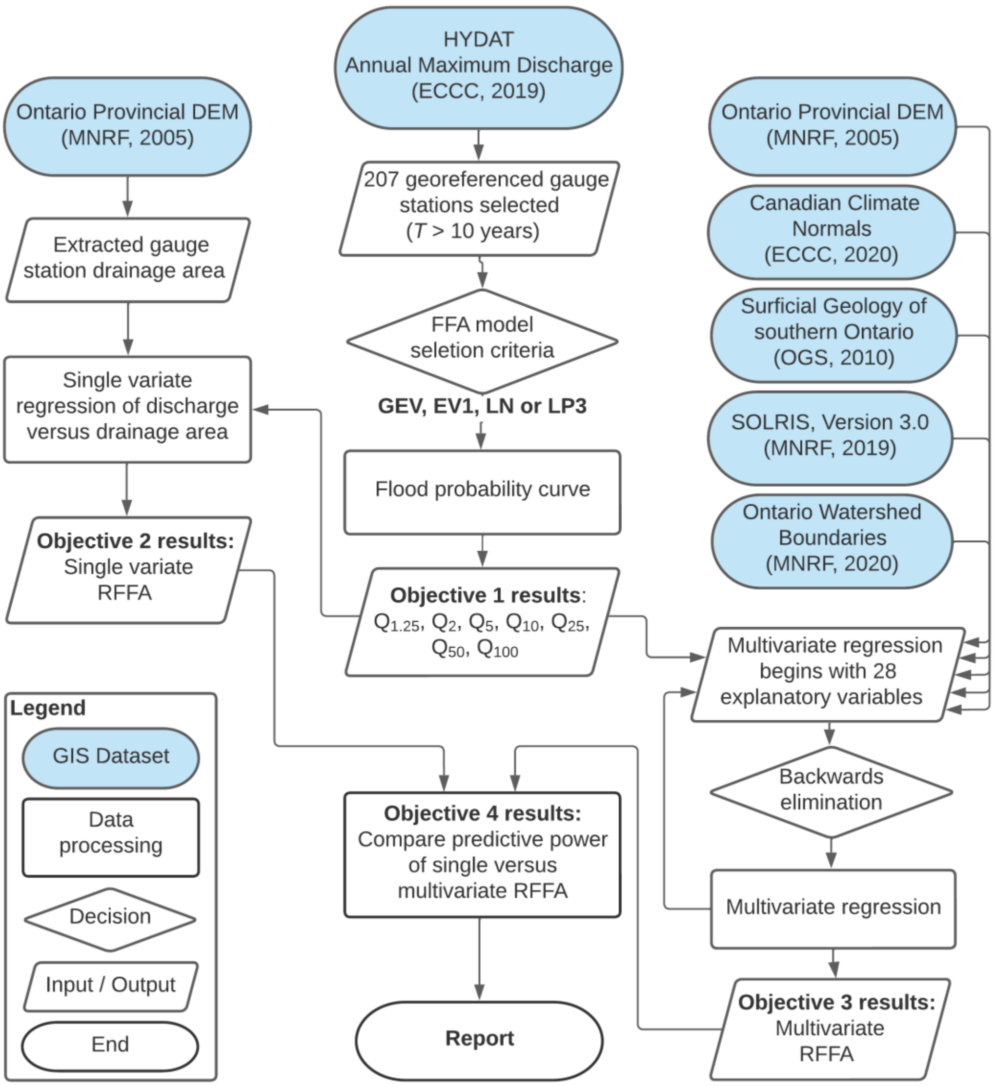

Figure 2 - Flow chart of FFA and comparison of common and multi-variate RFFA. The single-variable RFFA uses a discharge to drainage area relationship. The regression-based multi-variate RFFA employs sub-basin characterization and backward elimination of explanatory variables to determine the most parsimonious model to predict discharge over seven (7) flood quantiles.

3.1 Flood frequency analysis

A Station Meta Data Index for 1188 Ontario stream gauges from the HYDAT database of the Water Survey of Canada (WSC) monitoring program is accessed online at https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html (Environment and Climate Change Canada (ECCC), 2019). The quality of gauge data depends on the selected measurement techniques, computation methods, and physical conditions at the monitoring sites (i.e., ice and other influences). However, the WSC performs regular audits of field activities and adheres to standard operating procedures to improve data quality (ECCC, 2019). Gauge locations are sorted by catchment and synthesized to identify gauges specific to the southern Ontario region.

Retention of station data is based on three criteria: (1) the gauge station lies within the peninsular region of southern Ontario (2) the gauge station exists for a fluvial system with known field survey data (i.e., Annable (1995, 1996) and Phillips (2014)), and (3) streamflow data represent a minimum of 10 years of operation, continuous (non-seasonal) year-round operation. These criteria yield 207 gauge stations within the study area. Although some research suggests that the instantaneous maximum discharge may command greater geomorphic significance, the mean daily discharges provide a larger dataset with fewer gaps in the discharge records. The annual maximum mean daily discharge (m^3/s) for each of the gauge stations is used for distribution fitting and flood recurrence computations. The annual maxima series approach uses the highest annual discharge from the recorded mean daily discharge values at a gauge, ensuring statistical independence of observations between years. This approach has been shown to be more efficient than the partial duration series approach that uses floods that exceed some base threshold discharge (q_0), regardless of the time distribution (Cunnane, 1973). The MSClaio2008 R function, part of the package nsRFA in R, is used to compare the LP3, EV1, GEV, and LN distributions to the annual maximum discharge data for each of the 207 gauge stations and determine the most appropriate statistical distribution, or model. No prior processing is implemented to fit the distributions. Four criteria determine goodness-of-fit: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Anderson-Darling Criterion (ADC), and a second-order variant of AIC, the Corrected AIC (AIC_C). These model selection criteria are shown to provide good operational strategy when applied to frequency analysis of hydrological extremes (Laio et al., 2009). To identify the distribution that best represents the data, the candidate models are assessed in the form of probability distributions with parameters estimated using the maximum likelihood method. For each of the 207 gauges, the selected optimal distribution for each gauge dataset is used to model flood recurrence using the cumulative probability for each model. Flood quantiles for seven recurrence intervals (RIs) of 1.25, 2, 5, 10, 25, 50, and 100 years (i.e., $Q_{1.25}$, Q_2, Q_5, Q_{10}, Q_{25}, Q_{50}, Q_{100}) are derived directly from individual gauge data and, therefore, reflect the upstream conditions of the corresponding drainage basin.

3.2 Single-variable regional flood frequency analysis

The site specific drainage area for each gauge station is evaluated based on Ontario's Provincial Digital Elevation Model (DEM) – Version 2.0.0 (OMNR, 2005). The provincial DEM – Version 2.0.0 is a hydrologically enforced tiled raster dataset

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences Discussions

with a 10 m cell resolution and 5 m vertical accuracy. The accepted single-variable relationship (Eq. 1) between discharge and drainage area is obtained by statistical regression identifying a hydrologic relationship for each of the seven flood quantiles.

3.3 Multi-variate regional flood frequency analysis

Catchment basins of the study area are delineated based on the hierarchical framework of the Ontario Watershed Boundaries (OWB). The digital geospatial datasets, published by the Ontario Ministry of Natural Resources and Forestry (OMNRF), are accessed online from https://data.ontario.ca/dataset/ontario-watershed-boundaries. Basins are first identified according to Tertiary level watersheds. To characterize the upstream conditions affecting channel discharge, further subdivision is influenced by the Quaternary level boundaries. The watershed boundaries are accurate to within 100 m (OMNRF, 2020). All mapping and spatial analysis uses a combination of standard GIS software. Maps are projected to the Universal Transverse Mercator (UTM, Zone 17N), referenced to the North American Datum (NAD1983).

Twenty-eight (28) basin attributes are selected to characterize the drainage area conditions representing the land use, precipitation, topography, and hydrological properties from a geomorphological perspective within the basins. Quantification of sub-basin characteristics is computed using digital geomatic mapping from multiple sources:

a) The contributing upstream drainage area and the topographic conditions are captured using Ontario's Provincial DEM – Version 2.0.0.

b) The surficial geology is characterized from the revised Surficial Geology of Southern Ontario (MRD 128–Revised) of the Ontario Geological Survey (OGS, 2010). Accessed online at http://www.geologyontario.mndm.gov.on.ca/mndmaccess/mndm_dir.asp?type=pub&id=MRD128-REV, the digital geospatial dataset provides a seamless, standardized map of the geology, primary material, genesis, and formation coverages for southern Ontario.

c) Regional land use is based on the southern Ontario Land Resource Information System (SOLRIS) Version 3.0, a comprehensive, digital landscape level inventory published by the OMNRF that identifies urban, rural, and natural features at a 15 m resolution (OMNRF, 2019). The SOLRIS digital geospatial dataset was derived from Landsat-8 OLI imagery acquired from 2014 to 2017 and is accessed online at https://geohub.lio.gov.on.ca/datasets/southern-ontario-land-resource-information-system-solris-3-0 .

d) The precipitation patterns for southern Ontario are evaluated based on the Canadian Climate Normals 1981-2010. Climate Normals are commonly used to assess regional climate. The Canadian Climate Normals adhere to the accepted standards of the World Meteorological Organization which recommends 30-year records to eliminate year to year variation (Environment Climate Change Canada (ECCC), 2020). Rainfall and precipitation (including snowfall water equivalent) data for 65 observation stations are accessed at https://climate.weather.gc.ca/climate_normals.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

A relationship between the explanatory variables and each of the quantile discharge datasets is assessed by applying OLS regression. OLS assumes that the set of explanatory variables (i.e., basin characteristics) and errors must be independent to avoid bias. To identify the final explanatory variables for regression, discriminant analysis is applied to assess covariance and the duplication of information among variables. Where high correlation between variables is identified, the variable with the weakest theoretical association to channel discharge is removed. Employing backward elimination, regression analysis is used to identify the most parsimonious discharge models for RIs of 1.25, 2, 5, 10, 25, 50, and 100 years, while also considering the residual error of models.

3.4 Comparison of single-variable versus multi-variate RFFA

To assess the predictive power of the single variable RFFA versus the multi-variate RFFA, a comparison is made based on model scatter, residual error, and an analysis of model variance.

4. **Results**

4.1 Selection of statistical distributions and establishing flood quantiles

This study subdivides 16 tertiary level catchments of southern Ontario into 45 sub-watersheds for the purpose of establishing regional hydrologic, geomorphic, and land use conditions. Gauge stations are clustered within sub-basin units to best represent the immediate upstream hydrological conditions influencing the channel discharge response at each gauge station. Analysis indicates the hydrometric records of the 207 gauge stations have a minimum operation period of 10 years, an average of 42.5 years (+/- 1.7 years, median = 42 years), and a maximum operation period of 106 years. For each flood dataset, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Anderson-Darling Criterion (ADC) are applied to each of four candidate models (i.e., GEV, EV1, LN, and LP3) to evaluate model fit. Where the sample size, n, is small with respect to the number of estimated parameters, p, such that $n/p < 40$, the corrected Akaike Information Criterion (AICc) is also applied. The criteria enable a systematic and objective mathematical test of model fit. Figure 3 illustrates the delineation of sub-basins and clustering of gauge stations and the results of the "best fit" statistical model testing. The model selection criteria determines that 42.5% of the 207 hydrometric records are most suited to an LN distribution, 31.9% to an EV1 distribution, 21.7% to an LP3 distribution, and 3.9% to a GEV distribution suggesting all four distributions tested are potentially suitable for modelling flood extremes in southern Ontario. For 74.4% of the gauge records tested, the model selection criteria chose a 2-parameter model (i.e., LN or EV1) over a 3-parameter model (i.e., LP3 or GEV). The 2-parameter EV1 model is found to be five (5) times more likely to be selected as the optimal distribution over its 3-parameter parent model, GEV. The GEV distribution is only selected in a limited number of cases. Generally, within the areal extent of a sub-basin unit, multiple statistical distribution types are identified as optimal with no "best fit" distribution type indicated based on geographic location.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

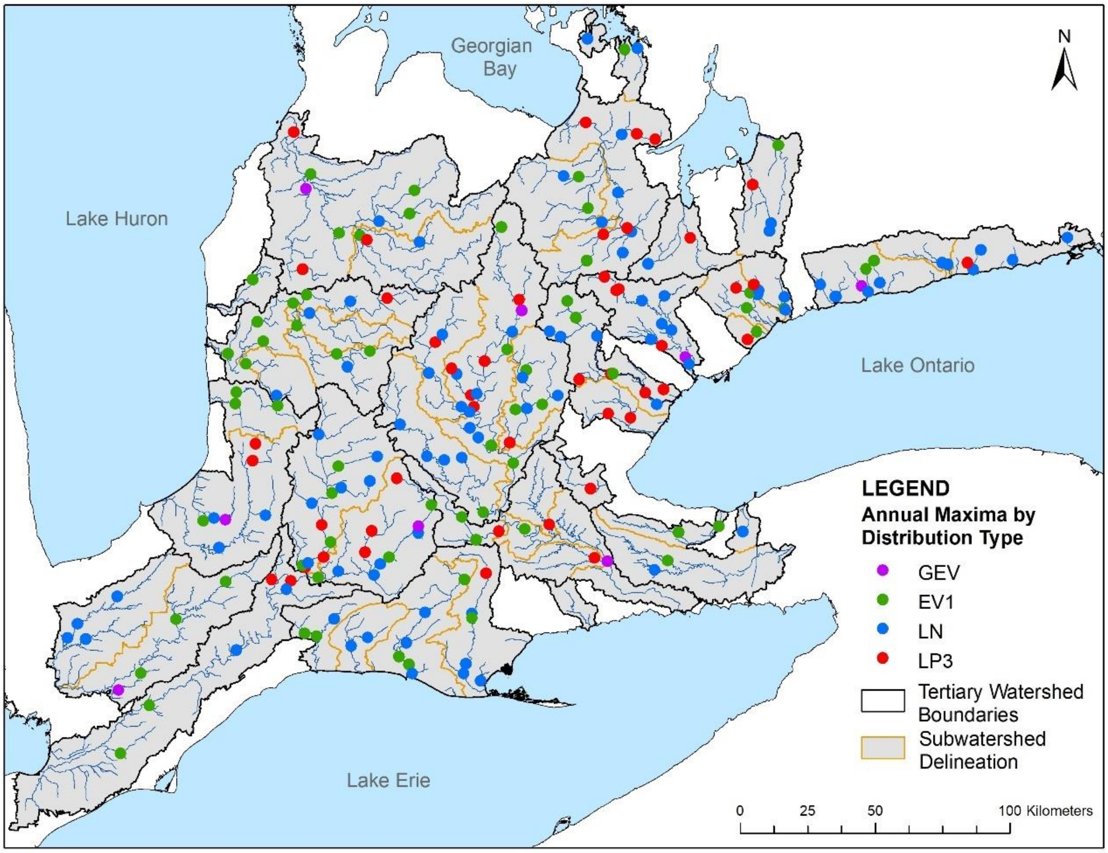

Figure 3 - Map identifying the geographic location of 207 gauge stations and spatial variability of the optimized statistical FFA distribution type based on the AMS data fit. Subdivisions of the tertiary level watershed boundaries are indicated. No sub-basin indicates a particular "best fit" distribution type. Forty-five (45) sub-watersheds are established to characterize basin conditions that represent the region's variable land use, precipitation, topography, and hydrological properties from a geomorphic perspective.

The selected optimal probability distribution curve for each gauge is used to estimate the flood quantiles for RIs of 1.25, 2, 5, 10, 25, 50 and 100 years for each one of the 207 gauge stations. These flood quantiles are consistent with return periods explored in other flood frequency analyses (Ahn & Palmer, 2016; Basso et al., 2016; Onen & Bagatur, 2017; Hollis, 1975). A Shapiro-Wilk analysis tests the null hypothesis that the flood quantile datasets are normally distributed (Table 1). The results (Test 1) indicate that the dataset for each flood quantile does not meet the assumption of normality ($p < 0.05$) and the null hypothesis is rejected. A logarithmic transformation is applied to all flood quantile values. The results of a Shapiro-Wilk test for each of the log-transformed flood quantile datasets fails to reject the null hypothesis ($p > 0.05$) suggesting the log-transformed flood quantiles are normally distributed (Test 2).

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences Discussions

Table 1 – Results of Shapiro-Wilks Normality Tests for each flood quantile, with and without logarithmic transformation of data

Test 1 – before transformation			Test 2 – after logarithmic transformation		
Flood quantile	W-stat	p-value	Log-transformed flood quantile	W-stat	p-value
$Q_{1.25}$	0.578	< 2.2e-16	Log $Q_{1.25}$	0.991	0.200
Q_2	0.580	< 2.2e-16	Log Q_2	0.992	0.299
Q_5	0.595	< 2.2e-16	Log Q_5	0.991	0.194
Q_{10}	0.604	< 2.2e-16	Log Q_{10}	0.990	0.135
Q_{25}	0.635	< 2.2e-16	Log Q_{25}	0.991	0.260
Q_{50}	0.634	< 2.2e-16	Log Q_{50}	0.992	0.351
Q_{100}	0.595	< 2.2e-16	Log Q_{100}	0.991	0.247

4.2 Single-variate regression RFFA

The upstream drainage area for each georeferenced gauge station is extracted from the hydrologically enforced DEM. Hydrological enforcement ensures that drainage occurs in a down-slope direction, facilitating the construction of a flow accumulation raster necessary to establish the upstream drainage area. A logarithmic transformation is applied to the drainage area variable values to ensure normality (W = 0.994, p-value = 0.522).

Regression of the logDrainage variable against each of the seven flood quantile datasets (i.e., $Q_{1.25}$, Q_2, Q_5, Q_{10}, Q_{25}, Q_{50}, Q_{100}) establishes seven single-variable power relationships (Table 2). Strongly significant logDrainage area relations are indicated across all RI's with a minor, but consistent, decrease in adjusted R2 values as RI increases.

Table 2 – Single-variate RFFA models for each flood quantile

Flood Quantile	Single Variable Models (i.e., logDrainage)		
	Equation	Residual SE	adjusted R^2
$Q_{1.25}$	$\log Q_{1.25} = -0.858 + 0.945(\text{logDrainage})$	0.217	0.867
Q_2	**$\log Q_2 = -0.746 + 0.957(\text{logDrainage})$**	**0.224**	**0.862**
Q_5	$\log Q_5 = -0.549 + 0.945(\text{logDrainage})$	0.219	0.864
Q_{10}	$\log Q_{10} = -0.384 + 0.917(\text{logDrainage})$	0.233	0.842
Q_{25}	$\log Q_{25} = -0.223 + 0.906(\text{logDrainage})$	0.245	0.824
Q_{50}	$\log Q_{50} = -0.084 + 0.885(\text{logDrainage})$	0.278	0.776
Q_{100}	$\log Q_{100} = -0.053 + 0.864(\text{logDrainage})$	0.321	0.713

Research indicates that the Q_2 flood quantile (highlighted in Table 2) represents a flow magnitude and frequency that is important to the maintenance of channel morphology and has, therefore, been used in a discharge- drainage area relation in numerous other studies of the southern Ontario region (Annable et al., 2011; Vocal Ferencevic & Ashmore, 2012; Phillips & Desloges, 2014; Thayer et al., 2016). Expressing the Q_2 results from Table 2 in a power-law format (Eq. 1), the Q_2 model is found to be similar to the findings of other southern Ontario models (Figure 4). The findings of Phillips and Desloges (2014) and Annable (1995) were similarly derived from annual maximum series datasets of the southern Ontario region. The Q_2 power

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

relationship identified in this study indicates a slightly lower estimate of Q_2 discharge for smaller drainage areas (<100 km^2) compared to the research of others. For larger drainage areas (>100 km^2), this study predicts similar discharge estimates compared to the relationship of Phillips & Desloges (2014) but greater discharge estimates than Annable (1995). Neither of those studies specified best-fit RFFA distributions so the relationship presented here is considered more robust.

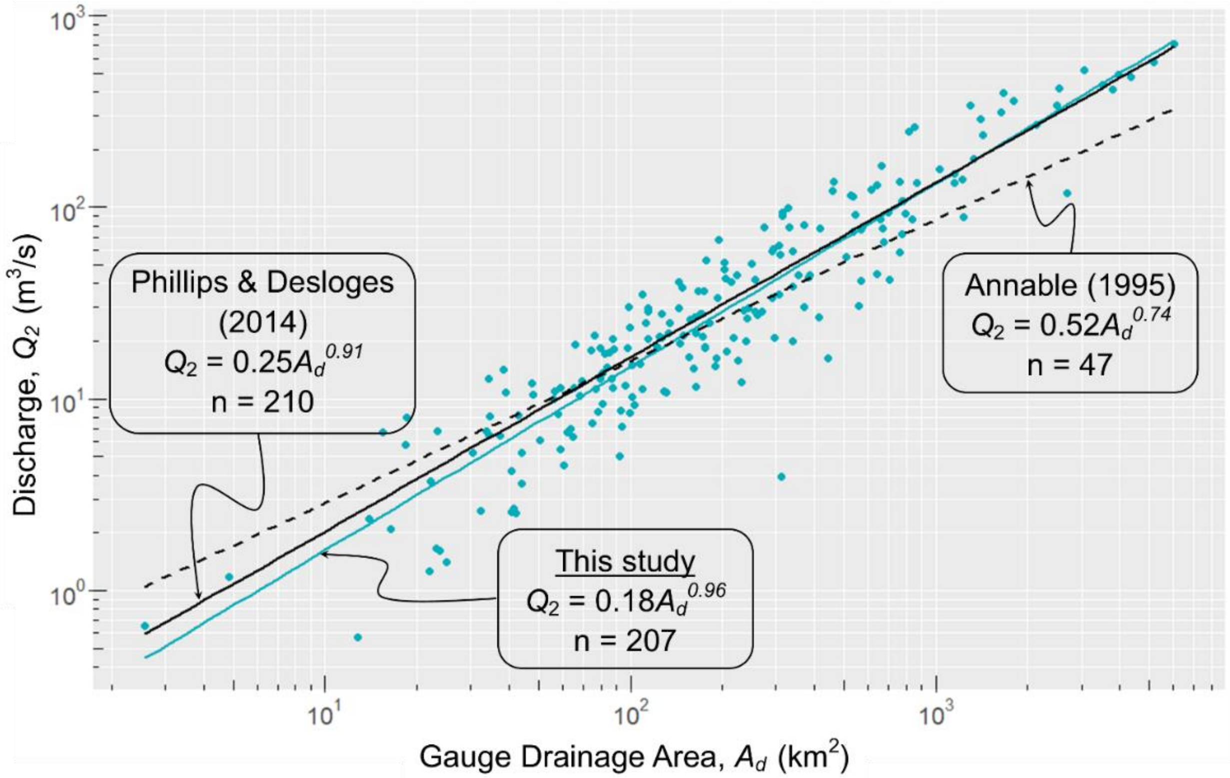

Figure 4 – The single-variable discharge-drainage area relationship power relationship for the Q_2 flood quantile of this study (in blue) compared to the findings of others. The Q_2 relationship of Annable (1995) is indicated by the dashed black line. The Q_2 relationship of Phillips and Desloges (2014) is given by the solid black line.

4.3 Multi-variate regression RFFA with parameterized basin characteristics

4.3.1. Regression model inputs

Geomorphic, hydrologic, land use and topographic basin attributes are related to best-fit RFFA model results from all 207 gauges/basins. Sub-divided basin characteristic variables are extracted from four geospatial raster datasets (Figure 5) using cell counts and zonal statistics. The geomorphic basin attributes are represented by the percentage of the dominant surficial material within the geographic area of each sub-basin (Figure 5(a)). The hydrological conditions are characterized by an interpolation of Canadian Climate Normals (Figure 5(b)). Point information for mean annual precipitation, annual number of precipitation days, mean annual rainfall and annual number of rainfall days from 65 observation stations is converted to raster coverage using several interpolation techniques. Inverse distance weighting (IDW) and ordinary kriging (OK) using a stable

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

model and an exponential model are compared. OK has been shown to produce accurate results when used to describe spatially heterogeneous natural phenomena (Bevan & Conolly, 2009) such as precipitation. The results of cross validation suggest fitting an OK exponential model for annual mean precipitation, annual mean rainfall, and the annual number of rainfall days, and an OK stable model for the annual number of precipitation days.

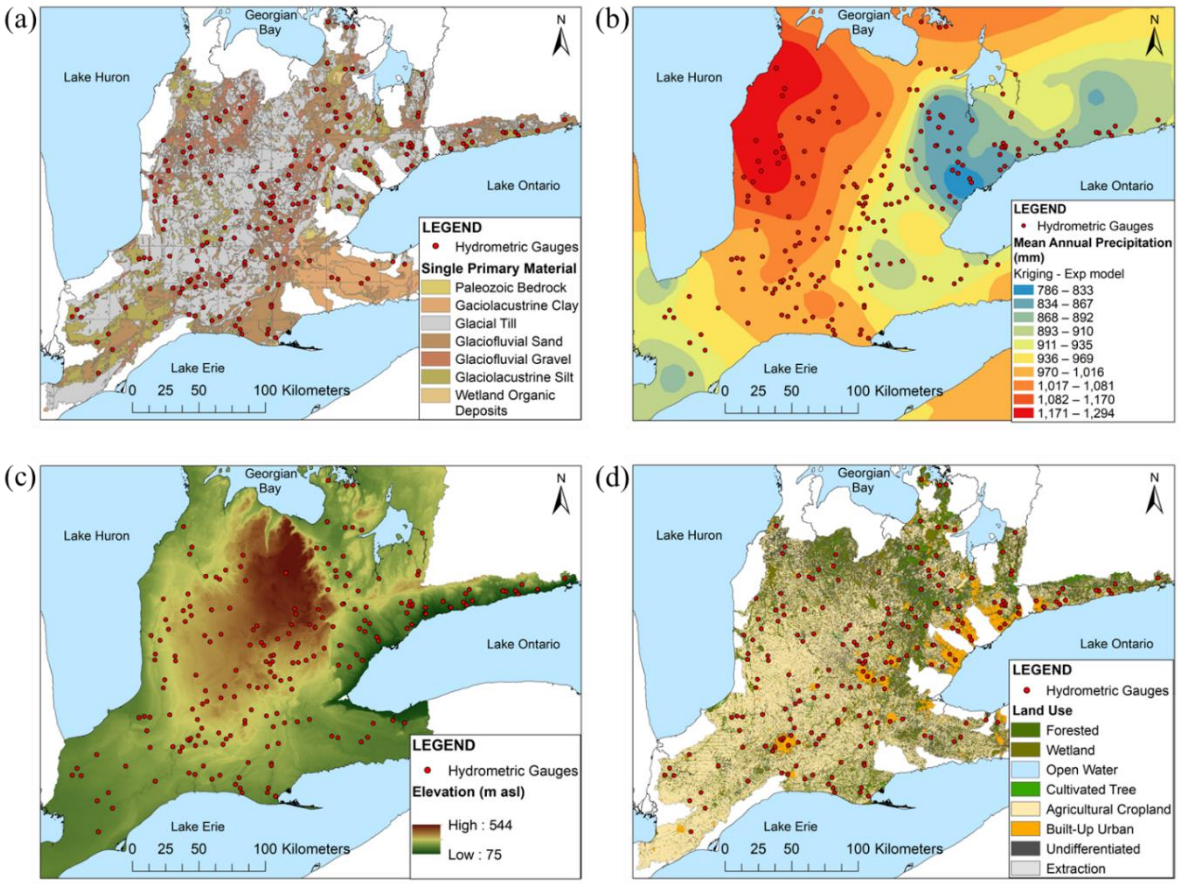

Figure 5 - Raster datasets illustrating (a) the dominant surficial material, (b) the mean annual precipitation, (c) the hydrologically enforced DEM, and (d) the categorical land use for the peninsular region of southern Ontario.

The topographic conditions of the sub-basins are extracted and quantified from the hydrologically enforced DEM (Figure 5(c)). Percent land use is quantified using the SOLRIS categories for each sub-basin (Figure 5(d)). For this study, three land use categories are established: percent urban, percent agricultural cropland, and percent naturalized area. Urban regions combine all transportation and built-up areas. Agricultural cropland is defined by tilled areas. Naturalized regions combine all tallgrass landcover, mixed forests, cultivated tree plantations, swamps, wetlands, and open water areas as indicated by the SOLRIS Version 3.0 dataset.

For each predictor variable, a single output value is produced and applied to the gauge(s) within a sub-basin. Since many sub-basins contain more than one gauge station, some gauges share the same topographic, geomorphic, climate, and land use values

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

but possess their own unique drainage area value. In total, 28 predictor variables are quantified (Table 3). The characteristics selected are considered to represent independent variables that influence channel discharge (across all 7 flood frequency quantiles) in terms of the regional geomorphic, hydrologic, topographic and land use properties.

Table 3 - Twenty-eight (28) sub-watershed characteristics to represent geomorphic, hydrologic, land use, and topographic variability between sub-basins.

No.	SUB-BASIN VARIABLES	ABBREVIATION	FILE TYPE	SOURCE
Geomorphic Variables: Dominant Surficial Material				
1	Glaciolacustrine Clay	Clay	Raster Dataset	OGS, 2010
2	Glacial Till	Daimicton	Raster Dataset	OGS, 2010
3	Glaciofluvial/Glaciolacustrine Gravel	Gravel	Raster Dataset	OGS, 2010
4	Wetland	Organic	Raster Dataset	OGS, 2010
5	Paleozoic or Precambrian Bedrock	Bedrock	Raster Dataset	OGS, 2010
6	Glaciofluvial/Fluvial/Glaciolacustrine Sand	Sand	Raster Dataset	OGS, 2010
7	Glaciolacustrine/Fluvial Silt	Silt	Raster Dataset	OGS, 2010
Hydrologic Variables				
8	Gauge Drainage Area	logDrainage	Raster Dataset	OMNR, 2005
9	Mean Precipitation	Mean_Precip	Point Shapefile	ECCC, 2020
10	Precipitation Days	Precip_Days	Point Shapefile	ECCC, 2020
11	Mean Rainfall	Mean_Rainfall	Point Shapefile	ECCC, 2020
12	Rainfall Days	Rainfall_Days	Point Shapefile	ECCC, 2020
Land Use Variables				
13	% Urban Land Use	Urban	Raster Dataset	MNRF, 2019
14	% Tilled Cropland	Cropland	Raster Dataset	MNRF, 2019
15	% Naturalized Land Use	Naturalized	Raster Dataset	MNRF, 2019
Topographic Variables				
16	Gradient Mean	Gradient_Mean	Raster Dataset	OMNR, 2005
17	Gradient Standard Deviation	Gradient_StDev	Raster Dataset	OMNR, 2005
18	Aspect Mode	Aspect_Mode	Raster Dataset	OMNR, 2005
19	Stream Length	Stream_Length	Raster Dataset	OMNR, 2005
20	Drainage Density	Drainage_Density	Raster Dataset	OMNR, 2005
21	Sub-basin Area	WS_Area	Raster Dataset	OMNR, 2005
22	Sub-basin Perimeter	WS_Perimeter	Raster Dataset	OMNR, 2005
23	Sub-basin Compactness	WS_Compactness	Raster Dataset	OMNR, 2005
24	Minimum Elevation	Min_Elevation	Raster Dataset	OMNR, 2005
25	Maximum Elevation	Max_Elevation	Raster Dataset	OMNR, 2005
26	Elevation Range	Elev_Range	Raster Dataset	OMNR, 2005
27	Elevation Mean	Elev_Mean	Raster Dataset	OMNR, 2005
28	Elevation Standard Deviation	Elev_StDev	Raster Dataset	OMNR, 2005

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

4.3.2. Basin characterization parsimony

When characterizing natural systems, the potential exists for some variables to correlate with other variables due to their representation of related natural phenomena. This is often indicated by high correlations between variables suggesting a duplication of information captured (Ahn & Palmer, 2016; Phillips & Desloges, 2015). A simple Pearson correlation (|r| > 0.6) suggests that 29 correlated relationships exist among eleven basin characteristics including Diamicton and Sand, Cropland and Naturalized, Gradient_Mean, Gradient_StDev, Stream_Length, WS_Area, Mean_Rainfall, Min_elev, Max_elev, Elev_mean, and Elev_StDev.

Principal components analysis (PCA) has been shown to be an effective tool for variable reduction that provides a statistical basis to discard redundant variables (King & Jackson, 1999). A PCA of the 28 explanatory variables indicates that the first 3 dimensions account for almost 58% of the total variability of the dataset and are the most interpretable. This suggests an absence of strong intercorrelations among many of the 28 variables. The correlation circles in Figure 6 illustrate the projections of the first three principal components (Dim1, Dim2 and Dim3). The first principal component (Dim1) tends towards a land use composition grouping with some loading from gradient variables and precipitation variables (Dim1 explains 27.7% of the variance). The second principal component (Dim2) tends towards an elevation cluster with additional loading from precipitation variables (Dim2 explains 19.8% of the variance). The third principal component (Dim3) is a weakly defined land surface grouping with loading from surficial geology classifications and basin geometry (Dim3 explains 10.3% of the variance). While elevation is a clear contributor to the variance of the dataset based on the PCA tests, the directional indicators suggest the presence of multicollinearity among the elevation predictors, reinforcing the results of Pearson correlation detection. PCA and Pearson Correlation support elimination of the elevation variables (i.e., Min_elev, Max_elev, Elev_mean, Elev_StDev, Elev_Range) except for Gradient_Mean which is retained as the sole predictor to represent the variability of topography among the sub-basins. An observed correlation between Gradient_Mean and Gradient_StDev also results in the elimination of Gradient_StDev from the potential predictor variables. Other correlated variables are removed from further analysis due their weaker theoretical association, including WS_Area and Stream_Length in favour of retaining WS_Perimeter and Basin_Compactness. The elimination of these basin geometry variables enables greater focus on basin shape rather than basin size, concentrating instead on the efficiency with which a fluvial system can evacuate precipitation from the region. Mean_Rainfall is also eliminated due to correlation with multiple variables (i.e., Mean_Precip, Cropland, Gradient_StDev). By eliminating nine (9) sub-basin characteristics, twenty-seven (27) of the 29 correlated relationships previously identified by the Pearson correlation tests (i.e., criterion |r| > 0.6) are removed.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Open Access
Discussions

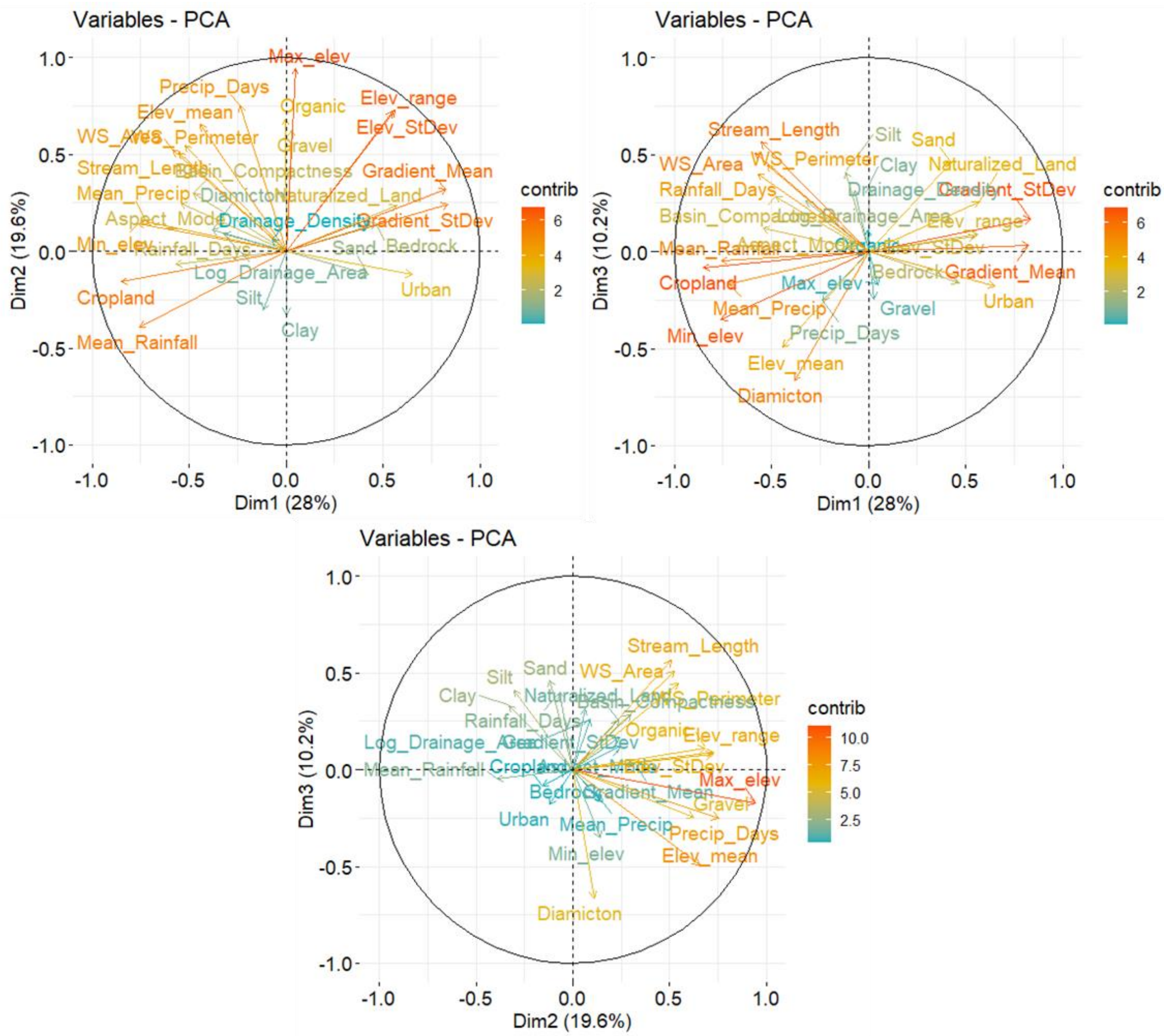

Figure 6 - Principal component analysis (PCA) highlighting the most contributing variables of the 28-variable dataset for each dimension and illustrating the correlation circles for principal components one, two, and three (Dim1, Dim2 and Dim3).

4.3.3. Multi-variate regression analysis with backward elimination to identify the most parsimonious models

It can be a practice to test and transform independent variables to ensure a normal distribution of a multi-variate dataset, however, tests for multi-variate normality are rarely performed (Tacq, 2010). Alternatively, the plots of standardized residuals from combinations of predictor variables are examined for a desired elliptically symmetric distribution. To enable model comparison, the gauge drainage area predictor variable included in the multi-variate regression is logarithmically transformed, consistent with the single-variable power model.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Multiple linear regression is applied to the remaining 19-variable dataset for each of the seven flood RIs, i.e., 1.25, 2, 5, 10, 25, 50 and 100 years, using an OLS approach. The fitted values of the model are compared to the dependent variables (i.e., $Q_{1.25}$, Q_2, Q_5, Q_{10}, Q_{25}, Q_{50}, and Q_{100}) to detect heteroscedasticity. Regression diagnostics are implemented to assess statistical significance or covariance among variables and ensure the model validity. While considering model error, a backward elimination strategy is applied for each RI to find the most parsimonious model. Predictor variables are eliminated based on t-tests and other information criterion statistics (i.e., examination of standardized residuals, added variable plots, variance inflation factors, and marginal model plots). F-tests indicate a linear association between a selected flood quantile and any of the predictor variables. Regression of all 19 predictor variables for each flood quantile reveals statistically significant relationships ($p < 2.2e\text{-}16$) suggesting at least one of the predictor variables is significantly related to the quantile discharge. The multiple coefficients of determination are greater than 0.8 ($R^2 > 0.8$) for all quantiles. Examination of the associated p-values for the t-statistic of each predictor variable indicates that, for all seven flood quantiles, logDrainage overwhelmingly contributes to the 19-variable prediction models (> 70%), although its importance decreases as flood frequency decreases. The 19-variable regression (log-drainage area included) indicates that either Naturalized or Gradient_mean is statistically significant for predicting flood quantiles $Q_{1.25}$, Q_2, Q_5, Q_{10} and Q_{25}. This is consistent with the results of the PCA which identifies the first principal component as a land use grouping and the second principal component as an elevation grouping. Other potential predictor variables do not indicate statistical significance in the 19-variable models. The retention of each variable is subject to regression diagnostics to ensure model validity and to identify the most parsimonious model. Added-variable plots (or partial regression plots) indicating low statistical significance are consistent with the t-statistic results confirming a lack of significance. Variables indicating no relationship are eliminated. Variance inflation factors (VIFs) are computed to further detect multicollinearity among variables. Factors exceeding 5 (VIF > 5) suggest high correlation and the variable with the weakest theoretical association to the dependent variable is eliminated. Contrasting geomorphic conditions between catchments (i.e., surficial geology) are represented by a high negative correlation ($|r| > 0.6$) between Sand or Diamicton. A high VIF (VIF > 5) confirms Diamicton is unsuitable as a predictor variable due to multicollinearity. Marginal model plots indicate somewhat linear but weak relationships for Organic, Sand, and Gravel. For both percent Clay and percent Bedrock, high incidences of zero values produce non-linear relationships suggesting a lack of significance as predictor variables. A high VIF is also indicated for Cropland. High negative correlation ($r = -0.775$) is also observed between Cropland and Gradient_mean resulting in the elimination of Cropland as a predictor. At each stage of backward elimination, reduced (or partial) models are compared for a change in the F-ratio of the fitted model.

During backwards elimination, varying three-predictor models demonstrate statistical significance; however, the third variable is not consistent over all seven flood frequency quantiles. Results indicate that including Mean_precip as a variable increases model fit for flood quantiles $Q_{1.25}$, Q_2, Q_5, and Q_{25}, whereas including Rainfall_Days improves the goodness-of-fit for the Q_{10} model. Alternatively, the inclusion of Organic is shown to have some statistical significance ($p < 0.05$) as a third predictor variable for flood quantiles Q_{50} and Q_{100}. For all seven flood quantiles, $Q_{1.25}$, Q_2, Q_5, Q_{10}, Q_{25}, Q_{50}, and Q_{100}, backward

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences Discussions
Open Access
EGU

elimination reduces to the same single independent variable, logDrainage. As the most parsimonious model, logDrainage is shown to significantly predict discharge (p < 2.2e-16) confirming the often used single-variable relationship between drainage area and discharge. However, all seven flood quantiles also identify a two-predictor model using the variables logDrainage and percent Naturalized where t-tests demonstrate that both logDrainage and Naturalized are statistically significant (p < 2.2e-16) (Table 4).

Table 4 – The most parsimonious two-variate RFFA models for each flood quantile

Flood Quantile	Two-variable Models (i.e., logDrainage and %Naturalized)		
	Equation	Residual SE	adj R^2
$Q_{1.25}$	$\log Q_{1.25} = -0.537 + 0.917(\text{logDrainage}) - 1.179(\text{Naturalized})$	0.177	0.911
Q_2	$\log Q_2 = -0.420 + 0.928(\text{logDrainage}) - 1.196(\text{Naturalized})$	0.184	0.907
Q_5	$\log Q_5 = -0.224 + 0.916(\text{logDrainage}) - 1.194(\text{Naturalized})$	0.179	0.910
Q_{10}	$\log Q_{10} = -0.068 + 0.889(\text{logDrainage}) - 1.158(\text{Naturalized})$	0.197	0.887
Q_{25}	$\log Q_{25} = 0.090 + 0.878(\text{logDrainage}) - 1.148(\text{Naturalized})$	0.213	0.867
Q_{50}	$\log Q_{50} = 0.223 + 0.858(\text{logDrainage}) - 1.127(\text{Naturalized})$	0.251	0.818
Q_{100}	$\log Q_{100} = 0.353 + 0.837(\text{logDrainage}) - 1.100(\text{Naturalized})$	0.299	0.750

4.4 Model evaluation

For all seven flood quantile models, the addition of the percent Naturalized predictor variable reduces the residual standard error (Res SE) and increases the adjusted coefficient of determination (adj R^2). Figure 7 illustrates plots of the estimated discharge, derived from the flood frequency curves, against the fitted one- and two-variable models for six of the seven flood quantiles. For all seven flood quantiles, less scatter is observed in the two-variate model (i.e., logDrainage and Naturalized) with an increased adjusted R^2 of approximately 0.05and a lower standard error using the two-predictor model suggesting the addition of the second variable (i.e., Naturalized) improves the goodness-of-fit for all seven flood quantile models (six illustrated). The two-predictor combination of variables provides an improved explanation for the variations in discharge by nearly 5%. Generally, an increase in model scatter is observed for both one-variable and two-variable prediction as the RI increases suggesting the predictive capability decreases moving from $Q_{1.25}$ to Q_{100}.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

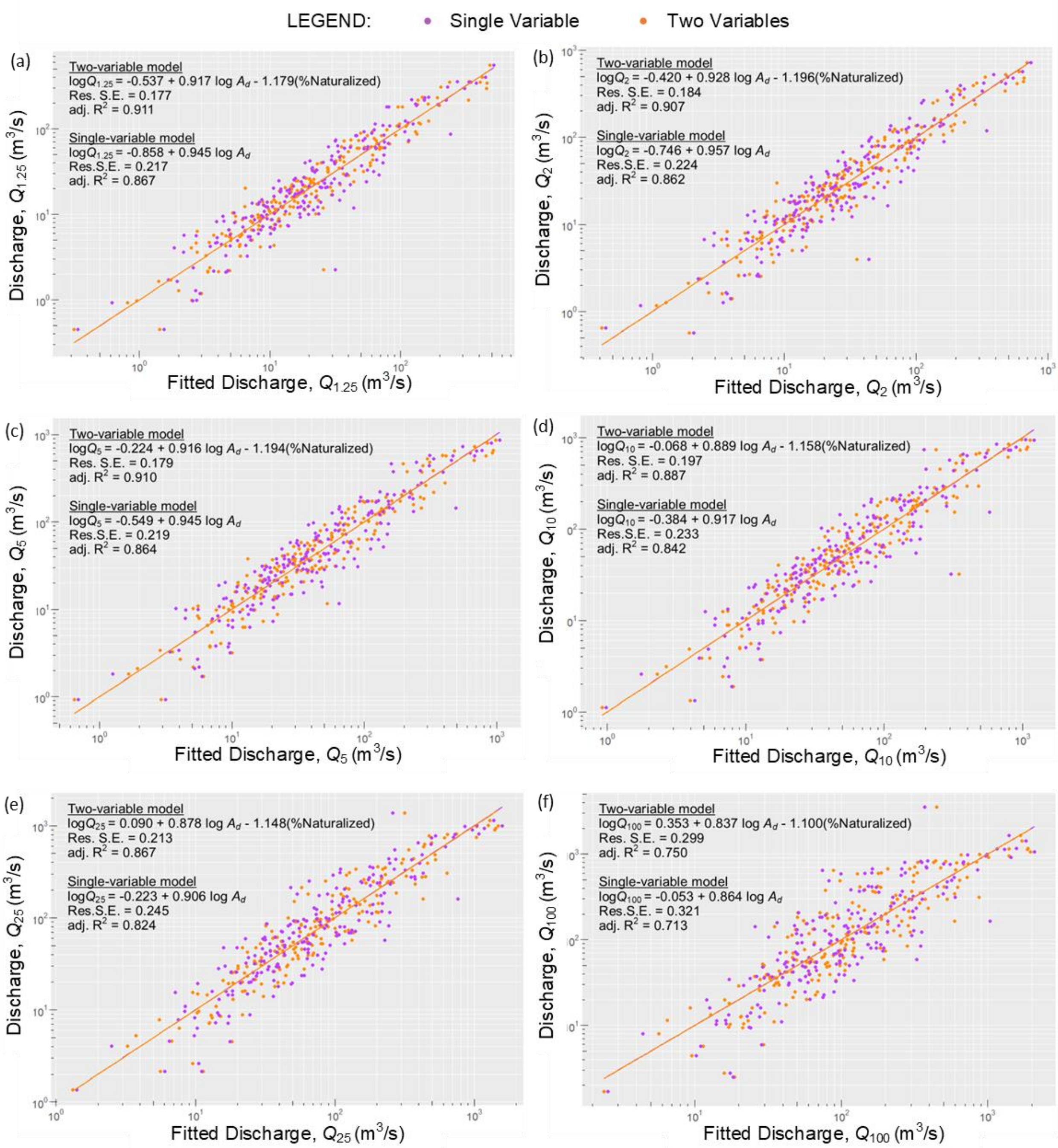

Figure 7 - The single-variable and two-variable fitted discharge models are plotted against the best-fit estimated discharge derived from the flood frequency curves for (a) a 1.25-year RI, (b) a 2-year RI, (c) a 5-year RI, (d) a 10-year RI, (e) a 25-year RI, and (f) a 100-year RI. The two-predictor log-drainage area and percent naturalized landscape model shows less scatter and a lower standard error than the single log-drainage area predictor for all six flood RIs shown. The explanation of variability in flood discharge is improved by nearly 5% using the two-predictor model for all seven flood quantiles.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences Discussions

This is consistent with the higher variance in discharge observed for large, infrequent flood events within the original gauge datasets. The predictability of larger flood events is limited by both the low frequency with which they occur, and the length of the gauge records analyzed. An analysis of variance (ANOVA) (Table 5) comparing the single-variable models to the two-variable models further indicates an improved prediction of discharge using the two-predictor model compared to the one-predictor model.

Table 5 – An analysis of variance (ANOVA) comparing the single-variable models to the two-variable models.

Flood Quantile	Analysis of Variance (ANOVA)				
	RSS Single-variable	RSS Two-variable	Sum of Sq	F-stat	Pr(>F)
$Q_{1.25}$	9.669	6.400	3.269	104.190	< 2.2E-16
Q_2	10.264	6.898	3.366	99.534	< 2.2E-16
Q_5	9.850	6.497	3.353	105.300	< 2.2E-16
Q_{10}	11.081	7.925	3.156	81.225	< 2.2E-16
Q_{25}	12.343	9.241	3.102	68.480	1.65E-14
Q_{50}	15.859	12.872	2.987	47.341	7.18E-11
Q_{100}	21.127	18.279	2.848	31.779	5.70E-08

For all seven flood quantiles, a decrease in the sum of squares of residuals is observed with the addition of the Naturalized predictor and an F statistic ($p < 0.001$) that demonstrates very strong evidence in favour of the two-predictor model.

5. Discussion

Flood magnitude, frequency and duration are primary drivers of channel erosion and stream morphology (Taniguchi & Biggs, 2015). High-magnitude, less-frequent floods will undoubtedly result in significant alterations to a channel's morphology and are more important when considering hazards, loss of life and infrastructure damage (Onen & Bagatur, 2017), however, the cumulative effects of more frequent, lower-magnitude floods can also be geomorphically more effective in altering channel form (Church & Ferguson, 2015; Wolman & Miller, 1960; Wolman & Gerson, 1978). Consequently, for effective risk management and hazard prevention, it is useful to model flows of different flood RIs when considering flood frequency as a predictive tool to better understand a river's morphological response to discharge (Basso et al., 2016*).* The best estimation of extreme flood events, however, is limited by the availability and accuracy of recorded gauge data, the length of the observed flood series, and the presence or absence of extreme flood occurrences within a flow record (Odry & Arnaud, 2017). This analysis uses a broad range of high- and low-frequency flood estimates from long-term historical flow data to develop a reliable RFFA for urban planning and infrastructure engineering. It is common practice to develop an RFFA relating the drainage area

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

of a catchment to channel discharge using a single-variable power-law relationship. Research suggests that physiographic features, such as those inherited by southern Ontario's glacial legacy, and anthropogenic land use, for example southern Ontario's clustered urbanization and widespread agricultural development, can influence a region's hydrogeomorphic response, particularly in smaller watersheds (Royall, 2013). The objective of this study was to derive a dependable RFFA using a multi-variate approach for a region influenced by glacial conditioning and varying land use, while also considering the hydrologic influences of climate and topography. The potential improvement on a generally accepted single-variate RFFA model is assessed.

In this study, rigorous goodness-of-fit testing of annual maximum mean daily discharge data series from 207 hydrometric gauge stations shows that 42.5% of gauge records are most suited to a 2-parameter LN distribution, 31.9% to a 2-parameter EV1 distribution, 21.7% to a 3-parameter LP3 distribution, and 3.9% to a 3-parameter GEV distribution in southern Ontario. This suggests that all four distributions are potentially suitable for modelling flood extremes in heterogenous regions. The model selection criteria favoured a 2-parameter model over a 3-parameter model in 74.4% of cases. This is consistent with other studies which found that selection criteria demonstrate a predisposition towards the most parsimonious model (i.e., fewest distribution parameters) (Onen & Bagatur, 2017; Laio et al., 2009; Farooq et al., 2018). Most notably, the 2-parameter EV1 model is optimal five (5) times more frequently than its 3-parameter parent model, the GEV distribution, which is only found appropriate for use in 3.9% of cases. This finding is similar to that of Laio et al. (2009) where the GEV distribution was only selected in a limited number of cases when modelling the annual maxima of peak discharge in 1000 United Kingdom basins.

Often, flood estimation will universally apply a fixed probabilistic model to historical gauge data (Di Baldassarre et al., 2009). Other southern Ontario studies have employed a blanket LP3 probability distribution to model the Q_2 flood frequency (Annable, 1995; Phillips & Desloges, 2014). However, the variation of statistical distributions identified as an optimal fit in this study suggests a need for careful, systematic model selection criteria when fitting observed flow data in regions with variable land use or other hydraulic influences (i.e., geomorphology, substrate materials, climate, or topography). To prevent an over-estimation or, more importantly, an under-estimation of discharge when predicting flood recurrence, model goodness-of-fit should be evaluated. The results of this study indicate that a 2-parameter LN statistical distribution will provide an optimal fit for 43% of the southern Ontario flood records when a broad range of flood quantiles are being examined.

The flood quantiles explored in this study (i.e., $Q_{1.25}$, Q_2, Q_5, Q_{10}, Q_{25}, Q_{50}, and Q_{100}) are consistent with return periods in other flood frequency analyses (Ahn & Palmer, 2016; Basso et al., 2016; Onen & Bagatur, 2017; Hollis, 1975). Other studies have explored a variety of novel regionalization approaches. Di Lazzaro et al. (2015) presented an RFFA using a single-variable parameterization of drainage density. Ahn and Palmer (2016) estimate flood frequency using the GEV distribution and then proposed regionalization methods using a spatial proximity approach. However, regionalization based on spatial proximity assumes that nearby sites are more similar than distal sites (Odry & Arnaud, 2017). In a glacially conditioned landscape, such as the southern Ontario region, the configuration of glacial deposits (Figure 5(a)) often forms drainage divides that segregate neighbouring catchments with diverse flood characteristics. This study, therefore, estimates channel discharge based on the

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

optimal statistical model for individual gauge data series and explores regionalization through a multi-variate regression-based approach to capture the variability of upstream hydrologic controls that are often dependent on the spatial arrangement of post-glacial physiographic features and, in the case of southern Ontario, the regionally clustered urbanization (Figure 5(d)). The mapping of surficial material, climate conditions, topography, and land use illustrates the variability of hydrologic influences on the region. Consistent with the agricultural land use of southern Ontario, analysis reveals a negative correlation is observed between Cropland and Gradient_mean. Regions of steep gradient are not typically associated with areas of high agricultural activity, whereas lower gradient regions provide much of the agricultural/cropping activity. Since crops are typically cultivated in areas with favourable conditions for growth (i.e., precipitation and gradient), this suggests that tilled cropland is a poor regional predictor for discharge if an elevation cluster is to be represented due to its collinear relationship with key elevation variables relevant to channel discharge. Similarly, the high spatial variability in surficial geology of southern Ontario (due to its glacial conditioning) is problematic. The attempt here to capture sub-watershed surficial materials using a single dominate material does not account for the detailed spatial variability of material heterogeneity (till, glaciolacustrine, glaciofluvial, etc.). Contrasting geomorphic conditions between catchments are represented by, for example, high negative correlations among Diamicton and Sand.

During the backward elimination process of variables in section 4.3.3, different land use, geomorphic, climatic, and topographic variables assume different importance in predicting channel flow depending on the flood magnitude being modelled. For example, precipitation shows a greater predictive relationship of channel discharge for lower magnitude, more frequent flood events, whereas surficial geology (e.g., Organic and Bedrock) has more predictive value for high magnitude, less frequent flood events. This may be due to the ability for precipitation to infiltrate the surface before contributing to surface runoff. During low-magnitude flood events, it is unsurprising that a fluvial system's hydrological response is more directly related to the amount of rainfall or snowmelt infiltration. However, during less frequent, high-magnitude or flash flood events, surface saturation is closely tied to surficial material properties that limit infiltration and contribute to surface runoff.

Although the most parsimonious model for estimating discharge is found to be the generally accepted and efficient single-variable relation between discharge and drainage area, when considering model variance, the two-predictor combination of upstream drainage area and the regional percentage of naturalized landscape (percent Naturalized) shows a 5% improvement when explaining variation in flood discharge for all RIs tested (i.e., 1.25, 2, 5, 10, 25, 50, and 100 years). An analysis of variance (ANOVA) further indicates a statistically significant improvement in prediction of discharge using the two-predictor model (i.e., logDrainage and percent Naturalized) compared to the single-predictor model (i.e., logDrainage). Percent Naturalized is important because it reflects areas within a catchment that have enhanced water storage compared to urban or agricultural areas. These findings are important for situations when it is necessary to reduce uncertainty in flood prediction. Plots comparing the single- and two-predictor models demonstrate less scatter for all seven flood quantiles. Generally, an increase in model scatter is observed for both one-variable and two-variable prediction as the RI increases suggesting the predictive capability lessens moving from $Q_{1.25}$ to Q_{100}. This finding is similar that of Basso et al. (2016) where model

Hydrology and Earth System Sciences Discussions

performance is better for short and intermediate return intervals. Any flood frequency analysis is limited by the length of the flow records being analysed. Since the average length of gauge records used in this study is 42.5 years, a decrease in model reliability is anticipated as the non-linear hydrological processes of the region are extrapolated. Despite careful selection of the candidate statistical distributions to "best fit" the observed flow records, the absence of large flood events captured within the sample data can skew the estimation of flood frequency for low-probability, low-frequency events (Odry & Arnaud, 2017).

Human landscape alterations that impact drainage density will influence rates of overland flow and channel flow, exerting additional influence on hydrological processes and stream response and, subsequently, impacting the magnitude and frequency of peak channel flows (Taniguchi & Biggs, 2015). Changes to land cover, such as deforestation, conversion to cropping and urbanization, typically decrease infiltration which increases discharge, and alters flood magnitude (Chin et al., 2013; Royall, 2013). It follows that the presence of reforested or natural areas will have a significant influence on modelled discharge. Since the early 1900s, select areas of southern Ontario have been reforested in recognition of wasteful clearing of marginal and submarginal agricultural lands by early settlers (Armson et al., 2001). The Naturalized variable includes tallgrass landcover, mixed forests, cultivated tree plantations, swamps, wetlands, and open water areas, representing areas of high infiltration or the surface storage of water. The negative coefficient for the percentage of naturalized area reduces the weight of the drainage area input. This is consistent with the theoretical expectation that drainage area of sub-basins with a high percentage of naturalized areas may be overemphasized without the appropriate correction for surface water storage. Although urbanization has been shown to have the most profound influence on fluvial system response, altering hydrological processes through a decrease in infiltration, an increase in overland flow and a potential decrease in groundwater recharge (Chin et al., 2013), the regional impact of clustered urban populations of southern Ontario is diluted by the expansive regions of cropland, grazing, and naturalized areas that separate them. Consequently, the percent Urban variable showed minimal significance in the multivariate regression. Similarly, the percent Cropland was shown to be a poor regional predictor for discharge due to a collinear relationship with other predictors. Since agricultural crops are typically cultivated in areas with favourable conditions for crop growth (i.e., precipitation and gradient), tilled Cropland failed to demonstrate statistical significance when modelling discharge. The statistical significance of percent Naturalized land use, however, suggests that the percentage of a sub-basin that is naturalized can be an effective variable to represent temporary surface water storage, limiting the impact to a channel during flood events.

6. **Conclusions**

The primary objective of this research is to explore a regional multi-variate flood frequency approach to transfer flood discharge information from gauged sites to ungauged sites in a low-relief, glacially conditioned landscape. The main conclusions of this analysis are:

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

1) When modelling the annual maximum mean daily discharge records for southern Ontario, 42.5% were most suited to a 2-parameter LN distribution, 31.9% to EV1, and 21.7% to LP3, and 3.9% to a GEV distribution suggesting all four distributions tested are potentially suitable for modelling flood extremes in a heterogeneous landscape. The variation of "best fit" probability distributions indicates that systematic model selection criteria is necessary when fitting observed flow data in regions with variable land use or other hydraulic influences (i.e., geomorphology, climate, or topography).
2) The percentage of tilled cropland is a poor regional multi-variate predictor for discharge if an elevation cluster is also explored. Extensive agricultural land use occurs in regions most favourable for crop growth (in terms of precipitation and topographic relief) resulting in a collinear relationship that favours the inclusion of key topographic variables over tilled cropland to explain channel discharge.
3) While land use, geomorphology, material type, climate, and topographic variables are variably important on the flood magnitude being modelled, the results here show the most parsimonious predictor for estimating discharge in ungauged streams is the accepted and efficient single-variable, drainage area.
4) However, when considering model variance, a two-predictor combination of upstream drainage area and the regional percentage of naturalized landscape shows a statistically significant 5% improvement when explaining variation in flood discharge for a broad range of recurrence intervals tested (i.e., 1.25, 2, 5, 10, 25, 50, and 100 years). The negative coefficient associated with the percentage of Naturalized area reduces serves as a correction to the drainage area relationship to account for surface water storage. This finding is important for situations when it is necessary to reduce uncertainty in flood prediction.

In summary, the findings suggest that applying a zonal two-variable model, which accounts for drainage area and the percentage of upstream naturalized land use, serves as a correction for surface water storage when modelling flood magnitude for high- and low-frequency flood events. This improvement is of value when considering the geomorphic response of channels (e.g., width to depth ratio's) to predicted channel discharge for a broad range of flood recurrence intervals.

Appendix A: Probability distribution functions

The GEV distribution uses a three-parameter probability distribution function such that

$$F(x) = \begin{cases} exp\left(-\left(1-\varepsilon\frac{x-\mu}{\sigma}\right)^{1/\varepsilon}\right) & \varepsilon \neq 0 \\ exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right) & \varepsilon = 0 \end{cases} \tag{A.1}$$

where μ, σ and ε are the location, scale, and shape parameters of the flow data, respectively. The location parameter describes the shift of a distribution along the horizontal axis, while the scale and shape parameters describe the spread (Zhang et al., 2020). The GEV blends the Gumbel (EV1), Frechet and Weibull distributions which are nested models within the GEV

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences Discussions

distribution (Laio et al., 2009). The simplified EV1 distribution uses the GEV function where the shape parameter, ε, is reduced to zero, giving the two-parameter probability distribution function

$$F(x) = exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{A.2}$$

where μ is the location parameter and σ is the scale parameter. Consideration of the three-parameter GEV distribution balances model bias versus model variance. The more complicated three-parameter GEV distribution reduces model bias compared to the two-parameter EV1 distribution, however, as the number of parameters increases, variance typically increases (Laio et al., 2009). The LN distribution is the log-transformed two-parameter Normal or Gaussian distribution represented by the probability distribution function

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{A.3}$$

also applying μ and σ as location and scale parameters, respectively. Similarly, the LP3 distribution is the log-transformed three-parameter Gamma or Pearson Type III identified by the probability distribution function

$$F(x) = \frac{1}{|\sigma|\Gamma\varepsilon}\left(\frac{x-\mu}{\sigma}\right)^{\varepsilon-1} exp\left(-\frac{x-\mu}{\sigma}\right) \tag{A.4}$$

where μ, σ and ε are the location, scale, and shape parameters, respectively. Pearson Type III and Normal distributions are converted to LP3 and LN distributions when the data are log-transformed at the outset (Di Baldassarre et al., 2009).

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Discussions

Data availability statement:
The data that support the findings of this study are available from the corresponding author, PET, upon reasonable request.

Author contributions:
Pamela E. Tetford: Conceptualization (lead); methodology (lead); investigation (lead); formal analysis (lead); writing – original draft (lead); writing – review and editing (equal).
Joseph R. Desloges: Conceptualization (supporting); methodology (supervision); investigation (supervision); formal analysis (supervision); writing – original draft (supporting); writing – review and editing (equal).

Competing interests:
The authors declare that they have no conflict of interest.

Acknowledgements:
This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSER-CGS-D graduate scholarship to P. E. Tetford), and research funding from the University of Toronto to J. R. Desloges. Many thanks to Esther Bushuev who helped compile data of the WSC flow gauges for analysis. Academic advice and insights are greatly appreciated from Marney Isaac, Carl Mitchell, and Michael Widener.

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

References

Ahn, K. H. & Palmer, R. (2016). Regional flood frequency analysis using spatial proximity and basin characteristics: Quantile regression vs. parameter regression technique. *Journal of Hydrology*, *540*, 515–526. https://doi.org/10.1016/j.jhydrol.2016.06.047

Annable, W. K. (1995). Morphological relationships of rural water courses in southwestern Ontario for use in natural channel designs. ProQuest Dissertations Publishing.

Annable, W. K. (1996). *Database of morphologic characteristics of watercourses in Southern Ontario*. Toronto, Ontario: Ontario Ministry of Natural Resources.

Annable, W.K., Lounder, V.G., & Watson, C.C. (2011). Estimating channel-forming discharge in urban watercourses. *River Research and Applications*, *27*(6), 1535-1459. https://doi.org/10.1002/rra.1391

Armson, K.A., Grinnell, W.R. & Robinson, F.C. (2001). History of reforestation in Ontario. In Wagner, R.G. & Colombo, S.J. (Eds.), *Regenerating the Canadian Forest: Principles and Practice for Ontario* (pp. 3-22). Ontario Ministry of Natural Resources. Fitzhenry & Whiteside Limited.

Ashmore, P. (2015). Towards a sociogeomorphology of rivers. *Geomorphology*, *251*, 149–156. https://doi.org/10.1016/j.geomorph.2015.02.020

Basso, S., Schirmer, M., & Botter, G. (2016). A physically based analytical model of flood frequency curves. *Geophysical Research Letters*, *43*(17), 9070–9076. https://doi.org/10.1002/2016GL069915

Bevan, A. & Conolly, J. (2009). Modelling spatial heterogeneity and nonstationarity in artifact rich landscapes. *Journal of Archaeological Science*, *36*, 956-964.

Chapman, & Putman, D. F. (1984). *The physiography of southern Ontario / by L.J. Chapman and D.F. Putnam.* (3rd ed.). Ontario Ministry of Natural Resources.

Chin, A., O'Dowd, A. P., & Gregory, K. J. (2013). *Urbanization and River Channels*. In *Treatise on Geomorphology* (Vol. 9). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-374739-6.00266-9

Church, M., & Ferguson, R. I. (2015). Morphodynamics: Rivers beyond Steady State. *Water Resources Research*, 51 (4), 1883–97. https://doi.org/10.1002/2014WR016862.

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, *18*(3–4), 257–271. https://doi.org/10.1016/0022-1694(73)90051-6

Desloges, J. R., Phillips, R. T. J., Byrne, M.-L., & Cockburn, J. M. H. (2020). Geomorphology of the Great Lakes Lowlands of Eastern Canada BT - Landscapes and Landforms of Eastern Canada. In O. Slaymaker & N. Catto (Eds.) (pp. 259–275). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-35137-3_11

Di Baldassarre, G., Laio, F., & Montanari, A. (2009). Design flood estimation using model selection criteria. Physics and Chemistry of the Earth, 34(10–12), 606–611. https://doi.org/10.1016/j.pce.2008.10.066

Di Lazzaro, M., Zarlenga, A., & Volpi, E. (2015). Hydrological effects of within-catchment heterogeneity of drainage density. *Advances in Water Resources*, *76*, 157–167. https://doi.org/10.1016/j.advwatres.2014.12.011

Donnan, J. A. (2008). *Economic implications and consequences of population growth, land use trends, and urban sprawl in southern Ontario*. Environmental Commissioner of Ontario.

Dunne, T., and Leopold, L.B. (1978). Water in environmental planning: New York, W.H. Freeman and Company, 818 p.

Environment and Climate Change Canada (ECCC) (2019). Historical Hydrometric Data. Retrieved from https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences Open Access EGU Discussions

Environment and Climate Change Canada (ECCC) (2020). Canadian Climate Normals. Retrieved from https://climate.weather.gc.ca/climate_normals/

Farooq, M., Shafique, M., & Khattak, M. S. (2018). Flood frequency analysis of river swat using Log Pearson type 3, Generalized Extreme Value, Normal, and Gumbel Max distribution methods. *Arabian Journal of Geosciences*, *11*(9). https://doi.org/10.1007/s12517-018-3553-z

Fulton, R.J., Karrow, P.F., LaSalle, P., & Grant, D.R. (1986). Summary of quaternary stratigraphy and history, eastern Canada. Quaternary Science Reviews, 5, 211-228. https://doi.org/10.1016/0277-3791(86)90188-5

Ghunowa, K., MacVicar, B. J., & Ashmore, P. (2021). Stream power index for networks (SPIN) toolbox for decision support in urbanizing watersheds. *Environmental Modelling & Software*, *144*(August), 105185. https://doi.org/10.1016/j.envsoft.2021.105185

Hollis, G. (1975). The effect of urbanization on floods of different recurrence interval. *Water Resources Research*, *11*(3), 431–435. https://doi.org/10.1029/WR011i003p00431

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, *29*(2), 271–281. https://doi.org/10.1029/92WR01980

Javelle, P., Ouarda, T. B. M. J., & Bobée, B. (2003). Spring flood analysis using the flood-duration-frequency approach: Application to the provinces of Quebec and Ontario, Canada. *Hydrological Processes*, *17*(18), 3717–3736. https://doi.org/10.1002/hyp.1349

King, J. R., & Jackson, D. A. (1999). Variable selection in large environmental data sets using principal components analysis. *Environmetrics*, *10*(1), 67–77. https://doi.org/10.1002/(sici)1099-095x(199901/02)10:1<67::aid-env336>3.0.co;2-0

Laio, F., Di Baldassarre, G., & Montanari, A. (2009). Model selection techniques for the frequency analysis of hydrological extremes. *Water Resources Research*, *45*(7), 1–6. https://doi.org/10.1029/2007WR006666

Larson, G., & Schaetzl, R. (2001). Origin and evolution of the Great Lakes. *Journal of Great Lakes Research*, *27*(4), 518–546. https://doi.org/10.1016/S0380-1330(01)70665-X

Natural Resources Canada (2019). Federal hydrologic and hydraulic procedures for flood hazard delineation. Federal Flood Mapping Guidelines Series. pp. 19-24. https://doi.org/10.4095/299808

Odry, J., & Arnaud, P. (2017). Comparison of flood frequency analysis methods for ungauged catchments in France. *Geosciences (Switzerland)*, *7*(3). https://doi.org/10.3390/geosciences7030088

Onen, F., & Bagatur, T. (2017). Prediction of Flood Frequency Factor for Gumbel Distribution Using Regression and GEP Model. *Arabian Journal for Science and Engineering*, *42*(9), 3895–3906. https://doi.org/10.1007/s13369-017-2507-1

Ontario Geological Survey (OGS). (2010). Surficial Geology of Southern Ontario (MRD 128 – Revised) [digital file]. Retrieved from http://www.geologyontario.mndm.gov.on.ca/mndmaccess/mndm_dir.asp?type=pub&id=MRD128-REV

Ontario Ministry of Natural Resources (OMNR). (2005). Provincial Digital Elevation Model— Tiled Dataset, version 2.0.0 [digital file]. Ontario Ministry of Natural Resources, Peterborough, Ontario, Canada.

Ontario Ministry of Natural Resources and Forestry (OMNRF). (2019). Southern Ontario Land Resource Information System (SOLRIS) – version 3.0 [digital file]. Retrieved from https://geohub.lio.gov.on.ca/datasets/0279f65b82314121b5b5ec93d76bc6ba

Ontario Ministry of Natural Resources and Forestry (OMNRF). (2020). Ontario Watershed Boundaries [digital file]. Retrieved from https://geohub.lio.gov.on.ca/datasets/mnrf::ontario-watershed-boundaries-owb

https://doi.org/10.5194/hess-2022-411
Preprint. Discussion started: 22 February 2023

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Oudin, L., Andréassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*(3), 1–15. https://doi.org/10.1029/2007WR006240

Paixao, E., Auld, H., Mirza, M., Klaassen, J., & Shephard, M. (2011). Regionalization of heavy rainfall to improve climatic design values for infrastructure: case study in Southern Ontario, Canada. *Hydrological Sciences Journal*, *56(7)*, 1067–1089. https://doi.org/10.1080/02626667.2011.608069

Pasternack, G. B. (2013). *Geomorphologist's Guide to Participating in River Rehabilitation*. In *Treatise on Geomorphology* (Vol. 9). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-374739-6.00268-2

Phillips, R. T. J. (2014). *Alluvial floodplain classification and organization in low-relief glacially conditioned river catchments* (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Phillips, R. T. J., & Desloges, J. R. (2014). Glacially conditioned specific stream powers in low-relief river catchments of the southern Laurentian Great Lakes. *Geomorphology*, *206*, 271–287. https://doi.org/10.1016/j.geomorph.2013.09.030

Phillips, R. T. J., & Desloges, J. R. (2015). Alluvial floodplain classification by multivariate clustering and discriminant analysis for low-relief glacially conditioned river catchments. *Earth Surface Processes and Landforms*, *40*(6), 756–770. https://doi.org/10.1002/esp.3681

Royall, D. (2013). *Land-Use Impacts on the Hydrogeomorphology of Small Watersheds*. In *Treatise on Geomorphology* (Vol. 13). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-374739-6.00341-9

Tacq, J. (2010). Multivariate Normal Distribution. In International Encyclopedia of Education (Vol. 7, pp. 332–338).

Taniguchi, K. T., & Biggs, T. W. (2015). Regional impacts of urbanization on stream channel geometry: A case study in semiarid southern California. *Geomorphology*, *248*, 228–236. https://doi.org/10.1016/j.geomorph.2015.07.038

Thayer, J.B., Phillips, R.T.J., & Desloges, J.R. (2016). Downstream Channel Adjustment in a Low-Relief, Glacially Conditioned Watershed. *Geomorphology* 262: 101–11. https://doi.org/10.1016/j.geomorph.2016.03.019.

United States Geological Survey (2019). *Guidelines for Determining Flood Flow Frequency Bulletin 17C Book 4, Hydrologic Analysis and Interpretation*. U.S. Department of the Interior, U.S. Geological Survey, Reston, VA, pp. 23–34.

US Army Corps of Engineers (2019). HEC-SSP Statistical Software Package Version 2.2 User's Manual. Retrieved from https://www.hec.usace.army.mil/software/hec-ssp/documentation/HEC-SSP_22_Users_Manual.pdf

Vocal Ferencevic, M. & Ashmore, P. (2012). Creating and evaluating digital elevation model-based stream-power map as a stream assessment tool. *River Research ad Applications*, *28*(9), 1394-1416. https://doi.org/10.1002/rra.1523

Winter, Thomas C. (2001). The Concept of Hydrologic Landscapes. *Journal of the American Water Resources Association* 37 (2): 335–49. https://doi.org/10.1111/j.1752-1688.2001.tb00973.x.

Wolman, M. G., & Gerson, R. (1978). Relative scales of time and effectiveness of climate in watershed geomorphology. *Earth Surface Processes*, *3*(2), 189–208. https://doi.org/10.1002/esp.3290030207

Wolman, M., & Miller, J. (1960). Magnitude and Frequency of Forces in Geomorphic Processes. *The Journal of Geology*, *68*(1), 54-74. Retrieved September 11, 2021, from http://www.jstor.org/stable/30058255

Zhang, Z., Stadnyk, T. A., & Burn, D. H. (2020). Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. *Canadian Water Resources Journal*, *45*(1), 43–58. https://doi.org/10.1080/07011784.2019.1691942